

Transport Gateways and Trade Corridors. Portes du transport et corridors de commerce.

Canadian Transportation Research Forum
Proceedings of the 31st Annual Conference
Winnipeg, Manitoba
May 26 - 29, 1996

VOLUME 2

Le Groupe de Recherches sur les Transports au Canada
Actes de la 31ieme Conference Annuelle
Winnipeg, Manitoba
26 au 29 mai, 1996

Volume II

Volume I

Opening and Plenary Session/Séance plénièré d'ouverture

Session/Séance 1: *Gateway Ports: Competing for International Commerce*

Session/Séance 2: *Performance and Structure of the Canadian Trucking Industry*

- *Rethinking Strategy for Global Competitveness: The Logistics Focus.* Dr. Vedanand, University of Manitoba 1
- *Best Practices in the North American Trucking Industry,* Jan Bowland and Don McKnight, KPMG Management Consulting 11
- *Trucking in a Borderless Market: A Profile of the Canadian Trucking Industry, 1988 to 1994,* Fred Barzyk , Transportation Division, Statistics Canada 24

Session/Séance 3: *Logistics Demand Modelling and Assessment of Economic Impact*

- *Macroeconomic Impact and Benefit/Cost Analysis of Transportation and Mining Developments in the Northwest Territories,* Russell D. Neudorf and Dr. Masood U. Hassan, Department of Transportation, Government of the Northwest Territories 42
- *Perspectives d'évolution de la demande de transport de merchandises au Québec,* Yvon Bigras, Université du Québec à Trois-Rivières, Serge Le Brun, Institut Maritime du Québec, and Jacques Roy, Université du Québec à Montréal 56
- *Impact of Transportation Costs Variations on the Relative Market Shares for Different Transportation Modes: A Network Model - the Case of Dry Bulk Transport on the Benelux/Germany to Spain Corridor,* Bart Jourquin, Laurent Demilie and Michel Beuthe, Facultés Universitaires Catholiques de Mons 71

Session/Séance 4: *Economic Development of Gateway Grain Ports*

- *Gateway North Marketing Agency - Defining the Start Up,* Don Taylor, SRP Management Consulting Associates Ltd 80
- *The Relative Competitive Position of the Port of Prince Rupert,* Eugene Ludwick, Prince Rupert Port Corporation 96
- *The Influence of the Sepetiba Port Complex on Brazilian Competitiveness in the International Grain Market,* Lieut. Col. Luis A. Silveira Lopes and Paulo C. Barroso Magalhães, IME - Instituto Militar de Engenharia 107

Session/Séance 5: *Panel: Trends in Transport/Logistics Education*

- *Trends in Logistics Professional Development: Skills vs. Concept: - Case Study: The Logistics Institute,* Victor S. Deyglio, Canadian Professional Logistics Institute and David Cape, Ryerson Polytechnic University 120
- *Integrating the Case Method in Transportation and Logistics Education,* Louis A. Le Blanc and Conway T. Rucks, University of Arkansas at Little Rock 134

Session/Séance 6: *The Historical Development of Gateways and Corridor Routes*

- *The Trade Corridor that is the Rhine-Main-Danube Waterway,* Barry Largent Brown, University of Exeter 144
- *Gateway to the Triangle: Railway Development in Western Canada,* John Hall, Transportation Planner 154
- *Winnipeg: Gateway to the Mid-Continent Corridor,* Barry E. Prentice, Transport Institute, University of Manitoba 171

Session/Séance 7: *Issues in Rail Transportation*

- *Implications jurdiques du projet de loi fédéral C-101, loi sur les transports au Canada,* Louis Doyle, Ministère des Transports du Québec 187
- *Demand Model Development to Assess High Speed Passenger Train Markets on Windsor to Quebec City Corridor - An Update,* Dr. Deepak Ekbote and Dr. Richard Laferriere, Research and Analysis, CP Rail System 195
- *VIA Rail's Strategic Plan for Success,* Gerry Kolaitis, VIA Rail Canada Inc. 200

Session/Séance 8: *Aviation - Competition for International Traffic Corridors*

- *The Canadian Aviation Charter Market: Opening New Transport Gateways or Niche Market Pricing?,* Gordon G. Baldwin, Statistics Canada 212
- *The 1995 Canada-U.S. Air Services Agreement: A Preliminary Analysis,* Michael W. Pustay, Texas A&M University 222
- *Cost Competitiveness of Airline: An International Comparison,* Tae Oum and Chunyan Yu, University of British Columbia (abstract only) 237

Session/Séance 9: *EDI/Information Technology: Evolution and Revolution*

- *New Customs EDI System Set to Change International Clearance Procedures,* Barry Frain, Expert Customs Brokers . . 240
- *Strategic Alliances Enabled by Electronic Commerce - An Overview of the Process,* Alison Lough Sunstrum, Consultant, Alternative Solutions., and F.H. (Joe) Howard, Transportation Consultant 252
- *Information Technology and Transformation in the Grain Industry: The Impact of Logistical and Market Intelligence on Industrial Structure,* Dr. Greg Mason, Prairie Research Associates Inc. 266

Session/Séance 10: *Coastal Gateways and Trade Corridors*

- *Un dervice maritime roulier sur le Saint-Laurent,* Réjean Leclerc, Ministère des Transports du Québec 280
- *Coastal Gateways: Costs and Subsidies on British Columbia Coastal Ferry Routes,* W.G. Waters II, James Evans and Joan Caravan, The University of British Columbia 293
- *Analyse et intégration des statistiques sur le transport multimodal des marchandises,* Louis Pierre, Statistics Canada . . 310

Session/Séance 11: *Farm to Market Grain Assembly and Shipment*

- *Estimating the Value of Rail Car Accessibility for Grain Shipments: A GIS Approach,* Eric L. Jessup, John L. Ellis and Ken L. Casavant, Washington State University 325
- *Farm to Market Grain Movement,* Arthur Friesen and Jerry Fruin, University of Minnesota 339
- *"Not with a Bang but a Whimper" -- The Birth, Life and Death of the Crow Rates: Lessons for Policy Makers in an Uncertain World,* Paul Earl, Western Wheat Growers 351

Session/Séance 12: *Competition and Opportunities of Airport Gateways*

- *The Deregulation of Airports and the Opportunities of Gateway Development,* Fredrick Petrie, Aviation Associates 369
- *Canada's Airports, Their Role as Cargo Gateways,* Andrea Mathieson, Statistics Canada 379
- *Airport Privatisation in Canada and Australia: Experiments in Structures of Ownership and Control,* John Street and David Smith, BTCE Australia and University of British Columbia . . . 393

Session/Séance 13: *Economics and Environmental Impacts of Transportation*

- *User Pay Transportation: Full Costs and Pricing Policy Implications,* Neal A. Irwin, Lee Sims and Joseph Jones, IBI GROUP 407
- *The Control of Carbon Dioxide Emissions from Freight Transportation: A Case Studies Approach,* David Hackston, Richard Lake and Louis-Paul Tardiff, Research and Traffic Group 422
- *From Isolation to Integration: Organizational Alternatives to Airport Environmental Planning,* Anthony Perl, University of Calgary 434

Session/Séance 14: *Panel : The East/West Movement of Grain*

- *Grain Handling and Transportation: Capacity, Rates and Tariffs,* Dr. John Heads and Dr. Arthur G. Wilson, Transport Institute, University of Manitoba 443
- *Thunder Bay - Grain Gateway to the East?* Dr. Arthur G. Wilson and Dr. John Heads, Transport Institute, University of Manitoba 459
- *Rail Rationalization and the Grain Industry in Southern Ontario* Dave Hackston, Research and Traffic Group 470

Session/Séance 15: *Infrastructure Investments and Corresponding Economic Impacts*

- *Highway Infrastructure Capital and Productivity Growth: Evidence from Canadian Goods-Producing Sectors,* Bilkis R. Khanam, Transport Canada 479
- *The Role of Transport in determining the Demand for Industrial Land,* Larry F. Loreth, City of Winnipeg 492
- *Highways Costs and Revenues in Québec: Evidence and Analysis,* Michel Boucher, Université du Québec 507

Session/Séance 16: *Sea Ports: Competition and Re-structuring under Privatization*

- *Analyse comparative des ports de Montréal et de Baltimore,* Yannick Unvoas, Université du Québec à Rimouski and Brian Slack, Université Concordia 523
- *Restructuring Canadian Ports: Prospects and Problems,* Michael C. Ircha, University of New Brunswick 539

Volume II

Session/Séance 17: *Grain Logistics from a Continental Perspective*

- *The Logistics of North American Wheat Exports,* Raquel Christie and Gary Warkentine, Agriculture Canada (abstract only) 553
- *Grain Transportation in a Continental Perspective,* Javier Caceres, Consultant and Darrell Richards, Transport Concepts . 555
- *A Short-Run Analysis of Inland U.S. Barge Rates for Grain,* James W. Dunn and Gregory L. Harnish, Pennsylvania State University 566

Session/Séance 18: *Transportation Research Methods and Results*

- *The Use of Hand-Held Computers in Roadside Surveys,* Greg Little and Doug Smith, Ministry of Transportation, Ontario ... 583
- *A Conjoint Analysis of Air Carrier Loyalty,* Edward Bruning, Barry E. Prentice and Patty-Jo Bellamy, University of Manitoba 599

Session/Séance 19: *Legal Issues in Cargo Handling: Safety, Pollution and Claims*

- *Apportioning the Risks in Intermodal Transport: Impacts of Proposed Revisions to the U.S. Carriage of Goods by Sea Act,* Hugh M. Kindred, Dalhousie University 613
- *Chemical Transportation Risk Management,* John Cowan, KPMG Management Consulting 625
- *Establishing Insurance Requirements for an Evolving Shortline Railway Industry: The Ontario Experience,* Mike Ronson and Ann Misetich, Ministry of Transportation, Ontario 638

Session/Séance 20: *Panel: Grain Transportation Challenges for the 21st Century*

- Tami Reynolds, Organizer

Session/Séance 21: *Trucking in the NAFTA Environment*

- *NAFTA and Truck Configurations: A Framework for Assessments,* Amiy Varma and Nabil Suleiman, North Dakota State University 647
- *Identifying Needed Transportation Investment, by Corridor, to Support NAFTA Trade Flows in Washington State,* Ken A. Eriksen and Ken Casavant, Washington State University 662

- *NAFTA Trade Corridors - The "Unlevel Playing Field" for International Carriers,* R. Ian Kingham, P.Eng., GMK Transportation Planning and Engineering Ltd. 677

Session/Séance 22: *Public Policy and Economic Performance*

- *In Defence of a Transportation "Systems" Approach To Transportation Policy Development,* Derek Lett, Ontario Ministry of Transportation 690
- *Public Transportation in the Face of Deteriorating Economic Conditions and Colonial-Dependent Spatial Policies in Cameroon,* Ambe Njoh, Ph.D., University of South Florida . . . 704
- *Performance Indicators of the Transport Sector,* Jean-Pierre Roy, Transport Canada 719

Session/Séance 23: *Panel: Regional Railroads and Short-line Rail Issues*

- *An Overview of the Short Line and Regional Railroad Industry in the United States,* Andrew Jennings, Mercer Management Consulting, Inc. 734
- *Characteristics of U.S. Grain Short-Line Railroads: Implications for the Prairie Provinces,* Denver Tolliver, John Bitzan, Trent Byberg, Upper Great Plains Transportation Institute, North Dakota State University 749

Session/Séance 24: *Trucking Safety and Driver Education*

- *Truck Safety - Perception and Reality,* Greg Little, Ministry of Transportation, Ontario 762
- *Canada's Trucking Industry Moves Toward a Nationally Certified Program of Driver Training,* Lloyd L. Ash, Trimac Consulting Services 776
- *Controlling Highway Safety – Commercial Driver Qualifications,* Michèle Delaquis and Fred P. Nix 792

Session/Séance 25: *Empirical Analysis and Forecasting Traffic Demand/Movements*

- *Estimating O-D Travel from Link Traffic Counts: Can We Really Get There From Here?,* Bob Leore, Transport Canada . . 806
- *Aggregate Air Passenger Forecasting - Winnipeg International Airport Activity Forecasts,* Darryl Howery, Ellen Taylor and Drina Culo, Applications Management Consulting Ltd. 820

Session/Séance 26: *Urban Transportation: Attracting Ridership*

- *Commercial Ventures at Park+Ride Termini,* David M. Cape, Ryerson Polytechnic University and Lancaster University 831
- *Application of Diesel Rail Car Technology,* Darrell Richards, Transport Concepts 841
- *Urban Transit Ridership in Canada: Nowhere to Go But UP!,* Mebs S. Kanji, University of Calgary 854

Session/Séance 27: *Equipment Challenges in the Truck Management*

- *Assessing Non-Conforming Vehicles for Net Transport and Safety Benefit,* John Woodrooffe, Roaduser Research International, and Douglas Johnson, John Palmer and David Smith, New Brunswick Transportation and Communications .. 869
- *Bridge Cost Allocation Issues,* Edward Fekpe, Battelle Memorial Institute 883
- *Performance-Based Truck Weight and Dimension Regulations.* Fred P. Nix, John R. Billing, and Michèle Delaquis 893

CTRF Student Paper Winners for Publication

David A. Saucy, The Role of Transportation in Canada - ASEAN Trade: Is it a Barrier to Trade? MBA Program, Dalhousie University. A.M. Stevens Award, First Prize, Masters, 1994.

Edward S.K. Fekpe, Ph.D., P.Eng., Uncertainties in Truck Regulatory Policy and Highway Infrastructure Management, Department of Engineering, University of Manitoba. Jim Davey Award, First Prize, Ph.D., 1994.

Hamidreza Soleymani, Sensitivity Analysis for AASHTO Flexible Pavement Design, Engineering Department, University of Saskatchewan. Second Prize, Ph.D., 1994.

Susan S. Dableh, Evaluation of Traffic Sign Retroreflectivity, Department of Civil Engineering, University of New Brunswick. First Prize, Undergraduate; and CSL Groyp Gold Medal (CIT), 1994.

Sharlene Wiseman, Instersection Analysis: Boulevard Decarie "Est" and Rue Jean-Talon Ouest, Department of Civil Engineering & Applied Mechanics, McGill University. Third Prize, Undergraduate, 1994.

Hélène Dufourd, Application de la Méthode de Recherche avec Tabous au Tracé d'une Linge de Métro, Département de Mathématiques et de Génie Industriel École Polytechnique de Montréal. A.M. Stevens Award, First Prize, Masters, 1995.

Darren M. Scott, The Journey to Work and Urban Spatial Structure: A Case Study of the Greater Toronto Area, 1986 to 1991, Department of Geography, McMaster University. Second Prize, Masters 1995.

El Bachir Belhadji, Using State Dependent Utility Functions in the Context of Road Safety: Theory and Evidence. Department of Economics, University of Montréal. Jim Davey Award, First Prize, Ph.D., 1995.

Curtis F. Berthelot, Numerical Load Impact Analysis of Typical Saskatchewan Roads, Department of Civil Engineering, University of Saskatchewan. Second Prize, Ph.D., and Canadian Trucking Association Award, Best Trucking Paper, 1995.

The Logistics of North American Wheat Exports

Although the United States (U.S.) and Canada account for under 20 per cent of total world wheat production, they account for over half of total world wheat exports and nearly all of the high-quality spring wheat traded. Distinctive North American geographic and climatic characteristics promote the production of a wide range of wheat classes and varieties. To accommodate this diversity, a complex transportation infrastructure has emerged to efficiently move wheat from the interior regions of production to the coastal regions for export. The objective of this paper is to examine the logistics of wheat exports in Canada and the U.S., highlighting the unique elements of each system, and to provide a map of the various wheat flows.

Canada produces primarily hard red spring and hard amber durum wheat, mainly grown in the Prairie provinces. The Canadian Wheat Board (CWB) is the sole marketing agency for wheat (and barley) grown in this region, destined for export or domestic human consumption. Canadian wheat is primarily transported by rail to one of three port destinations, mainly in an east-west direction. The majority of wheat moves west, crossing the Rocky Mountains destined for Vancouver and Prince Rupert on the Pacific coast. A smaller amount moves east, to Thunder Bay and across the Great Lakes for export from the St. Lawrence Seaway, but this avenue is limited by ice in the winter months. Some smaller volumes are transported north to the Hudson Bay and south to the United States.

In the United States, five classes of wheat are grown in four distinct growing regions. The largest class, in terms of production and exports, is hard red winter wheat grown in the Mid-west region. The second largest class is soft red winter wheat grown in the lower Eastern states, in competition with corn and soybeans. Spring and durum wheat are produced in the Great Plains region and white winter wheat in the Northwestern states. Wheat destined for export is transported mainly by rail and barge, in roughly equal amounts, with trucks moving smaller volumes close to port locations. The majority of U.S. wheat moves south to the ports of the Gulf of Mexico and west to ports of the Pacific Northwest. Smaller quantities move east through the Great Lakes/St. Lawrence Seaway and ports of the Lower Atlantic.

To demonstrate the path of wheat flows, the various growing regions and geographical constraints will be illustrated on a detailed map of Canada and the

Raquel Christie and Gary Warkentine
1 Agriculture and Agri-food Canada

United States. The map will illustrate the flow of each class of wheat from the distinct growing regions to the port destinations for the 1994-95 crop year, highlighting the relative volumes per class of wheat, transportation mode, port and overseas destination. Accompanying the map will be a paper providing additional detail on each transportation system. The flow of wheat will be followed from the farm, through the primary elevator system, along the various transportation routes, and to the final port position. The combination of the map and text will provide a complete overview of wheat flows and the rationale behind the logistical systems.

The report, with map and text, will be available from Agriculture and Agri-Food Canada, Policy Branch, Adaptation and Grain Policy Directorate, Market Analysis Division, in an upcoming issue of the *Bi-weekly Bulletin*.

Raquel Christie and Gary Warkentine
2 Agriculture and Agri-food Canada

Grain Transportation in a Continental Perspective

Javier Caceres and Darrell Richards
Transport Concepts

Summary

The Western Grain Transportation Act was much more than a subsidy. It established the relationship between grain production, transportation, handling, and marketing. Demise of the WGTA on August 1, 1995 has unleashed dynamic social, political, and economic forces that will lead to a continental grain and transportation market, and the demise of the Canadian Wheat Board and Canadian Grain Commission in their present form.

Many had assumed that WGTA reform would force Canadian institutions such as CN, CP, the Canadian Wheat Board, the Canadian Grain Commission, and the grain companies to become more competitive. But restructuring will occur on the basis of a North American market, not necessarily on Canadian terms. Some of these institutions may no longer be relevant in a continental market.

Most impact studies of WGTA reform have paid more attention to Canadian routes than continental routes. They usually assumed that CN and CP remain transcontinental Canadian carriers without access to U.S. Pacific ports, the continuation of the Board and Grain Commission, and public control of hopper cars. These assumptions are no longer valid. The current pace of mergers suggest there could be only two or three main line railways in the North American continent. These companies will rationalize rail transportation irrespective of borders. Grain grading and inspection will be downgraded in order to reduce the cost of using U.S. rail routes and ports and to conform to globalization of international grain markets.

Introduction

Throughout history, people have tried to forecast the future: first by using the stars, then by the crystal ball, and now by econometric models. This paper is not the product of empirical or quantitative analysis. We have used, what appears to us, logic and past experience to extrapolate current trends into the future. Those who venture into forecasting have sometimes been rewarded by being burned at the stake. We seek a less dramatic response.

Background

In order to project where we are headed, we must review from where we have come. Canada's grain industry and associated industrial structure was built on a nationalist policy to forge east-west transportation links and encourage the use of Canadian railroads and ports. Much of the transportation and industrial infrastructure of Western Canada was founded by government intervention and support to Canadian grain movement contrary to the more natural geographic and economic north-south alignment. The Canadian grain industry was based on three pillars which constrained the penetration of foreign grain companies in the export of Canadian grain:

a) farmer owned grain handling system

b) orderly marketing system

c) heavy government intervention promoting east-west movement of grain

The political-economic forces that created this policy and infrastructure began to decline in the 1970's as the overall force of Canadian nationalism declined. Major leaders of nationalist sentiment in the Conservative Party (for example John Diefenbaker) lost their influence on the party. This coincided with the transformation of the prairie grain cooperatives into commercial enterprises, and the entry of multinationals into the Canadian export grain industry. The political influence of market-oriented grain groups picked up while the political influence of the cooperative philosophy of grain pools declined.

The process of change usually is driven by endogenous (internal) and exogenous (external) forces. In Canada for many years the internal forces were unable to induce the economic restructuring of the grain industry. The external force of globalization has now accelerated the dismantling of Canadian grain regulations. The fundamental concept of globalization is the least-cost sourcing of similar commodities. Grain, from the transnational perspective, requires uniform grading standards in the major exporting countries. Thus, globalization will increase pressure to downgrade Canada's standards which give our grain its identity.

Canada is far more dependent on international grain markets than is the U.S. Canada's domestic market cannot absorb more than 20% of its grain production, whereas the U.S. domestic can absorb 80% of its production. Canada deals with food as a commercial economic commodity. For example, Canada opened the Chinese market and trades with Cuba. The U.S. uses the export of food as a political-economic tool (food as a weapon). For example, the U.S. has in the past embargoed grain to Russia and soybeans to Japan. Canada is exposing itself to a continental grain market, but the American market is distorted by political interference. After continental integration, this U.S. political influence could affect the Canadian export grain system. Development and harmonization of a continental grain market may also mean the end of Canada's independent and unique grain export system.

Continental Perspective

The Crow Benefit was more than a subsidy. It was the cornerstone of public policy that structured the relationship between different sectors and upon which infrastructure was based. The demise of the WGTA has set in motion a dynamic process that will lead to a continental grain and transportation market. Much of the discussion about repeal of WGTA and deregulation of the grain industry appears to assume that restructuring will occur on Canadian terms to increase the effectiveness of Canadian institutions and systems. Repeal of WGTA is unleashing dynamic economic, social, and political forces that will restructure the industry on a continental basis. Other events (future of hopper cars, privatization of CN, review of Wheat Board, rail mergers) are hastening the continental restructuring.

Railway Industry Trends

The Canadian east-west railroad infrastructure owes its existence in part to the deliberate policies of Canadian Governments to force east-west links against the grain of more natural -- and economical -- north south access. A continental perspective suggests more integration of railroads on a north-south basis. An admittedly extreme example of what may be in store for Western Canadian infrastructure is the restructuring of CP and CN in Atlantic Canada. CP now serves Montréal and Toronto out of U.S. ports rather than Saint John.

The size and scope of CN and CP have been similar, and in some cases even larger, than the Western U.S. railroads with which they compete. CN and CP have responded to the U.S. competition by cutting costs and infrastructure to increase density and productivity. But the goal posts have moved. Rather than benchmarking CN and CP to Burlington Northern and Union Pacific, we must now look at Burlington Northern Santa Fe and at a Union Pacific that has taken over North Western, that will probably take over Southern Pacific, and that has access to Mexico.

There are now two major main line railroads in the Western U.S. There soon may be only two in the Eastern U.S. Union Pacific is examining the purchase of Elgin Joliet and Eastern, which may be an interim step to a true Atlantic to Pacific U.S. merger. Conceivably there could be just two or three main railroads in the U.S. within five years (although it is always possible that political concerns about monopoly could interfere). In the context of these giant railways operating in a continental market CN and CP would be feeders or short lines.

The east-west orientation of CN and CP is changing rapidly. CP has already split off an eastern (which really is central since the section east of Montréal is no longer part of CP) business unit that faces an uncertain future. Integration of central and eastern Canadian rail infrastructure with the U.S. system is likely. Similarly, integration of the western U.S. and Canadian rail systems will become an economic imperative. In part this will be driven by the desire of Canadian exporters to gain access to the U.S. Pacific Coast.

Privatization of CN and passage of the Canada Transportation Act will enable the restructuring of railroads on a continental basis. The majority ownership of CN shares is now foreign. The reorganization and integration of CN infrastructure on a continental basis is not only possible, it is the logical outcome of economic geography. The Canada Transportation Act removes any role for the Canadian Government to interfere with the sale or disposal of Canadian rail infrastructure, as long as action on rail lines is segmented between provinces.

The future control and ownership of hopper cars could be influenced by the forces of continentalism. Currently there is a proposal to create a farmer controlled Canadian hopper car company. We see this as something that would postpone but probably not stop the integration of hopper cars into the continental grain and transportation market. The potential returns that can be made buying and selling grain would be enhanced by a market driven hopper car system. These economic benefits will be used to buy access to hopper cars or finance a supply of hopper cars. As the railroad infrastructure is integrated on a continental basis, the integration of hopper cars will not respect national boundaries.

The legislative/regulatory framework for railroads is different in Canada than in the United States. U.S. legislation has explicit goals to ensure adequate revenues for the railways to encourage investment, maintain fair wages, and meet the demands of interstate commerce and national defense. Note the inclusion of national political and strategic objectives. Pricing protection mechanisms such as the adequate revenue principle and the lack of competitive access features for shippers contribute to U.S. railroad viability and stability. The Canada Transportation Act provides no political objectives for railroad infrastructure, nor does it provide mechanisms to ensure railroad infrastructure viability through adequate revenues. Rather, Canadian railways face competitive access features such as competitive line rates. In addition, U.S. railroads also pay lower taxes: 8.1% of revenues compared to 14.2% for Canadian railroads. The east-west Canadian railway infrastructure cannot be sustained under these conditions.

Continental Grain Market

A continental grain market is becoming a reality. On June 22, 1995 the Pools decided to take over domestic and North American marketing of non - board grains from XCAN Grain Pool Ltd. by the end of August. XCAN grains will handle mainly offshore sales. The change means that the Pools will now deal directly with North American buyers, provide prices to country elevators, finance inventories, arrange transportation and handle other logistics.

In its mission statement, the Canada - US Joint Commission on Grains clearly indicated that it will strive to transcend perceived national self-interest and to more fully realize the spirit of existing trade agreements. Its recommendations are looking for the compatibility and the harmonization of the respective systems.

The Commission recommended that Canada and the USA should continue to define the conditions for trade that provide, among other objectives, for a more open and competitive infrastructure and handling system with greater compatibility and reciprocal access for producers of both countries.

The report also indicated that "The Commission has concluded that grade and quality control differences can contribute to competitive pressures and trade problems in the North American market. Differences in regulatory regimes can create significant market distortions, particularly in certain years, due to shifts in production patterns, yields and growing conditions. The long-term goals include enhancing the compatibility of the two grading systems to avoid future difficulties and facilitate trade."

To the above end the Commission recommended that the grain inspection authorities in both countries standardize their methods and develop a common basis for the science of measurement.

In the area of infrastructure, the Commission indicated that both countries should move toward a more coordinated grain transportation and handling system. The Commission recommended that both countries pursue the long term goal of providing reciprocal access to the other's grain infrastructure.

Impacts of Restructuring

Reduced Grain Shipments

As freight rates for farmers increase due to loss of subsidy, the price of grain on the prairies will decline. Less grain will be produced, different kinds of grain will be produced, and more grain will be fed to livestock. As the cost of shipping grain to export position increases, it will become more attractive to sell into the U.S. domestic market. The eventual deregulation of the Canadian Wheat Board and the Canadian Grain Commission will facilitate this diversion. There is, however, a potential backlash by U.S. farmers to limit penetration of Canadian grain in the U.S. domestic market.

Changes in types of crops produced, the processing of grain, and sales of grain to the U.S. will change the transportation system. Trucking is much better placed to handle north south grain movements than are the railways. Canada's rail infrastructure is oriented east-west rather than north-south. For example, there is one effective rail crossing of the Saskatchewan to U.S. border, but several highway crossings. Provinces and States will begin working on projects to upgrade heavy haul cross border truck routes. Higher value crops and processed products are more time sensitive and suitable for truck transport. The supply of truck transport for grain is also increasing. Farmers buying tractor semi-trailers or B-Trains (double trailer trucks) to haul their grain are offering commercial trucking services to grain companies and farmers. The additional availability of labor and tractors to haul other commodities may further destabilize the financial viability of both trucking and rail in Western Canada.

In summary, there will be less grain to ship by rail. This predicted impact on grain production has been softened, however, by the strong world demand and high prices for grain currently. Nevertheless, grain shipments through the Canadian Pacific ports declined 31.4% in the first 21 weeks of the current crop year (coincides with elimination of WGTA) compared to the same period in 1994-95. Grain stocks were depleted late in the 94-95 crop year in order to take advantage of the WGTA subsidy that was about to expire.

Elevator/Branch Line Consolidation

Repeal of the WGTA will accelerate the consolidation of elevators and branch lines that has already been occurring for twenty years. The pace of consolidation will increase even more when the rate ceiling is lifted which will result in a greater spread of rates between branch line/main line and between single car/trainload. Restructuring will occur faster in the south than in the north because grain diversion to the U.S. will be more pronounced. Northern grain will be more captive to the Canadian railroad system. The farther north grain is produced, the more expensive it will be to truck it to the United States for sale or export. As volumes of southern grain are diverted to the U.S. system, northern grain will also have a bigger share of carrying the overhead costs of the Canadian system.

Diversion of Grain Exports to U.S. Routes

"The transportation subsidy (WGTA) sucked grain out of Canada east and west. Now it will move more naturally into the U.S."

John Miller, President, Miller Milling Co., Minneapolis, Minnesota, Telegraph Journal, June 26, 1995

Loss of the WGTA rail subsidy will make U.S. rail lines and ports an attractive alternative for Canadian exports. Even though Canadian routes currently have an advantage because of our low dollar, (i.e. the cost of paying to use U.S. routes is higher when our dollar is low) New Orleans becomes an attractive alternative for eastbound exports once the WGTA subsidy is eliminated.

A report by Western Economic Diversification Canada (WEDC) found that eastbound shipments were more at risk of diversion than westbound shipments. Currently, westbound grain rail rates in Canada, even without subsidy, are lower than westbound rates in the U.S. On the other hand, elevator charges are about one-half as expensive in the U.S. as in Canada. On balance, the cost of shipping westbound via the U.S. and Canadian routes is about the same. If the value of the Canadian dollar increases, the U.S. routes would be more competitive.[1]

Even if the costs of the U.S. and Canadian routes are currently similar westbound, grain shippers may wish to open the U.S. routes in order to introduce competition and drive down elevation charges in Canada. This is part of the unpredictable political/cultural variable that econometric models are unable to measure. Farmers are witnessing dramatic restructuring of branch lines and country elevators. Some have suggested that diverting grain to the U.S. system would deliver a political protest message to Canadian railroads and the Federal Government.

The WEDC report notes that "If the potential traffic were there, Burlington Northern has ample leeway in its existing freight rate structure for North Dakota and eastern Montana to lower rates to compete for Canadian traffic." (p.5-55). It states that capacities of elevators in the U.S. near the border are being increased, and rates from these elevators are attractive. This suggests that cross border trucking of grain to the U.S. export route is anticipated. (p.5.43,44)

Some segments of the Canadian grain industry consider that annexation to the U.S. system provides the best chances for survival. For example, Alberta Wheat Pool is already planning to protect its market share by using the U.S. system. In partnership with General Mills, Alberta Pool is building a grain handling facility in the border town of Sweatgrass, Montana. Pool will truck grain to the Montana elevator where it will be shipped by Burlington Northern deeper into the U.S.[2] The terminal will handle wheat, durum and barley destined for US mills and malt houses. Most of the grain will be trucked directly from southern Alberta. This is in accordance with the expected north-south movement of grain. Other grain companies on the Canadian prairies may be expected to secure elevator access on the U.S. side of the border.

Some Canadian farm groups are holding discussions with the port of Portland to ship grain through those American facilities. Such a move would sting the port of Vancouver. Canpotex is building a potash terminal in Portland, Oregon. The potash movement could facilitate grain movement to the U.S.

Currently, Burlington Northern is suffering congestion on its main line through Washington State. The peak time of grain shipments coincides with peak demand for other traffic. Thus, Burlington Northern may not currently have spare capacity to handle a lot of Canadian grain. The company is currently considering a proposal to re-open the abandoned Northern Pacific Stampede Pass route through Washington State in order to increase capacity. Washington State lawmakers are pushing legislation to encourage reopening of the Stampede Pass line. Inland waterways in the U.S. Pacific Northwest may also provide a conduit for Canadian grain from Alberta to the Pacific coast.

The price of shipping grain by rail to the Pacific Coast is currently higher via the United States than via Canada. This is offset, however, by the lower cost of using elevators in the U.S. Many of the elevators on the U.S. Pacific Coast are owned by the public ports authority, and leased at low cost to grain companies. Part of the cost of running grain terminals in the U.S. is paid for out of the business of buying and selling grain, which results in low prices for handling the grain. In Canada for the most part, the elevator companies are handling the grain as an agent of the Wheat Board. Since they do not make as much money buying and selling grain (the Wheat Board captures the margins for farmers), they must recover costs from handling fees. This results in higher elevator charges in Canada than in the U.S.

Currently there are some constraints on exporting Canadian grain via U.S. routes. Canada has higher grading and inspection standards, but proposals are being made to harmonize our standards with those of the U.S. as a cost-cutting measure. It was politically difficult to allow publicly owned hopper cars to be used to serve foreign ports, but the Government has announced its intention to dispose of the hopper car fleet. Currently there is a proposal to create a farmer controlled hopper car company, but there is no indication that farmers are loyal to Canadian ports. Most of the institutional constrains to using U.S. routes will unravel in the deregulation that will follow elimination of WGTA. It is a question of when rather than if.

Rail Rates

Canada's grain freight rates are a paradox. The railroads are exposed to a deregulated continental transportation market, but there is a ceiling on freight rates for grain. This situation is not sustainable. As the railways start to lose volumes of grain to U.S. routes, their financial viability will be affected, and pressure will build to deregulate the rate ceiling.

The grain rate ceiling tends to limit the difference in rates based on volume and geographic area. Deregulation of the rate ceiling will increase the spread in rates between main lines/branch lines, between north and south, and between unit trains and single car loading. Based on current contributions and rates for other bulk commodities, rates from inland grain terminals in the south could be $3-$4 lower than the current maximum. Based on current deregulated grain rates in the U.S., rates for single cars or branch lines could be $11 higher than the current maximum, particularly in the north which has less opportunity to use U.S. routes. The price of access to hopper cars during peak demand could add to these costs, depending on the method of hopper allocation chosen.

Rates will be higher for grain shipped from the northern prairies compared to the southern prairies. Rates on southern grain will be limited as farmers have the option of trucking the grain for sale in the U.S. or to load for export via U.S. elevators. The U.S. railways can compete for southern grain. Northern grain will be more captive to the Canadian system. Trucking south to the U.S. is less feasible because of longer distances. As southern grain is diverted to the U.S. system, northern grain will face a bigger share of the overhead cost of the remaining Canadian system.

Conclusion

Railways are rapidly being reorganized on a continental basis. There could be as few as two or three main North American Railroads within five years. Many short line railways will depend on these main line carriers for revenue sharing and car supply. Globalization will result in downgrading of Canadian grain grading and inspection standards. Changes to the Canadian Wheat Board marketing system and to hopper car ownership and allocation will be introduced over time. The end result will be harmonization of the Canadian and U.S. grain systems in a continental grain market. Canada's export grain system will be integrated into that of the major transnational companies.

[1]Manitoba Transport Institute, International Competitiveness of Western Canadian Transportation, prepared for Western Economic Diversification Canada, November 1994

[2]John Urquhart and Scott Kilman (The Wall Street Journal), "End of Grain Subsidy Will Reshape the Prairies," Telegraph Journal, June 26, 1995

A Short-Run Analysis of Inland U.S. Barge Rates for Grain

by James W. Dunn and Gregory L. Harnish[1]

The economic structure of the inland U.S. inland barge system is studied. A short-run model for barge rates for grain on the Mississippi-Ohio-Illinois-Missouri-Tennessee River system is estimated. Variables considered are volume of trade, supply of barges, rail rates, fuel prices, labor rates, and coal rates.

1. Introduction

The grain transportation system for the center of North America is heavily dependent on water transport on the Mississippi River and its tributaries. The presence of the water alternative takes at least 50 million of tons of grain off of the highways and railways of the United States, freeing these routes for cargo that cannot reasonably use barge transportation. The congestion this eliminates is enormous. With more open borders between the United States and Canada, the Mississippi may become a more important outlet for Canadian grain as well. Access to the Mississippi near Minneapolis provides could make Winnipeg a gateway for southbound movements rather than just east-west movements.

This paper provides background on this important transportation market and an economic model that attempts to explain the rates for barge transport.

2. Background

Mississippi River System

The Mississippi River System is particularly for important grain transportation. The Mississippi River System includes the Arkansas River, the Illinois Waterway, the Mississippi River, the Missouri River, the Ohio River, and the Tennessee River.

---

[1] Department of Agricultural Economics and Rural Sociology, Pennsylvania State University.

The volume of grain passing through locks measures the relative importance of individual waterways within the Mississippi River System (Bertels, 1992). The United States Army Corps. of Engineers collects data on the volume and type of grain passing though locks on the Mississippi River System. Table 1 shows tons of Southbound grain passing though selected locks on the Mississippi River System. Arkansas No.1 is the last lock on the Arkansas River. Illinois No. 8 is the final lock on the Illinois River. Mississippi No. 15 is located at Davenport, Iowa. Mississippi No. 25 is the final lock on the Upper Mississippi before the confluence of the Mississippi and the Illinois Rivers. Mississippi No. 26 is the last lock on the Upper Mississippi River after the confluence with the Illinois River but before the confluence of the Mississippi River and the Missouri River. Mississippi No. 27 is the final lock on the Mississippi River. The Missouri joins the Mississippi River above Mississippi No. 27. Ohio No. 52 is the last lock on the Ohio River. Ohio 52 is also the first lock south of the confluence of the Ohio River and Tennessee River.

Table 1
Southbound Grain Movements Mississippi River System Locks (1995)
(Thousands of Tons & Percent of Total)

Lock	Mile[2]	Corn	Soybeans	Wheat	Other	Total
Arkansas No. 1	10	55	375	863	10	1,302
		(4.19)	(28.80)	(66.24)	(0.77)	(100)
Illinois No. 8	80	14,254	2,817	310	43	17,424
		(81.81)	(16.17)	(1.78)	(0.25)	(100)
Mississippi No. 15	1,437	16,092	2,084	405	130	18,712
		(86.00)	(11.14)	(2.17)	(0.70)	(100)
Mississippi No. 25	1,195	20,629	3,421	595	199	24,844
		(83.04)	(13.77)	(2.39)	(0.80)	(100)
Mississippi No. 26	1,157	36,269	6,462	956	269	43,956
		(82.51)	(14.70)	(2.17)	(0.61)	(100)
Mississippi No. 27	1,139	36,622	6,724	1,240	302	44,888
		(81.59)	(14.98)	(2.76)	(0.67)	(100)
Ohio No. 52	42	1,650	778	285	5,529	8,242
		(20.02)	(9.43)	(3.46)	(67.08)	(100)

Source: National Ocean Service (1993), United States Army Corps of Engineers (1995)

[2] Miles from mouth of river.

Table 1 indicates the bulk of grain transported by barge on the Arkansas, the Illinois, the Missouri, the Ohio, and the Upper Mississippi Rivers originates on the Upper Mississippi River. In 1995, 24.8 million tons of grain passed through Mississippi No. 25, compared to 17.4 million tons passing though Illinois No. 8, 8.2 million tons passing though Ohio No. 52, and 1.3 million tons passing through Arkansas No. 1. The volume of grain originating above Mississippi No. 15 was slightly greater than the amount originating on the Illinois River. The tonnage increase between Mississippi No. 26 and Mississippi No. 27 indicates less than one million tons of grain originated on the Missouri River. The majority of barge traffic on the Missouri River occurs below Kansas City, Missouri (Fruin and Halbach, 1994). Intra-river movements of sand, gravel, and rock account for 67 roughly percent of barge tonnage on the Missouri River.

Table 1 also indicates the relative volume of different types of grain originating on the Arkansas, the Illinois, the Missouri, the Ohio, and the Upper Mississippi Rivers. Corn and soybeans are the predominant grains originating on the Illinois and Upper Mississippi Rivers. Wheat accounts for one third of grain tonnage originating on the Arkansas River. On the Ohio River, "other grain" is mainly corn and soybeans that is not classified in the statistics.

Towboats

Desai (1977, p.10) reports that towboats are typically between 117 and 160 feet long and 30 to 40 feet wide. Towboats have 7.5 to 8.5 feet drafts and 3,000 to 8,500 horsepower engines. Towboats operating on the Lower Mississippi River have relatively higher horsepower engines than those operating on the Upper Mississippi River (Woolverton, 1978, p.93). Survey results from the early-1980's indicate towboat costs represent 64 percent of variable costs for firms transporting grain on the Mississippi River System (Miller and Hill, 1986). Fuel and labor expenses are the major operating costs for towboats.

Fuel expenses represent the largest portion of towboat variable costs: 44.4 percent to 57.6 percent depending on the engine size of the towboat (Woolverton, 1978, p.99). Miller and Hill (1986) find fuel costs account for slightly over 50 percent of towboat variable costs, whether fuel costs are expressed per barge operated, per barge loaded, or per 1,000 ton-miles of transportation. Fuel costs per towboat are a function of distance, the usage rate, and fuel prices (Woolverton 1978, p.99). Towboats use No.2 Diesel Fuel. The Inland Waterways Act of 1978 instituted a fuel tax for towboats on inland waterways effective in 1980.

Between 1980 and 1995, the fuel tax increased from 4 to 20 cents per gallon

Labor costs account for 27.5 to 35.4 percent of towboat variable costs (Miller and Hill, 1986; Woolverton, 1978, p.94). A 1995 survey of towboat labor rates suggests a towboat crew earns between $1,500 and $1,800 per day.[3] Crew sizes range between eight and 12 people depending on the firm and waterway. Crew sizes appear slightly higher for towboats operating on the Lower Mississippi River than towboats on the Upper Mississippi River.

Barges

Barges used on the Mississippi River System are rectangular steel boxes with flat bottoms and shallow drafts (typically nine feet). These barges can be used only on inland waterways. Types of barges operating on the Mississippi River System include open hopper barges, covered hopper barges, and liquid cargo (tank) barges. Open hopper barges transport non-perishable cargo including coal, sand, gravel, and limestone (Wood and Johnson, 1989, p.183). Covered hopper barges typically transport perishable cargo including grain and fertilizer. Liquid cargo barges transport liquids or gases, including petroleum. Table 2 lists the number of barges by type operating on inland waterways.

Jumbo and standard are the two major classifications of open and covered hopper barges. Jumbo covered hopper barges hold 1,500 tons and measure 195 feet in length and 35 feet across. One open or covered hopper barge holds the equivalent of 15 railroad cars or 60 semi-tractor trailers. As a general rule, jumbo covered hopper barges hold 5,485 bushels of grain per foot of draft (Sorenson, 1993, p.94). Stumbo open hopper barges measure 195 feet in length and 26 feet across. Standard open hopper barges carry 1,000 tons of cargo and measure 175 feet in length and 26 feet across. Barge carriers prefer jumbo-sized barges because fewer barges per a given tow size reduces the time spent assembling and disassembling tows. Fewer barges per a given tow size also improves navigation by increasing tow stability (Case, 1968, p.86).

[3] A barge firm conducted the survey and released the results to the authors conditional upon anonymity.

Table 2
Types of Barges (1993)

Barge Type	Number in Fleet	Percent	Average Age (Years)
Standard Open Hopper Barges	848	4.0	14.3
Stumbo Open Hopper Barges	497	2.3	14.3
Jumbo Open Hopper Barges	6,237	29.4	14.7
Jumbo Covered Barges	10,694	50.4	15.8
Small Tank Barges	83	0.4	40.9
Jumbo Tank Barges	1,238	5.8	17.6
Semi-Integrated Tank Barges	1,208	5.7	20.2
Specialty Barges	427	2.0	23.0
Total	21,232	100.0	

Source: Lambert, 1994.

Case (1968, p.94) notes that jumbo covered hopper barge designs changed little since WWII. Covered hopper barges are essentially open hopper barges fitted with removable, watertight fiberglass lids. Thus, covered hopper barges are capable of transporting perishable cargo and non-perishable cargo. Barges must be cleaned between uses. For example, a covered hopper barge transporting coal must be cleaned before it may transport grain. Cleaning costs run several hundred dollars per barge.

Between 1973 and 1982 the number of hopper barges approximately doubled (Diehl and Philip, 1985). Miller and Hill (1986) find the number of covered hopper barges increased from 4,762 in 1973 to 11,550 in 1984. From the 1950's through early-1970's, annual hopper barge construction averaged fewer than 500 barges (Lambert, 1994). Constructions roughly averaged 1,000 barges annually between 1974 and 1978. From 1978 to 1981, annual construction rose dramatically, reaching over 2,500 barges in 1981. Since 1982, construction collapsed to 500 barge per year or fewer.

Because the average life of a hopper barge is 20 to 25 years, the industry expects barge carriers will retire many of the barges built in the late-1970's and early 1980's within the next 10 years. Between 1995 and the Year 2000, Hubbard (1995) estimates barge firms will retire 3,200 barges. Barge firms are expected to retire eight hundred barges per year between the Year 2000 and 2005. To meet future demand, Hubbard projects the covered hopper barge fleet needs to increase to 20,440 barges by the Year 2000. This implies barge firms will order 5,400 new barges between 1995 and the Year 2000—the fastest construction rate

since the 1970's. With the prices of a new jumbo open hopper and covered hopper barges at $235,000 and $275,000, respectively, the industry will require roughly 1.5 billion of capital to finance new barge building during the next 10 years (Hubbard, 1995).

Miller and Hill (1986) report barge variable costs account for 36 percent of variable costs incurred by barge carriers hauling grain on the Mississippi River System. Maintenance (40.9 percent) and insurance (for the barge and its contents) (53.9 percent) are the largest categories of barge variable costs (Woolverton, 1978, p.95).

Locks and Dams

The United States Army Corps of Engineers (Corps) operates, maintains, and improves inland waterways, including the provision of locks and dams to help maintain an adequate channel depth and facilitate navigation. Although the standard dimensions for locks on the Mississippi River System are 600 feet long and 110 feet wide, lock chamber dimensions vary considerably.

Lock size has important implications for the time it takes for a tow to pass through a lock. Depending on the lock size, a towboat may need to separate the tow into smaller groups of barges in order to pass through the lock. For a towboat and tow consisting of 15 barges, it takes two lockings to pass through a 600 foot lock. These additional lockings result in substantial delays for towboats and tows traveling on inland waterways.

Due to projected barge traffic increases, a preliminary study by the United States Army Corps. of Engineers estimates it is necessary to lengthen three locks on the Upper Mississippi River by the Year 2000 and 11 additional locks by 2050 (Fruhling, 1993). Estimated costs to lengthen a lock run $250-300 million.

Terminals and Coverage Area

Barge terminals typically have barge loading, unloading, and fleeting facilities[4] (Sorenson, 1993, p.97). Many of these terminals are small but some hold several million bushels. Most do not have rail connections. A terminal will attract grain directly from farms in the vicinity and indirectly via country elevators within 150 miles. Both the farmers and the country elevators will evaluate their alternatives and

[4] Barge companies use fleeting areas to increase or decrease the number of barges in a tow.

choose the option that is the most profitable. Since rail lines also run to the Gulf, shippers farther from the rivers but near a rail line will use rail when the rates are competitive. This competition should keep rail and barge rates somewhat in line.

Mississippi Gulf terminals stretch from Port Allen, Louisiana (slightly north of Baton Rouge) to Myrtle Grove, Louisiana (roughly 25 miles south of New Orleans). Storage at these terminals ranges from 2.1 to 7.7 million bushels, most of which is short-term storage. All terminals have railroad connections. Railroads serving Mississippi Gulf elevators include the Illinois Central Gulf, the Union Pacific, and the Southern Pacific. Rail car capacity at these terminals varies between 75 and 277 cars. All of these terminals maintain berthing and fleeting areas for barges. The size of fleeting areas ranges from five to 46 barges while berthing space ranges from 12 to 86 barges.

Grain terminals at the Mississippi Gulf are capable of rapidly moving massive quantities of grain. For example, the Cargill terminal at Reserve, Louisiana can unload 180,000 bushels per hour. Mississippi Gulf grain elevators are also capable of loading between 60,000 and 120,000 bushels per hour onto ocean vessels (United States Department of Agriculture, 1991). At a rate of 120,000 bushels per hour, it takes roughly 17 hours to load a Panamax size ship with 52,000 metric tons of corn (slightly over two million bushels).

Backhauls

Fertilizer is often considered a "perfect" backhaul since its Mississippi Gulf origin and upstream destinations are exactly the opposite of southbound grain movements. However, up to 80 percent of northbound covered hopper barges on the Mississippi River System travel empty (Halbach, Fruin, and Wulff, 1985). Some companies actively seek-out backhauls while others do not. Carrying cargo upstream decreases the turnaround time between grain shipments to the Mississippi Gulf. Factors contributing to decreased turnaround time include extra time to load, unload, and clean barges and additional transit time to the backhaul's origin and destination. Administration and scheduling costs are also important. Thus, backhaul rates are a function of all additional variable costs and southbound grain rates (Halbach, Fruin, and Wulff, 1985).

Coal Transportation

Since jumbo covered hopper barges are capable of transporting nonperishable commodities, grain harvest and marketing patterns

produce seasonal influences on barge rates for coal. (Gordon,1995, p. 49). Coal is the most important commodity hauled on inland waterways, accounting for roughly 30 percent of tonnage. In terms of Class I railroads, coal also accounts for 40 percent of tonnage (more than three times as much as farm products) and 22 percent of revenues. Coal shippers use barges to transport coal to power plants within the United States to export terminals. Both barge and rail transportation compete for long-haul coal transportation. However, rail and truck transportation compete to deliver coal to river terminals.

Barge Industry Structure

The barge industry is quite concentrated. The concentration varies across types of barges and cargos, but as one indicator, in 1993, the largest three firms controlled over 50 percent of jumbo covered hopper barges and the largest eight firms control 77.5 percent of jumbo covered hopper barges (Lambert, 1994).

The barge industry is also becoming increasingly vertically-integrated. In 1984, Hubbard (1995) estimated the largest five grain firms controlled 61 percent of covered hopper barges hauling grain and the largest 10 grain firms controlled 99 percent. In 1995, the largest five grain firms controlled 99 percent of covered-hopper barges hauling grain. Thus, the largest operators of jumbo covered hopper barges are also the largest purchasers of grain barge transportation.

Although grain companies control several major barge firms, the barge firm does not necessarily transport grain exclusively for the parent company. Miller and Hill (1986) find 35 percent of grain company-owned barge firms receive at least 99 percent of their grain traffic from the parent company. However, seventy-six percent of grain company-owned barge firms receive more than 50 percent of their grain traffic from their parent company.

Grain Barge Rates

Market forces determine grain barge rates. Three methods of price discovery exist for grain hauled on the Mississippi River system: privately negotiated long-term contracts, forward and spot trading on the St. Louis Merchants Exchange, and privately negotiated spot rates (for immediate shipment). Each method of price discovery is a type of formula pricing relative to a benchmark schedule of freight rates. Hubbard (1995) estimates compensatory barge rates are 175 percent of the benchmark tariff.

Long-Term Contracts

Long-term contracts specify the number of barges per week or per month a barge carrier will supply to a shipper (Woolverton, 1978, p.64). The duration of these long-term contracts runs from three to 12 months. Over 80 percent of barge rates are set by privately negotiated contracts (McGarry and Schmitz 1992, p.472). Woolverton (1978, p.65) estimates 60 to 80 percent of grain barge shipments are done under long-term contracts. Long-term barge contracts may have fuel price, user charge, and general price level escalators.

Most grain shippers anticipate their barge space needs for the upcoming year and use long-term contracts to purchase the majority of this amount of barge space. Woolverton (1978, p.65) reports barge carriers typically sell between 75 and 90 percent of their available grain barge space under long-term contract. By failing to commit their grain capacity in total, barge carriers retain flexibility to adjust to scheduling problems, river conditions, and opportunities to transport other commodities. The same is true for grain shippers. Barge carriers and grain shippers either purchase additional barge space or sell surplus barge space as needed on the spot market.

St. Louis Merchants Exchange

The St. Louis Merchants Exchange is a centralized market for southbound grain barges.[5] The Exchange holds a southbound Call Session daily at 9:30 a.m. (Central time). Exchange members bid, offer, and trade standardized spot and forward contracts for southbound grain freight on the Mississippi River System.

The standardized contract specifies transporting 50,000 bushels (1,400 to 1,500 tons) of heavy grains from a specified origin to the Mississippi Gulf. Trades may specify immediate (spot) shipment or shipment up to one year in advance. Spot trades specify a particular origin (e.g. Memphis). Future trades may specify a particular segment of a river as the origin with the buyer identifying a specific origin at a letter date. Contract prices are quoted as percent of the Waterways Freight Bureau Tariff No. 7[6] published in 1968 (Sorenson, 1993, p.112). The

---

[5] The information in this section is based on an interview with Morris Larson of the St. Louis Merchants Exchange.

[6] The Waterways Freight Bureau published barge tariffs for various commodities and represented the barge industry before the Interstate Commerce Commission. Although grain was exempt from Interstate

buyer transfers ninety-five percent of the freight bill with the bill of lading with the remainder to follow. Barge companies use brokers to participate in the Daily Call Session. However, the largest barge company (ACBL) does not participate in the Daily Call Session.

Spot Rates

Market participants establish spot barge rates at the St. Louis Merchants Exchange and through private negotiation. Rates are quoted as a percent of the Waterways Freight Bureau Tariff No. 7. The tariff rate for St. Louis is $3.99 per ton. Barge freight rates for grain at St. Louis are extremely volatile. Sorenson (1993, p.108) attributes volatile barge rates to relatively inelastic short-run supply and fluctuations in short-run demand.

Resellers are the most active participants in the spot grain barge market (Woolverton, 1978, p.66). Resellers are primarily firms with more barge space contracted than grain to ship. Brokers also participate in the spot market. Brokers earn a fixed fee per barge or percent of the freight rate to match grain shippers with barge carriers. Although barge carriers drop-out of the spot markets if the freight rates drops below 70 to 80 percent of the benchmark tariff, the same is not necessarily true of resellers. All costs become fixed once a firm contracts for barge shipping space. The alternative to reselling barges is to incur demurrage costs and default penalties (Woolverton, 1978, p.67). A grain shipper may resell grain barge space to another grain shipper.

Barge rates for grain display a pattern of seasonality consistent with the fall harvests. Farmers in corn-surplus states market 60 percent of corn sold off farm during October and November (Sorenson, 1993, p.99). Farmers move 85 to 90 percent of off-farm wheat sales during wheat harvest.

3. The model

This research project relies on spatial equilibrium theory (Samuelson, 1952) and studies by Woolverton (1978) and Dunn and Gianoulades (1985). As such, it views the transportation demand as being derived from the trade flows between spatially separated markets. The

---

Commerce Commission jurisdiction, Tariff No. 7 historically served as a framework for rate negotiation. After the Interstate Commerce Commission canceled Tariff No. 7 (1974 supplement) in 1976, the barge industry continued to use the rate schedule as a basis to negotiate grain barge rates.

determinants of these trade flows are therefore the determinants of the transportation demand. Several possible transportation demand curves are shown in Figure 1.

Short-Run Supply Function

Dunn and Gianoulades (1985) report that economists traditionally view the supply of transportation services as perfectly elastic. This implies transportation rates are independent of the level of trade. However, they propose ocean freight rates are also dependent on the level of trade because the short-run supply of transportation may resemble the curve depicted in Figure 1.

The short-run supply curve shown in Figure 1 has three regions. At Q_1, the supply curve has a slightly positive slope because excess shipping capacity exists. This implies shifts in demand induce minimal changes in freight rates. At Q_2, marginal costs increase moderately because the most efficient equipment is already hauling grain. Thus, demand shifts cause moderate changes in freight rates. At Q_3, marginal costs increase rapidly since the fleet is at full capacity. Here, slight changes in demand result in highly volatile freight rates.

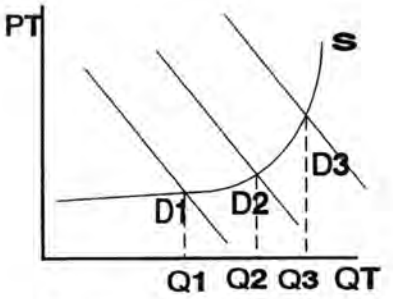


Figure 1. The short-run barge transportation market

A similar argument can be made for the short-run supply of southbound grain barges. Like ocean vessels, the fleet of covered hopper barges is fixed in short-run. Depending on the backlog of new covered hopper barge orders, the time elapsed between ordering a new covered hopper barge and taking delivery of the barge ranges between several months to several years. Thus, the supply of southbound grain barges may resemble the supply of ocean shipping described by Dunn

and Gianoulades (1985). This implies southbound grain barge rates may be dependent on the volume of grain transported via barges.

Like ocean vessels, the short-run supply of southbound grain barges is positively related to Southbound grain barge rates and negatively related to barge rates of alternative cargoes. Coal is an example of an alternative cargo for grain barges. Because the lids on covered hopper barges are removable, grain barges are capable of transporting both perishable and nonperishable commodities, including coal.

Dunn and Gianoulades (1985) also include variable costs in their short-run supply curve. Recall that both fuel and labor costs are the major components of barge transportation variable cost. Thus, both fuel and labor costs should be negatively related to the supply of southbound grain barges. Finally, the short-run supply of southbound grain barges is negatively related to the number of covered hopper barges already busy transporting grain.

Econometric Model

The rate for hauling grain by barge on the Mississippi River system will be represented by the rate for hauling grain from the St. Louis area of the Mississippi to the Gulf. Rates from other locations, whether on the Mississippi or one of its tributaries, will be related to this rate. Research in progress is studying broader rates to see their resemblance to these rates. A reduced form will be estimated, where the rate is assumed to be the result of quantity supplied equalling quantity demanded for barge transportation for a particular week. This rate can then be explained by the exogenous determinants of the supply and demand for transportation. The resulting equation is

$$R_t = f(G_{t-2}, RR_t, RC_t, F_t, L_t, B_{t-2}, D_t),$$

where R_t is the rate, G_t is the quantity of heavy grains exported from the Gulf, RR_t is the railroad freight index for grain transported in the United States, RC_t is the barge rate for hauling coal, F_t is the fuel price, L_t is the labor rate index, B_t is the number of covered-hopper barges loaded with grain at the Gulf, and D_t is a set of dummy variables for months and crop years, all for each week t during the period 1988 to 1992. All of the variables except the dummy variables should be positively related to rates. Finally, the model includes a dummy variable representing the effects of the 1988 Drought.

4. Data

The period of study (1988-1992) is a function of data availability and quality. The barge rate is the average spot barge rate for transporting heavy grains to the Mississippi Gulf measured in dollars per ton. The exports are the quantity of heavy grains exported from the Mississippi Gulf measured by export inspections. (Grain Marketing News) The railroad freight rate index for grain transported in the United States is a component of the Producer Price Index (PPI) estimated by the United States Department of Commerce. Weekly estimates of the rail freight index were obtained by interpolation. Coal rates are the annual contract rate to transport export coal from the Big Sandy River to Davant, Louisiana measured in dollars per ton. The fuel price is the spot price of No. 2 heating oil per gallon plus the appropriate fuel tax rate. The labor cost is an index of the daily labor cost per towboat.

5. Results

The results of the estimation of the model using ordinary least squares are shown in Table 3. The estimated coefficients are all of the correct signs and are generally significant. The price flexibilities for the regular variables are 0.21 for grain export inspections, 0.24 for fuel price, 4.63 for rail rates, 45.94 for labor rates, 0.08 for number of loaded barges at the Gulf, and 1.58 for coal rates. This means that a 1% increase in export inspections would be related to a 0.21% increase in barge rates. The quantity variable, exports has a significant, yet small effect. This implies that during this period the barge shipping market was in the elastic portion of the supply curve. Efforts at nonlinear estimation in order to discover curvature in the supply of barge transport curve, such as is seen in Figure 1, showed no curvature. Apparently the period in question was never a period of scarce transport.

The labor flexibility is extremely high. This variable may be correlated with an omitted variable and be representing the impact of both. An increase in labor rates should have a dramatically different effect than an increase in fuel costs, with an upper bound for the flexibility of perhaps 1.0.

The rates for competing modes and uses are important in determining spot barge rates for grain. These flexibilities suggest that barge rates are more affected by external market forces than costs of operating the barges (if the huge labor fleibility is discounted).

Implications

The estimated model is similar to models for other routes (not reported) on the river in structure, size of coefficients, and price flexibilities. As a result, the Mississippi River system barge transportation market for grain can probably be viewed as a single interrelated market. The rates were primarily a function of cost of operation variables and the cost of alternative modes of transport (rail) or alternative uses of the equipment (coal). The quantity to be shipped was less of a factor. This may be reflective of the market power of the barge owners and the role that the big grain companies play in the market. When the major customers also control much of the fleet, a great sensitivity of rates to volume to be shipped would be less likely.

6. Concluding Comments

This preliminary study shows that the barge transportation market behaves like any economic market. When costs of supplying the service go up, the price rises. When the rates for alternative uses for the equipment rise, the price for the primary use rises also. When the quantity to be shipped rises so do the rates. And finally, when the price of a competing mode rises, the rates of barges rise.

Table 3: Coefficient Estimates for Barge Rates (1988-92)

Variable	coefficient	t statistic
Intercept	-280.302	-3.977
Exports	0.981	2.358
Fuel Price	0.018	1.503
Rail Rates	0.224	2.793
Labor rates	2.352	3.520
Barges at Gulf	0.309	0.921
Coal rates	1.248	2.645
1987 Crop year	20.964	3.706
1988 Crop year	13.227	3.578
1989 Crop year	6.807	4.245
1991 Crop year	-5.869	-2.955
1992 Crop year	-11.819	-3.083
January	2.006	2.518
February	1.458	2.153
March	1.342	2.370
April	0.781	1.727
May	-0.253	-0.669
July	-1.015	-2.567
August	-1.333	-2.597
September	6.104	4.287
October	6.464	5.231
November	3.603	3.383
December	2.246	2.382
1988 Drought	2.816	3.674
R^2	0.61	

7. References

Bertels, Paul J. 1991 Annual Barge Summary. Agricultural Marketing Service, United States Department of Agricultural, 1992.

Case, Leland S. "Estimation of Production and Cost Functions for Inland Waterway Transportation." Ph.D. 1968.

Desai, Kiran. "Scheduling of Tows Through Locks on an Inland Waterway: A Simulation Study." Ph.D. 1977.

Diehl, Neil N. and Craig E. Philip. "Critical Issues facing the U.S. Barge Industry in the 1980's." Transportation Journal. Fall, 1985: 5-10.

Dunn, James W. and Andrea Gianoulades. "A Short Run Analysis of Ocean Freight Rates for U.S. Grain Exports." Proceedings of the Transportation Research Forum. 26 (1985): 573-578.

Fruhling, Larry. "The Mississippi's Murky Future." The Des Moines Register. May 10, 1993

Fruin, Jerry and Dan Halbach. Barge Traffic on the Illinois and Missouri Rivers: 1972-1992. Staff Paper P94-20 University of Minnesota, Dept. of Agricultural and Applied Economics, Oct., 1994.

Gordon, Richard. Export Coal Transportation. Unpublished paper. 1995.

Halbach, Dan, Jerry Fruin, and Scott Wulff. 1984 Barge Rates for Upper Mississippi River Commodities. Staff Paper P85-13 University of Minnesota, Dept. of Agricultural and Applied Economics, April, 1985.

Hubbard, Peter E. Midland Enterprises Inc., 1995.

Lambert, Jack. Barge Fleet Profile of Inland River Equipment for the Mississippi River System and Connecting Waterways. St. Paul: Phalanx Publishing Co., 1994.

McGarry, Michael J., and Andrew Schmitz, ed. The World Grain Trade: Grain Marketing, Institutions, and Policies. San Francisco: Westview Press, 1992.

Miller, Patricia H. and Lowell D. Hill. Organization and Structure of the Barge Industry Transporting Grain and Oilseeds. AE4613, University of Illinois at Urbana-Champaign, May 1986.

Samuelson, Paul A. "Spatial Price Equilibrium and Linear Programming." *American Economic Review*. 42 (1952): 283-303.

Sorenson, Orlo L. "Grain Transportation." Grain Marketing. Ed. Gail L. Cramer and Eric J. Wailes. 2nd ed. San Francisco: Westview Press, 1993.

United States Army Corps of Engineers. Lock Performance Management Database. World Wide Web, 1995.

United States Department of Agriculture. Federal Grain Inspection Service. Export Elevator Directory. Washington DC: July, 1991.

Wood, Donald F., and James C. Johnson. Contemporary Transportation. New York: Macmillan Publishing Co., 1989.

Woolverton, Michael W. "Analysis of Grain Barging Rates and Costs on the Mississippi River." Ph.D. 1978.

THE USE OF HAND-HELD COMPUTERS IN ROADSIDE SURVEYS

Greg Little and Doug Smith
Ministry of Transportation, Ontario[1]

1.0 INTRODUCTION

The Ministry of Transportation, Ontario (MTO) has conducted roadside surveys of trucking for many years. MTO recently undertook a review of the Commercial Vehicle Survey Program to ensure that the program would continue to meet the needs of the users of the survey. One of the results of the review was the use of hand-held, pen-based computers in the collection of survey data. These computers were used in the province-wide Commercial Vehicle Survey which was undertaken in the summer of 1995. Software was designed to utilize state of the art technology to enter data directly into hand-held computers using pen-based technology.

The paper will review the capabilities of the technology used in the 1995 Commercial Vehicle Survey, provide details on the operational experience and evaluate the technology to assess to what extent benefits were realized.

2.0 BACKGROUND

The Commercial Vehicle Survey (CVS) is a valuable source of data on the characteristics of trucking activity. Information is collected on a wide range of characteristics including: driver, carrier, vehicle, commodity, and trip. The objective of the CVS is to obtain a profile of inter-city trucking activity on Ontario highways. The information is particularly valuable as it is the only source of information which provides the relationships between these characteristics. In effect, the CVS provides information on who is moving what, where, when, and how. This information is valuable for a broad range of purposes within the ministry as well as for external purposes. Uses of the information include:

- Freight Policy Analysis e.g. harmonization of commercial vehicle weights and dimensions
- Planning Studies e.g. developing regional transportation plans and truck traffic patterns
- Economic Forecasting e.g. developing trends in freight movement

---

[1] The views expressed are those of the authors and not necessarily those of the ministry.

- Strategic Decision Support e.g. developing and maintaining a Strategic Transportation Network for Capital Investment Planning

The Commercial Vehicle Survey requires a significant investment in both resources and time. With a budget of almost $600,000, the 1995 CVS will take close to 2 years from the beginning of planning to the final report. A survey staff of 23 was in the field for more than four months.

3. CVS PROGRAM REVIEW

The Ministry of Transportation has undertaken a province-wide Commercial Vehicle Survey approximately every five years. The last survey prior to the 1995 CVS was in 1988. After the last province-wide survey there were a number of concerns about the survey methodology, data expansion and aggregation, statistical soundness of the data, and the value of the data for specific types of analysis. While the survey addressed most of the needs for planning purposes, the survey did not meet all of the requirements in the area of freight policy development. The surveys were not being conducted on a sufficiently frequent basis to be useful for trend analysis.

As a result of these concerns, a project was initiated to develop a statistically sound methodology to improve data reliability and to enhance the value of the commercial vehicle survey. The activity involved:

- Determining Data Needs
- Improving the Survey Methodology
- Designing the Questionnaire
- Conducting Pilot Surveys to Test the Redesigned Survey Methodology and the Direct Data Entry System.

A) Data Needs

Based on consultation with stakeholders, the data needs were categorized into the following areas of interest: carrier, driver, commodity, origin/destination of the trips and goods, vehicle characteristics. These data had to be tied to a time and location. The areas of questioning were designed to answer the following questions:
Who is moving what, where, and when, using what equipment?

- What are the relationships among these factors?
- How and why have these factors, patterns and impacts changed over time?
- On what routes do these freight movements occur and how does the road network handle the flows?
- What are the annual volumes of commodities being transported?
- What are the safety, social, economic and environmental impacts of these movements?

B) Survey Methodology Development

A review and evaluation of the survey methodology identified the framework upon which future surveys should be built:

- select the strategic survey locations;
- determine control data and survey duration requirements;
- randomize selection of survey sites, days and hours;
- improve quality control through intensive interview training;
- develop methodologies to expand data, to combine data from several sites and to reduce bias (trip length, driver and vehicle when combining data);
- ensure high quality truck count/classification data are collected;
- develop a provincial network and zone system;
- investigate direct data entry to improve data quality and ensure data is available soon after the interviewing is complete.

C) Questionnaire Development

A component of the determination of data needs exercise was to review and evaluate the questions used on previous surveys. During the evaluation some questions were eliminated, some added and other questions reworded or restructured. Some suggestions made in the survey methodology study were incorporated into the questionnaire. Questions related to bias elimination were incorporated into the questionnaire. The intent was to keep the questionnaire length to less than 10 minutes.

The most difficult elements of the questioning were related to the origin and destinations. There are many origin destination possibilities: commodity, driver, truck, the load and a last stop and next stop. Complicating the situation are different types of trips such as line hauls and peddle runs.

It was decided that three origin destinations would be collected:

- Commodity Origin Destination: This was the origin and destination of the commodity based on examination of waybills or information provided by the driver
- Trip Origin Destination: The origin of the trip is the location where the driver started with the truck in its loaded/empty condition when interviewed. For example, in the case of a loaded truck going from New York to Winnipeg, and the driver being interviewed after picking up the truck in Buffalo to go to Winnipeg, the trip origin would be Buffalo and the destination is Winnipeg.
- Last Stop/Next Stop: These origin destinations are defined as locations where the driver last stopped and the expected next stop. The purpose of the

stop can be for service, sleep, customs inspection or loading/unloading.

The questionnaire was reviewed for consistency of layout, style and coding conventions.

D) Pilot Survey

The purpose of the pilot survey was to implement, test and refine the survey methodology and refine the questionnaire.

Results of the 1993 CVS pilot survey motivated in part the application of Direct Data Entry (DDE). For the 8,377 interviews collected using the pencil and paper technique, it cost $4.42 per interview and took six months for the interviews to be entered electronically, edited, cleaned and coded. Identifiable error rates by the interviewers and data entry were unacceptably high (in the order of 10% - 15%) due to:

- control information (location, date, time) being poorly reported;
- interviewers missing questions;
- poor spelling;
- the interviewer writing in ineligible codes;
- the interviewer accepting inconsistent answers;
- the data entry person not being able to interpret the interviewer's handwriting;
- typing errors by data entry personnel.

4. 0 DIRECT DATA ENTRY (DDE)

Computer assisted interviewing has been used for survey data collection for more than ten years. The technique has been used primarily for telephone surveys where the interviewer follows a questionnaire on the computer screen and inputs the responses directly into the computer. This technique is continuing to replace the technique of filling in a paper questionnaire and having the data transformed into electronic format at a later date.

MTO has had considerable experience with direct data entry type systems. In 1984, a Share-A-Ride Matching program was designed for direct entry of information while the applicant is on the telephone. The 1991, Transportation Tomorrow Survey employed direct data entry using a custom designed system.

The expected benefits associated with the direct data entry technique were lower data costs and better quality data. These benefits are achieved because the technique provides:

- on line entry of data, therefore eliminating the costs to have the written responses entered electronically;
- built-in range and logic checks that allow the interviewer to correct data or clarify a response with the respondent;
- the ability to use external data bases to reduce data entry;
- the system can automatically conduct coding internally;
- the ability to monitor interviewers to improve response quality and productivity.

These benefits were expected without any appreciable increase in interview time.

The Commercial Vehicle Survey was identified as a survey in which DDE techniques could be used to advantage.

The Direct Data Entry system developed for the Commercial Vehicle Survey has taken the DDE into a new generation of systems. The DDE technique has been enhanced for field conditions where keying in data is considered impractical. Basically, the system simulates the paper and pencil technique. The system employs pen-base technology that uses a pen to touch the screen to input data rather than using a keyboard. Data are entered by clicking boxes or by writing answers on the screen. The system recognizes handwriting and inputs the characters as if they were typed.

The significant challenge was to take the DDE concept into field conditions where environmental factors can play havoc with equipment and where there is continuous repositioning of interviewers. To achieve this, the computers had to be portable (battery powered), light weight, durable, have a full size screen for visibility, operate pen-based software, have significant computing speed and storage space and be reliable in a hostile environment (cold, heat, full sun, night, dust, fog and light rain).

The hardware selected for the survey was the Fujitsu - Stylistic 500 with the following features:

- light weight (2 kg); interviewers had no difficulty carrying the computer for extended periods;
- two lithium ion batteries; in survey conditions the batteries lasted about two hours (this varied based on lighting conditions and weather), a second set of batteries is recharged while the first set is in use;
- protective casing to minimize operational impact - computers proved to be very durable in the normally rough use in field surveying environment ;
- the full sized screen (12 cm by 16 cm) was essential for readability and it allowed several questions to be on each screen;

- the computer operated with pen based software;
- 486 operating system with 105 megabytes of hard disk storage (the software and data bases require about 50 megabytes).

Features of the Direct Data Entry software for the Commercial Vehicle Survey include:

- control data (date, time) are entered by the computer;
- the system does not proceed unless all questions on a screen have been answered;
- system guides the interviewer through the questionnaire (system automatically skips over questions that are not applicable);
- all answers to questions are coded in a consistent manner;
- over 100 logic/range checks and error messages are incorporated into the system;
- interviews can be reviewed and edited on site (allows for quality control by supervisor);
- built-in external data bases are accessed through pop-up lists;
- a built-in data base of more than 216,000 place names in North America is used to identify correct origins and destinations;
- input of correctly spelt place names allows for automatic geocoding and the calculation of travel distance;
- other databases built into the system include survey station locations, carrier names, Canadian placename geocodes, province and state names and border crossing locations;
- coding of place names to a county level zone system is done automatically;
- totalling of weights is done automatically;
- the system contains a help section, a sketch pad and a calculator;
- interviewer can flip to the French version of the questionnaire;
- the system calculates interview time and interviewer productivity;
- data from the system are output in a DBF format for direct entry into RBASE;
- backup saves data in three locations.

The system has been developed in modules so that shortened versions of the questionnaire can be developed at relatively low cost.

The efficiency of existing hand writing recognition software is very dependent on the hand writing style and neatness of the person doing the writing, especially when alpha and numeric data can be entered into the field. When the recognition system does not decipher the hand writing correctly, delays are created because interviewers have to rewrite the response. These delays would break up the flow of the interview and frustrate the interviewer. The DDE system for the CVS was designed to

minimize the amount of writing required by having the interviewer select the answer from a list whenever possible. For questions with a small number of possible answers, a list of answers was displayed on the screen and the interviewer would touch a box beside the correct response. However, when there were more than about ten possible answers a pop-up list based on a built-in data base was displayed. Since all names in the list could not be displayed on the screen at the same time, an alphabetical search program was developed to aid in finding the correct response. For example, when the interviewer was inputting a place name, the interviewer would write in the first three letters of the answer. The system would display in alphabetical order the names starting with those letters. If the name did not appear in the displayed list, the fourth letter was added and so on until the name appears (if it is in the list). The interviewers would then select the name from the list. In other words, the interviewer could input a place name such as Peterborough by writing in only three or four letters, and then selecting the name from the pop-up list, rather than writing in all twelve letters. The interviewer had the option of completing the entry of the name and inputting the name as entered (this option was to be used if the answer was not in the list.

When the information entered by hand is selected for input it will not necessarily be standardized to the data base, i.e. a place name such as Toronto could be entered by hand and input rather than selecting Toronto from the data base. In addition, the internal zone coding was not completed. The standardized spelling also facilitates the use of distance software for calculation of trip distances.

As part of the development of the software, five data bases were assembled to reduce the data entry time of names and places, to have standardized spelling of names to reduce data cleaning (place names are used by the distance calculator program), also coding of the locations to a zone system could be done on-line within the system. The five data bases assembled are: province/state, placename, survey locations, carrier name, and border crossings.

The province/state data base contains all the province and state names and an associated zone system. The data base is used for coding of licence plate jurisdiction and accessing the place names within a jurisdiction.

The placename database contains a total of 216,000 place names in Canada, United States and Mexico and the names of 229 other countries. Also, included are the associated province/state/country zone numbers. For the 38,735 Canadian locations, their latitude and longitude location are included. Places in Ontario are associated with a county level zone system. Locations of places with the same name are distinguished by a descriptor of their location. This data base was assembled from Statistics Canada data and place names in the distance calculator program.

The survey locations' data base contained all the locations where the survey was to be conducted. It contains the station number, highway location and a locational descriptor.

The carrier names data base is derived from the Ministry of Transportation Commercial Vehicle Operators Registrant (CVOR) file. It contains the carrier name of every holder of a CVOR in Ontario (79,800 carrier names) and therefore is a reasonably complete record of vehicle operators in Ontario. Some carriers operating infrequently in Ontario may not be included in the list.

The border crossing data base contained information on all international and all Ontario inter-provincial border crossings. Each border crossing was represented in various ways (by bridge name, border cities, province/state name and road number); this list attempted to cover any description the driver may provide to the interviewer. In total, there were 118 descriptions for the 35 border locations.

5.0 EFFECTIVENESS OF HAND-HELD COMPUTERS

As noted above, one of the significant innovations in the 1995 Commercial Vehicle Survey was the use of hand-held pen based computers to enter the survey data on site. The computer program for the survey questionnaire was developed by a consultant[2] for the Ministry of Transportation.

5.1 Planned Benefits

The use of this technology was undertaken to provide the following benefits:

- increased accuracy of data
- increased completeness of data
- improved control of survey operations
- immediate access to interview and observation data in electronic format
- increased cost efficiency

5.2 Actual Benefits

An analysis of the 1995 CVS results was undertaken to determine the extent to which the expected benefits were realized. The 1993 Commercial Vehicle Survey Pilot was analyzed as a comparison to the 1995 CVS. This 1993 CVS was appropriate for this purpose as a similar methodology was used for the data collection. The hand-held computers were not used to collect data in 1993; data was

[2] Dartek Systems Inc, Markham Ontario

collected manually using paper forms and data entered in bulk after the fact. In addition, survey staff, dedicated to the survey, were utilized in both cases. Similar staff training was conducted for both surveys.

There were, however, some significant differences between the two surveys which have been accounted for where possible. These differences are:

- duration
  - 1993 CVS August
  - 1995 CVS May to September
- number of surveys
  - 1993 CVS 8,377
  - 1995 CVS 31,866
- the survey instrument used in 1995 contained fewer questions/observations than in 1993
  - 1993 CVS 70 questions/observations
  - 1995 CVS 38 questions/observations

In almost all cases, the survey questions used in 1995 were used in 1993.

Records from the 1993 Commercial Vehicle Survey were examined to identify problems. A sample of approximately 3% of total records were examined in detail. These records were examined for completeness and accuracy. Data entry errors were identified by comparing the hard copy survey questionnaire and the electronic record.

The 1995 Commercial Vehicle Survey data base was examined to determine the completeness of records. A comparison of errors for common data elements for the two surveys was undertaken. In both cases, data accuracy could only be determined for obvious error. Generally, it is only possible to check the accuracy of the driver's responses at the time of collection.

5.2.1 Increased Accuracy

The use of the hand-held computers increases the accuracy of the data collected by three means:

- elimination of data entry errors
- utilization of built-in data bases to standardize information
- built-in range and logic checks

The utilization of these measures ensured consistent structure and coding practices throughout the survey.

Elimination of Data Entry Errors

Without the use of on-site data entry, paper surveys must be completed by hand and the information entered into electronic form at a later date. This inevitably leads to error regardless of how much control and care is taken. In the 1993 Commercial Vehicle Survey Pilot data entry was undertaken by a consultant.

A total of 86 data elements were examined for the 1993 survey. Of the data elements examined, 36 (49%) had data entry errors (a difference between what was on the paper questionnaire and in the data base). The average error rate for those data element with errors was 2.4%. These errors did not occur in 1995 as the data entry was done on site and not transferred from hard copy.

Utilization of Built-in Data Bases

As noted above, the computer system utilized five built-in data bases.

1. Survey Sites

The Survey Site data base allowed the surveyor to select the survey location at which whey were operating. The correct site had been a significant problem in previous surveys. The survey location was significantly improved in 1995 compared to the 1993 survey (1993 17.1% error, 1995 5.6% error). However, it remains a significant problem as the surveyor was required to select the correct site at the beginning of each shift or when the computer was shut off. In the 1995 CVS data base, all station names entered were from the data base. However, problems appeared to stem from interviewer carelessness, or not being aware of the correct identification of the station.

The traffic direction was also entered in conjunction with the station number. However, this was not correlated with the station and any direction could be entered without reference to the station. This resulted in a significant number of incorrect entries. Bi-directional stations were a particular problem as the direction of traffic could legitimately vary and was incorrectly entered in numerous occasions.

2. Carrier Name

As noted above, the Carrier Name data base is a copy of the CVOR file of the ministry. As such, it should be a reasonably complete record of vehicle operators in Ontario. It would not have information on some carriers e.g. out-of-province operators who are infrequent operators in Ontario.

The data element COMPNAME in the 1995 CVS data was tested from a sample of

2,666 names representing 8.6% of entries with the following results:

records examined	2666
names selected from data base	1811 (67.9%)
names entered manually (not selected):	
names in data base	365 (13.7%)
names not in data base	490 (18.4%)

The interviewers were instructed to either enter the company name by hand or select the name from the data base, as it was stand alone information and would not affect other data. In total, the data base used was comprehensive, containing approximately 85% of carriers surveyed.

3. Place Name

The place name data base contained over 216,000 place names in the United States, Canada, and Mexico. These were used in seven different areas of the survey:

- commodity origin
- commodity destination
- trip origin
- trip destination
- last stop
- next stop
- vehicle base

A significant number of place names occur twice in the place name data base in various jurisdictions, e.g. Mississauga. In all cases, each place name has a zone number. In addition, for the Ontario place names which were duplicated, a description of the city is also provided to assist in selection. Despite this, a preliminary investigation indicates that the selection of the wrong place name was a significant problem.

The Commodity Origin data element was present in 21,828 records. A sample representing approximately 60% these records were examined in detail with the following results:

records examined	13839
records selected from data base	13050 (94.3%)
records entered manually:	
in data base	602 (4.4%)
not in data base	181 (1.3%)

This shows that the data base was very comprehensive, containing almost all locations encountered in the survey. The data base was utilized very successfully by the surveyors, with only 4% of records having a place name manually entered which existed in the data base. In addition, at the beginning of the survey there was confusion about the necessity of selecting the place names from the data base which was not clarified until the third week. This would account for many occurrences.

4. Border Crossing

The Border Crossing data base contained information on all international and interprovincial border crossings in Ontario. Each border crossing was represented various ways, i.e. by Ontario and other jurisdiction perspective to cover any description the driver may provide to the surveyor. In total, there were 118 descriptions of all border crossings. The Border Crossing data was examined with the following results:

records examined	15222
records selected from data base	15094 (99.2%)
records entered manually:	
in data base	32 (0.2%)
not in data base	96 (0.6%)

5. Province/State Name

The province or state was required to be entered whenever a place name was to be selected. The jurisdiction selected restricted the place name to those in that jurisdiction. The program did not allow the jurisdiction to be entered free hand, it had to be selected from the list. The two digit code for the jurisdiction was automatically entered into the data base.

Built-In Range and Logic Checks

The computer program has edit checks built into the data. In total there are 30 built-in logic checks. All fields where a selection was to be made ensured there were no selections out of range. The logic checks included:

- number of axles in a group, e.g. axle group 1 must be less than 3 axles,
- maximum weight for axle groups, e.g. axle group 1 must be less than 15000 kgs
- number of stops for peddle runs
- requiring border crossings to be identified if a trip origin and/or destination is outside Ontario

The checks were designed to either disallow data outside the specified limits or allow it by exception. These checks all performed as designed, and data was not allowed where prohibited or by exception.

There is significant latitude possible in the types of checks which can be developed. However, care must be taken not to make them too restrictive or problems can arise in the field. For example, the program originally would not allow a border crossing if the trip origin and destination was in Ontario. However, it was soon evident that a few trips crossed a border even though the trip origin and destination were in

Ontario e.g. Sarnia to Sault Ste. Marie. The program was revised to allow this.

5.2.2 Increased Completeness of Data

The computer program was designed not to allow a question to be missed by the surveyor. This was a significant problem in previous surveys (whether inadvertent or deliberate). A sample of data variables from the 1993 and 1995 survey data bases were examined for missing data.

1993 CVS:	Average rate of missing information per variable	3.7%.
1995 CVS:	Average rate of missing information per variable	0.3%

This shows a significant reduction in both the number of data elements where data was missed, and the percentage of records with missed data. However, there should not have been missing data in any of the surveys, as the computer did not allow questions not to be answered. These cases appear to have been occurrences in which the computer was not working properly, e.g. was shut down during the survey for some reason.

5.2.3 Improved Control of Survey Operations

Entering the survey information during the interview process provides an opportunity to review the data in a timely manner and correct problems or errors which are occurring. In the 1995 CVS, this was done to a limited extent due to logistic and systems problems arising from assembling the initial data base. Approximately three weeks into the survey process, all survey staff were visited on site and advised of specific problems which were occurring. However, additional follow up was not possible.

In future surveys, this feature should be given more attention to improve data quality.

5.2.4 Improved Timeliness of Data

The direct data entry allowed the 1995 CVS data base to be assembled within one month of the end of the survey (September 30th). Cleaning and editing of the data could begin at this point. The data base should be available by the end of July 1996. A significant improvement over the 1988 CVS experience of almost three years. In addition, the assembled data incorporated the use of standardized data bases and coding of the data. This provided an initial data base which was already a partially cleaned data base, which was enhanced by coding with a county level zone system.

The fast availability of the raw data does enable some preliminary analysis to be conducted. To date, the 1995 CVS data has been used for preliminary analysis on a number of projects and will continue to be utilized as it is edited and cleaned. This would not be possible without data entry on site.

5.2.5 French Version of Questionnaire

The computer system facilitated the administration of the survey questions in the French language. A French version of the questionnaire was built into the computer and could be accessed by the interviewer as required. This was used nominally during the survey with 471 (1.5%) of questionnaires being conducted in French.

5.2.6 Increase Cost Efficiency

Direct Data Entry eliminated the need to enter survey data after the fact. This is a direct saving of the survey cost. There was a requirement to purchase the hand-held computers. However, this was a capital cost and the equipment is available for use in future surveys. In addition, the computers were rented to Transport Canada which recouped some of the capital cost. A comparison is outlined below:

1995 CVS	
Saving from utilizing computers	
- cost savings of not requiring data entry (31,800 @ $3.00 estimated)	$95,400
Cost of utilizing computer	
- computer costs	
annual cost of computers (30% amortization)	$21,000
less computer rental revenue	$ 9,300
Total Computer Costs	$11,700
Contract staff for data assembly and cleaning (part time)	$15,000
Total cost	$26,700

The utilization of computers in the 1995 CVS saved a total of $68,700 (comprised of the estimated saving from the elimination of data entry, $95,400 less expenses incurred of $26,700).

3.3. Productivity

It was not expected that the use of the computers would increase the productivity of the interviewers, i.e. it would not lessen the time required to complete a questionnaire. There was a possibility it might take longer to complete a questionnaire by computer, compared to completing a paper document.

It is not possible to make a direct comparison with the 1993 Pilot Survey, as the 1995 survey had fewer questions. In addition, the number of surveyors used at one time differed, with three surveyors per shift being the standard in 1993 and one per shift normally being used in 1995. However, examining the average number of questionnaires completed per shift, it would appear that the productivity did not suffer as a result of the use of computers. The average number of completed questionnaires per shift per surveyor is outlined below by month (each shift was 8 hours long):

1995 CVS	Avg/shift
May	27.6
June	33.8
July	33.0
August	34.5

1993 CVS	Avg/shift
August	21.1

From the 1995 data, it was also apparent that the efficiency of surveyors increased after a learning curve, as the average for May was significantly lower than other months. If we look at the data for the first four weeks of each survey, we see a somewhat different picture for each survey.

1995 CVS	Avg/shift
week 1	23.3
week 2	26.5
week 3	24.8
week 4	33.2

1993 CVS	Avg/shift
week 1	21.0
week 2	21.3
week 3	24.5
week 4	19.0

The 1993 survey has a discernable increase in week three with a decrease in week four. The decrease in week four may be due to a slacking off in the final stage of the survey. This may indicate that the optimal efficiency was reached in week three.

For the 1995 CVS, optimum efficiency was not reached until the fourth week. This does not mean that it took three weeks of constant computer use to reach optimum efficiency. There were numerous disruptions in the schedule during the first month due to problems coordinating survey staff with compliance staff. Therefore there was some staff down time which lengthened the time required to reach optimal efficiency. In addition, there were, as noted above, some problems with the computers locking up during the first two weeks of the survey. However, even with this, approximately two weeks was required for the staff to become familiar with the use of the computers, as well as the survey questionnaire and operation.

6. SUMMARY/CONCLUSIONS

The use of hand-held computers for direct data entry in the 1995 Commercial Vehicle Survey was a successful application of technology in an operational environment. The following conclusions can be reached from the operational

experience:

1. The computers fulfilled expectations in all areas:
   - accuracy of data was increased by:
     - elimination of data entry errors
     - utilization of standardized data bases, built-in edit/logic checks, and control data
   - completeness of data improved significantly
   - control of survey operations was utilized to a limited extent - however, the system provides opportunity for greater utilization of this function
   - the timeliness of the data was significantly increased, i.e. a data base was compiled quickly - the finalized data base should be available within 10 months after the end of the survey
   - the computers were cost effective - eliminating data entry and coding costs produced a savings of $68,000

2. There is a significant amount of time required for survey staff to become familiar with the operation of the hand-held computers and survey procedures. To minimize the impact of this, training should include as much on-site familiarization as possible. In addition, during the first few weeks of the survey, extra attention should be given to quality control. This could be a combination of on-site supervision and checking data quality from computer files which would be checked on a daily basis.

3. Extra attention should be paid to ensuring the correct station information is recorded at the time of the survey. The utilization of the hand-held computers increased the accuracy of this data to a great extent. However, there remained a significant problem of incorrect site information. This data is critical to having good quality data. Therefore, there should be further investigation into ways and means of improving the accuracy of this data. This should include automated solutions as well as increased attention to quality control.

4. The use of the hand-held computers did not significantly change the productivity of the survey operation. There may have been a somewhat longer familiarization period required due to the use of the computers.

CONJOINT ANALYSIS OF AIR CARRIER LOYALTY

Edward R. Bruning, Transport Institute, University of Manitoba
Barry E. Prentice, Transport Institute, University of Manitoba
Patty-Jo Bellamy, Faculty of Management, University of Manitoba

I. Introduction

In the spring of 1996, the governments of Canada and the United States liberalized the two countries' cross border aviation policies. The *Open Skies* agreement allows a greater number of North American cities to be served by air carriers from bordering jurisdictions. Greater international competition between carriers - Canadian and American - will encourage more travel, support expanding trade relations and, indirectly, stimulate innovation and market responsiveness for the airlines.

Sometimes it is difficult to forecast whether significant policy changes are opportunities, or threats. Will Canadian travellers remain loyal to Canadian carriers in a deregulated transborder air travel market? This question focuses directly on the extent of country-carrier loyalty - a vital element of competitive advantage that carriers seek to possess. Put another way, will Canadian travellers remain "nationally loyal" in the face of changing international competitors' product offerings, and product attribute trade-offs? To the extent that Canadian carriers are able to retain "loyal" customers, their ability to compete effectively with larger and more financially able American carriers will be greatly enhanced. If, on the other hand, loyalty is readily traded off for lower fares, better services, and so forth, then the ability of Canadian airlines to maintain a competitive position in local markets may be diminished.

The marketing literature is replete with examples of firms who have been able to stave off competitors not because they have better prices or superior products, but because customers like them, are used to them, and remain loyal when making choices. Thus, the impact of the *Open Skies* agreement depends on the ability of Canadian carriers to 1) capture new travellers in new markets, and 2) retain established customers in markets they presently serve.

The empirical portion of the paper makes the point that 'national loyalty' is a significant factor influencing consumer choice in international air carrier. Based on relationships espoused in the theoretical literature in social psychology, we argue that one's social self-identity influences the extent of 'felt nationalism' which, in turn, is linked to the strength of national identity and, ultimately, economic nationalism--that is, the preference for local versus foreign made goods and services which impacts upon choice behaviour. Our empirical model relies on the accumulated scholarly research in psychology and marketing to argue for the legitimacy of these linkages.

In section two we present a review of the literature concerning our hypothesized relationships and an empirical model to test these relationships. Section three develops the logic of the empirical model, the sample used in testing the model, and the study design in carrying out the empirical research. We discuss the findings in section four. In the final section of the paper we address the implications of our findings with respect to airline firms, government policy makers, and researchers.

II. Conceptual Development

Self Identity Theory

Tajfel (1981) and Turner (1987) have presented compelling arguments and empirical evidence suggesting that individuals operate within various levels of identity as the situation or context in which they function dictates. This line of enquiry has been explored for several decades and has been subsumed under the rubric of Social Identity Theory. Using experimental settings, researchers have acknowledged that groups are formed based on very few similarities and, furthermore, with extended contact the degree of identification among members within the group becomes more pronounced. Thus, we can discuss personal identity, identity with family, ethnic group, religious group, all the way to identification with mankind and the universe. Each level of identity ushers forth a unique set of norms, behavioral patterns, and expectations for the individual to follow.

National identity is the level of identity that we are interested in assessing. We hypothesize that individuals with tendencies to identify strongly with others at the nation level also possess strong inclinations to make economic decisions that are in the 'national' interest as opposed to basing them on strict individualistic or utilitarian motives. Individuals who tend to identify themselves strongly at one of several levels of social identity (nation in our case) tend to view their choices and actions as members of a larger entity.

Thus, decisions on economic matters are dependent upon the implications to individual well-being in the strict utilitarian sense, and the impact that the decision has upon other members of the relevant social group with whom the individual identifies. This point has found empirical support in experiments dealing with minority group behaviour (Ethier and Deaus 1994), ethnocentrism (Bruning, Lockshin and Lantz 1993), and in consumer behaviour dealing with reference group influence on consumption choices (Bearden, Netemeyer, and Teel 1989).

Country of Origin

An area of study more closely aligned with our research question is referred to as Country-of-Origin (COO) research. COO emanates from scholars in marketing and international business and has grown significantly since the early 1970s. Briefly, COO focuses on identifying and describing specific cues that consumers respond to in making consumption decisions in situations where local and foreign products compete. Specifically, researchers question whether, say, a Canadian would prefer a Canadian-made product relative to a foreign-made product, an American is biased toward U.S.-made products, a Mexican is influenced positively by "Hecho en Mexico" markings, etc. COO has been found to operate in several ways. Consumers employ the COO cue as an additional variable in forming product evaluations (Hong and Wyer 1989). Second, the COO cue establishes a 'halo' effect that alters consumer's attention to and evaluations of product attributes (Han 1989). Third, the COO is conceptualized as a form of country stereotyping that consumers employ when other product-specific information is not available (Bilkey and Ness 1982; Hong 1987). The strength of the evidence appears to be with the latter two interpretations (Bruning, 1995).

Consumer Ethnocentrism

Consumer ethnocentrism has been found to be an important factor in international consumption behaviour (Shimp and Sharma (1987). In short, consumer ethnocentrism is a notion that posits that a consumer's patriotic emotions have a significant effect on attitudes and purchase intentions. Consumer ethnocentrism "gives the individual a sense of identity, feeling of belongingness, and...an understanding of what purchase behaviour is acceptable or unacceptable to the ingroup (Shimp and Sharma 1987, p. 280). A relevant manifestation of an ethnocentric consumer's tendencies is a propensity to buy home country goods at a higher price relative to foreign goods of equal quality (Hooley, Shipley, and Krieger 1988). This phenomenon was identified by Heslop and Papadopoulos (1992) who

suggested that domestic preference was partly a function of the strength (or weakness) of the nationalism of the people involved. Ethnocentrism has been found to influence quality evaluations (Shimp and Sharma 1987), price (Wall and Heslop 1988), and purchase intention (Han 1988).

In summary, the traditional COO research focuses on identifying the role performed by the country cue in product evaluation and choice. Researchers studying COO seek to explain the existence of a preference for own-country-made goods and services, but shy away from asking the more fundamental question of what factors and within what process do consumers use the COO cue in forming preferences? Research in the area of consumer ethnocentrism has focused on a person-specific psychological trait that appears to directly impact economic choice behaviour. In this study, we are integrating social self identity, COO, and consumer ethnocentrism research in developing the logic for our model of international airline choice. Our argument posits that a preference for local versus foreign products or services is a function of one's strength of national identity that is dependent upon the influence of social identity in defining one's self. The test carried out on the international air carrier market is an attempt to confirm our theoretical views.

Research Hypotheses

From the literature focusing on COO, consumer ethnocentrism, and social self identity several testable hypotheses emerge for the study:

H1: Social Self-Identity of airline travellers, moderated by age and gender, is positively related to their degree of Nationalism.

H2: Air travellers degree of Nationalism is positively correlated to the degree of Consumer Ethnocentrism.

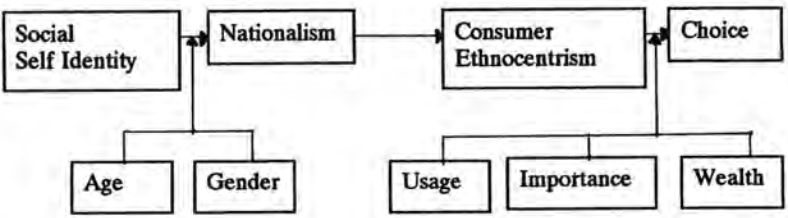
H3: Air Travellers degree of Consumer Ethnocentrism is positively related to their preference for Own-Country Carrier.

III. Methodology

In the conceptual model depicted in Figure 1, social self identity is linked to nationalism, however, both age and gender moderate the relationship. Nationalism, in turn is antecedent to consumer ethnocentrism that is tied to purchase intentions and choice behaviour. Three factors are included as moderating influences in the third relationship: historical usage levels, the importance or salience of product or service attributes, and a person's

wealth position which is critical in determining choice in an economic context.

Figure 1
A Model of Own-Country Air Carrier Preference



Research Instruments

Conjoint analysis is employed in the empirical portion of the study. Widely used in marketing research, the technique represents a consumer side approach to evaluating product attributes, individual characteristics, and situational factors on preferences and choices. The approach in this study is to derive partworth estimates of the several attributes and their levels. We then employ country of carrier partworth estimates as a dependent variable explained by measures of consumer ethnocentrism and several moderating factors. In turn, consumer ethnocentrism is linked with nationalism, and nationalism is explained by social self identity and two moderators.

There are two instruments used in securing data for the study. As already mentioned, one is a conjoint exercise where subjects are presented with multiple attribute bundles and asked to rate preferences for each of the bundles. The focus of the attribute choices is an hypothetical trip between two points in North America. The second instrument is a questionnaire that subjects complete concurrently with the conjoint exercise. The questionnaire collects information on personal demographic information and measures attitudes related to national identity, social self-identity, and consumer ethnocentrism. The development and implementation of the two instruments are discussed below.

Conjoint Experiment. The instrument for the conjoint exercise involves twenty-seven scenarios with various trip attribute combinations and a rating form for subjects to indicate trip bundle preferences based on a ten-point

preference scale. Attributes selected for inclusion in the experiment are suggested by previous research (Bruning 1994; Bruning and Hu 1984; and Green and Wind 1975). Eight trip attributes are selected for use in the experiment: price, in-flight service, number of stops, departure time, on-time performance, country of carrier, type of aircraft, and whether air miles are offered on the flight. The country of carrier attribute is our proxy for the relative importance of the country of origin (COO) factor in the air carrier selection process. Levels of each of the attributes are identified in Table 1.

Table 1: Conjoint Experimental Attributes and Attribute Levels

	Attribute Levels		
Attribute	Low	Moderate	High
Price	$560	$685	$779
In-flight Service	Low	Medium	High
Number of Stops	2 Stops	1 Stop	Non-Stop
Departure Time	Morning	Evening	Noon
On-time Performance (% on time)	50%	75%	95%
Country of Carrier	Mexico	US	Canada
Aircraft Type	707	727	DC-10
Air Miles Program	No		Yes

Three price levels are selected based on a 1500 mile flight between Canada and the United States. Pre-tests of the price range confirm that subjects identify $560 as the low price, $685 the moderate level price, and $779 the high price. In-flight services are defined as low, moderate, and high based on the quantity and quality of service descriptions presented in the instructions to the conjoint exercise. Low in-flight service is defined as the scenario of poor selection of magazines, no newspapers, no meals, too-few attendants for quick service, poor music quality, noisy aircraft, and inhospitable staff. Moderate service is defined as the scenario of at least one interesting magazine, no newspaper, cold sandwich and dessert, satisfactory speed of service, reasonable music quality, aircraft not too noisy, and congenial staff. The high in-flight service is characterized as having a good selection of newspapers and magazines, a hot meal, quick service, clear music and a movie, quiet aircraft, and excellent staff. All other attributes are self-explanatory. Pre-tests of sample subjects indicate that the levels of each of the attributes are appropriate for the study.

Attribute level combinations are based on a fractional factorial design yielding 27 combinations (Cochran and Cox 1957). Orthogonal arrays are used to determine the appropriate combinations necessary for effectively estimating main effects. Individual ratings based on a ten-point preference scale serve as the dependent variable while the various attribute levels are effect coded (1, 0, -1) independent variables representing the seven travel specific factors and the one country of origin factor. OLS regression is employed to estimate individual as well as aggregate partworth coefficients associated with each attribute and level.

Survey Questionnaire. A questionnaire is employed to solicit usage, attitudinal, and demographic information from respondents in order to capture social self identity, nationalistic, and ethnocentrism measurements and link them to the country of origin measure. The questionnaire comprises three sections. Section one captures information relative to the respondents frequency of travel, self-reported knowledge of air travel, assessments of air travel attributes, and air travel usage levels. The second section of the questionnaire includes four questions related to the extent of respondents' social self identity (Chronbach alpha = .70), eight questions related to national affiliation (Chronbach alpha = .74), and sixteen questions concerning consumer ethnocentrism (Chronbach alpha = .94). All scales exhibit acceptable levels of internal consistency. The final section of the questionnaire includes items concerning respondents' age, income, gender, and participation in air mileage programs. Data were collected at a major international airport in Western Canada. Respondents participating in the study represent all western Provinces (Manitoba, Saskatchewan, Alberta, and British Columbia), the central provinces of Ontario and Quebec, and the eastern province of Nova Scotia). A convenience sample design was employed to identify and collect information from the sample units. A total of 253 respondents took part in the conjoint exercise and follow-up survey during a three-week period in 1995.

IV. Results

Characteristics of the sample units and key variables employed in the analysis are reported in Table 2. Average age of the respondent falls in the 41-50 range with 15% frequenting the youngest age group and 19% in the 61 and over categories. Approximately 41 percent of the sample are males and 59 percent females. Interestingly, the sample units are equally split between travelling for pleasure and business. Self-reported levels of knowledge about international air travel is rated slightly above the mid-point of the seven-point scale. The typical sample respondent travels

approximately 3 to 4 times per year, although the percent of international air travel is considerably lower at less than 1 trip per year.

In terms of the importance ratings of the various air travel features, Table 2 indicates that price is the most highly rated attribute, followed by safety, on-time performance, and arrival and departure times. Several other travel features that did not rate as highly are the importance of travel agents, in-flight services, nationality of air carrier, and availability of air miles. In fact, it is worthwhile pointing out that availability of air mile programs is the least important attribute among the nine rated in the survey.

Table 2: Descriptive Statistics for Selected Profiling Variables

Traveller Attribute	Central Tendency	Importance Ranking
Age	41 - 60 years	
Gender		
Male	41%	
Female	59%	
Trip Purpose		
Business	51%	
Pleasure	49%	
Air Travel Knowledge	4.9 (7-point scale)	
Average Annual Trips		
Domestic	3	
International	1	
Travel Attribute		
Price	6.26	1
Safety	6.19	2
On-time performance	6.11	3
Departure Time	6.07	4
Arrival Time	5.93	5
Role of Travel Agent	5.22	6
In-flight services	4.93	7
Nationality of Carrier	4.56	8
Air mile Program	4.48	9

Conjoint Results

Table 3 reports flight attributes and partworth coefficients associated with each attribute level. In contrast to the attribute rating exercise reported in Table 2, the two most important attributes resulting from the conjoint exercise are 1) Number of Stops and 2) Country of Carrier. Price was considered the third most important factor followed by On-Time Performance. In-Flight Service, Departure Time, Aircraft Type and the availability of an Air Miles program are considered relatively low in importance based on the results of the conjoint exercise.

Test of Model Hypotheses

The conceptual model presented in Figure 1 generates several hypotheses that we assess with the data. In order to conduct the tests, we rely on the results from the conjoint analysis to construct a measure for the choice variable in the model. We calculate partworth coefficients for the price attribute for each sample respondent using OLS regression. The partworth coefficient for the own-country level of the country of carrier attribute (in our case Canada) is employed as a proxy measure for carrier choice preference. Partial correlation coefficients linking Social Self-Identity to Nationalism ($r=.347$, $p<.01$), Nationalism to Consumer Ethnocentrism ($r=.384$, $p<.01$), and Consumer Ethnocentrism to Choice ($r=.500$, $p<.01$) are positive and statistically significant. As indicated in the model, the relationship between Social Self-Identity and Nationalism is moderated by age and gender while usage level, individual wealth, and flight attribute importance moderate the relationship between Consumer Ethnocentrism and Choice.

Table 3: Full-Sample Partworth Utilities

Attribute	Partworth Utilities	Attribute Range	Percent Importance
PRICE		1.23	.19
$560	.79		
$684	.44		
$779	-1.23		
IN-FLIGHT SERVICE		.55	.09
Low	.55		
Medium	-.19		
High	-.36		
NUMBER OF STOPS		1.40	.22
2 Stops	-1.40		
1 Stop	.82		
Non-Stop	.58		
DEPARTURE TIMES		.36	.03
Morning	.13		
Evening	.23		
Noon	- .36		
ON-TIME PERFORMANCE		1.01	.16
50%	-1.01		
75%	.15		
95%	.86		
COUNTRY OF CARRIER		1.34	.21
Mexico	-1.34		
United States	.37		
Canada	.97		
AIRCRAFT TYPE		.71	.09
707	-.71		
727	.13		
DC-10	.58		
AIR MILES PROGRAM		.01	.01
No	.00		
Yes	.01		

V. Discussion and Implications

The empirical results of our analyses support the contention that country of carrier impacts consumer choice preferences. The relatively strong effect size of the partial correlation coefficients conforms with results reported in

COO research (Peterson and Jolibert 1995). Unlike most COO research, our findings go beyond indicating the presence of a country effect. The results of this study identity that the effect is sizeable and significant, but they also indicate that the effect is linked to person-specific traits that generate a propensity for individuals to prefer local versus foreign-provided goods or services.

While researchers exploring consumer ethnocentrism acknowledge the importance of the person-specific trait that biases assessment of quality and choice, they have not linked the ethnocentrism construct to antecedents that may perform a critical role in explaining the formation of attitudes, beliefs, or consumption behaviour. Rather, their contribution has been to operate outside the COO framework in conducting their enquiry. Our results suggest that the two approaches with their respective research questions may be integrated the results of which provide a richer perspective of the process consumers use in arriving at consumption decisions. We find that the strength of social self identity is a critical feature in understanding consumer ethnocentrism through an individual's identity with nation. Consumer ethnocentrism, therefore, is operative largely for individuals predisposed to identifying themselves with a defined national group. Those not predisposed to forming a sense of self identity at a macro-national level will likely not reflect strong nationalism attitudes, nor strong ethnocentric tendencies in consumption situations. Consequently, they will not indicate a strong preference for country of origin as a cue.

Our findings reach beyond the generalizations of previous COO research which posits that country serves as a separate cue affective in the information processing functions that result in choice behaviour. COO is a cue, and it is employed as a stereotype whenever more specific information is not known; however, it is more than a piece of information which individuals respond to. Depending on each person's "linkage" with nation, COO will play either an important or immaterial role. The point is that an emotion is evoked whenever a fundamental value or identity is challenged. For individuals who have a strong linkage between nation and self-identity, COO will prove to a strong factor in international consumption choice. On the other hand, weak self-identity-nation linkages will result in a much less significant role for COO.

The theoretical generalisations outlined in the previous paragraph have significant implications for air travel competition between Canada and the United States. Will Canadians remain loyal to Canadian carriers in a competitive North American air travel market? The answer is it depends

on the strength of the affiliation between individual consumers' self identity and the notion of Canada as a nation. Travellers possessing strong bonds between nation and self will tend to value own-country carriers above foreign-based carriers even when other flight attributes favour the foreign carrier. The results of the conjoint analysis confirm this finding. Furthermore, our analysis indicates that the degree to which loyalty for own-country carrier will prevail also depends on individual's age, gender, usage level, wealth, and involvement in the flight situation.

While our research has developed a clearer perspective as to the important linkages within the COO framework, it is clear that airlines need to understand the role of national loyalty, and the extent to which it favours them relative to foreign competitors. Based on this knowledge, they can develop programs that support the formation of nationalistic attitudes favouring own-country versus foreign-based carriers. In the case of Canadian air carriers, it is probably worthwhile to consider joint advertising programs that emphasize Canada and the idea of supporting the Canadian economy rather than simply supporting a particular airline as a distinct entity. Intense advertising competition that denigrates one carrier to the benefit of the other only serves to diminish a sense of national unity and, therefore, national affiliation among international air travellers. Furthermore, government programs that are designed to promote the interests of air carriers, tourism, and other economic sectors may improve their effectiveness by promoting the concept of a unified air travel program that includes joint participation by both Canadian air carriers.

Recently, the issue of merging the two Canadian carriers became a very strong policy issue. Our research and the logic of the model we tested indicates that, in open international competition, loyal customers will be more forgiving to local firms when price or other travel attributes are not on par with competitor's. An important question: would consumers find it easier to relate to a single international carrier rather than two as we find at the present moment? Does inter-carrier competition between Air Canada and Canadian Airlines International diminish the impact of national loyalty because the competition tends to be along regional lines? Future enquiry into COO and loyalty in aviation markets would be useful in sorting out these issues.

The emergence of strategic alliances and partial ownership between Canadian and American airlines presents a further topic for investigation. Does the public view code-sharing between airlines as desirable from a COO perspective? Should airlines emphasize the international "partnership", or downplay the commercial linkage? Further examination

of the COO concept to determine whether a "NAFTA" identity might rival a "national" identity could prove to be very useful.

Academic researchers can add to the understanding of loyalty and local airline competition. Issues that the present research has not addressed relate to appropriate definitions of the major constructs employed in modelling COO, definitions as to best ways of measuring these constructs, and the appropriate designs for sampling and analyzing measures of the constructs. Shimp and Sharma's consumer ethnocentrism (CET) scale is a reliable instrument for depicting consumer ethnocentrism; however, CET may not represent the true essence of consumer ethnocentrism. That is to say their conceptualization of consumer ethnocentrism may not be a valid representation of the phenomenon that impacts choice. More research is required to understand these conceptual and measurement issues.

References

Bearden, W.O.; Netemeyer, R.G.; and Teel, J.E. (1989). "Measurement of Consumer Susceptibility to Social Influence," Journal of Consumer Research, 15:473-481.

Bilkey, W.J. and Ness, E. (1982). "Country of Origin Effect on Product Evaluation," Journal of International Business Studies, 13:89-99.

Bruning, E.R. (1995). "The Role of Consumer Ethnocentrism in the Choice of International Air Carrier," Journal of the Transportation Research Forum, 34:1-12.

Bruning, E.R. and Hu, M.Y. (1984). "Misperceptions of Airline Services and the Degree of Prior Experience," Logistics and Transportation Review, 20:213-222.

Bruning, E.R.; Lockshin, L.S.; and Lantz, G.E. (1993). "A Conjoint Analysis of Factors Affecting Intentions of Canadian Consumers to Shop in U.S. Retail Centres," Proceedings, Administrative Science Association of Canada, Lake Louise, Alberta,

Ethier, K.A. and Deaux, K. (1994). "Negotiating Social Identity When Contexts Change: Maintaining Identification and Responding to Threat," Journal of Personality and Social Psychology, 67:243-251.

Green, P.E. and Wind, Y. (1975). "A New Way to Measure Consumers' Judgements," Harvard Business Review, July-August: 107-117.

Han, C.M. (1989). "Country Image: Halo or Summary Construct?" Journal of Marketing Research, 26:222-229.

Han, C.M. (1988). "The Role of Consumer Patriotism in the Choice of Domestic Versus Foreign Products," Journal of Advertising Research, 28:25-32.

Heslop, L.A. and Papadopoulos, N. (1992). "But Who Knows Where or When? Reflections on the Images", in Product Country Images, N. Papadopoulos and L. Heslop, eds. USA: Hawthorn Press.

Hong, S.T. and Wyer, R.S. (1989). "Effects of Country-of-Origin and Product Attribute Information on Product "Evaluation: An Information Processing Perspective," Journal of Consumer Research, 16:175-187.

Hooley, G.; Shipley, D. and Krieger, N. (1988). "A Method for Modelling Consumer Perceptions of Country of Origin," International Marketing Review, 5:67-76.

Shimp, T. and Sharma, S. (1987). "Consumer Ethnocentrism: Construction and Validation of the CETSCALE," Journal of Marketing Research, 24:280-289.

Tajfel, H. (1981). Human Groups and Social Categories. Cambridge, England: Cambridge University Press.

Turner, J.C. (1987). Rediscovering the Social Group: A Self-Categorization Theory. Oxford, England: Basil Blackwell.

Acknowledgement

This research on consumer ethnocentrism in the airline industry has been supported by a grant from the Social Sciences and Humanities Research Council.

APPORTIONING THE RISKS IN INTERMODAL TRANSPORT: IMPACTS OF PROPOSED REVISIONS TO THE U.S. CARRIAGE OF GOODS BY SEA ACT

Hugh M. Kindred*
Dalhousie Law School
Halifax, NS B3H 4H9

Introduction

Door-to-door delivery as now regularly supplied by seamless intermodal service providers is a great convenience of modern transportation. But when it fails and a casualty occurs, difficulties can be encountered in apportioning liability. Typically the shipper will have been issued with a combined transport bill of lading or similar multimodal document that allows it to hold the contracting carrier or operator responsible for the loss. But not infrequently it is more expedient for the shipper to look to the underlying carriers and cargo handlers for compensation. Perhaps the contracting carrier is a one-ship company that has escaped the jurisdiction, or a non-vessel owning multimodal operator that has no assets at all. Maybe the shipper wants to try and overreach the limitations on claims and liabilities in the agreement with the contracting operator. In all these cases, the shipper may pursue the performing carriers by road, sea, air or rail as well as the terminal operators, stevedores, warehousing companies and freight forwarders, including all their employees, agents and independent contractors.

Each of these service providers, however, wants to protect itself from this risk of unexpected liability. Each typically has a contractual arrangement that regulates its terms and conditions of service, and thus controls its exposure to liability. If a shipper may claim directly against the subcontracted service providers, it will disturb the equilibrium of intermodal relationships that make door-to-door delivery possible, and subject particular carriers and cargo handlers to risks they can neither foresee or control.

Attempts have been made both contractually and judicially to prevent abuse of these situations. Himalaya and circular indemnity clauses are two common approaches used to protect the intermodal service providers. But shippers may also expect that those parties who actually handle their goods must exercise

reasonable care and, if they do not, must answer for any resulting losses. Thus the conditions exist for confrontations between service providers, who wish to limit their risks of liability, and shippers, who want full accountability.

The proliferation of intermodal services has so increased the number of occasions when this tension arises that, as argued at the CTRF last year,[1] they need to be directly addressed. The central problem is that a seamless multimodal movement from one business premises to another is regulated by a series of segmented legal regimes peculiar to each mode or stage of transit. The obvious solution is to develop one, uniform legal regime for all modes of transport. Such a proposal has recently, though somewhat indirectly, been brought forward by the US Maritime Law Association (US MLA) in the course of promoting the revision of the US Carriage of Goods by Sea Act (US COGSA).[2] This initiative is important, to Canada in particular, in at least three distinct ways. First, it would amend and update American marine transport law so that it would be more in line with more modern international standards, including Canada's. Second, it would extend US COGSA to all the other stages of an intermodal movement and thus produce a single multimodal liability regime. Third, it would be bound to affect North American traffic which constitutes such a large part of Canadian trade and is so important in its export dependent economy.

This paper will explore the impacts of the proposed revisions to US COGSA along each of these lines of enquiry. Emphasis will be given to the effects on intermodal transport and on Canadian traffic, but first, in order to comprehend them, a general description of the changes to US maritime law will be provided.

Proposed Revisions to U.S. COGSA

US COGSA[3] was enacted in 1936 and applies the original international maritime legal regime known as the Hague Rules. These rules are clearly out-dated and many countries, including Canada, have moved beyond them, but there has been disagreement about how extensive the changes should be. Some countries, mostly OECD members, have implemented the Hague/Visby Rules while others, mostly developing countries, have gone beyond amendment of the Hague Rules and adopted the Hamburg Rules. Canada, in 1993, enacted the Hague/Visby Rules for the present and the Hamburg Rules for some time in the future when the Minister of Transport sees fit to substitute them.[4]

American opinions are apparently marked by the same uncertainty. The US MLA Report affirms that agreement has now been reached that COGSA is outdated, but disagreement remains about how to update it.[5] The middle way

proposed by the Report seems to have three objectives. The principal purpose of the reforms is to move US marine transport law from the Hague Rules towards the Hague/Visby Rules and in some respects beyond them, in the direction of the Hamburg Rules. The new COGSA is also intended to be extended in application so as to govern all stages of an intermodal movement that includes a sea leg. In addition, it would be enlarged to regulate the risks and liabilities of all the carriers and cargo handlers throughout a door-to-door delivery by the same set of rules.

The latter enlargements in the scope of application of US COGSA would be achieved by simple definitional changes. The period of application would be extended beyond the ship's tackle between loading and discharging ports (the tackle-to-tackle rule) to the whole time the cargo is in transit from the moment of its initial receipt by a carrier until its final delivery.[6] This period might only be from port-to-port if the contract of carriage is simply for a sea movement, but it might equally extend inland from door-to-door if the agreement with the shipper is for complete (multimodal) delivery. This interpretation is buttressed by the new definition of "carrier" to include both contracting and performing carriers.[7] The contracting carrier is the party who makes the agreement for carriage with the shipper. This might be the sea carrier but it might equally be a land carrier or a non-vessel/vehicle owning multimodal operator, such as a freight forwarder that is prepared to accept the responsibilities of procuring the whole carriage. A performing carrier actually moves or handles the cargo and thus may include "ocean carriers, inland carriers, stevedores, terminal operators, consolidators, packers, warehousemen, and their servants, agents, contractors and sub-contractors." In other words, everyone is a performing carrier who plays a part in the execution of an intermodal movement. Finally, the current documentary straitjacket that limits COGSA to contracts of carriage covered by bills of lading would be broken open. COGSA would be extended to all contracts for the carriage of goods "either by sea or partially by sea and partially by one or more other modes of transportation" documented in any printed or electronic form.[8] Thus, ocean bills of lading, sea waybills, combined transport documents[9] and EDI[10] would all be sufficient evidence of a contract of carriage to which the US COGSA might apply. Thus by a deft reinterpretation of carriers under a contract of carriage of goods at least partially by sea, the marine regime of US COGSA would potentially become applicable to the full extent of a complete intermodal movement.

On the substantive matter of risk allocation during carriage, the proposed reforms of COGSA would move US maritime law towards the Hague/Visby Rules in two substantial ways. First, the limits of a carrier's liability would be raised from $500 US per package to 666.67 SDR[11] or about $970 US per package, or 2 SDR (roughly $2.90 US) per kilo, whichever is the higher.[12]

These are the limits of ocean carrier liability under the Hague/Visby Rules and as enacted in Canada. Second, the time limit of one year in which to bring suit for a claim would be continued, but an opportunity for three months thereafter would be opened to a defendant carrier to bring an action for contribution or indemnity against another defaulting party to the carriage.[13]

But the application of the limits of liability could be interrupted in several different circumstances. First, the limits may be set aside where the defendant's conduct was intended to cause the loss incurred or was recklessly careless that cargo loss would probably result.[14] This principle of unlimited liability is found in the Hague/Visby Rules,[15] but not so the rider that such conduct must be "within the privity or knowledge" of the defendant carrier.[16] It is probably a retrogressive step to protect the employer-carrier from the full vicarious liability for the deliberate or reckless conduct of its employees.

Second, the carrier's right to limit its liability would be extended to loss in the event of a deviation by the ship. Traditionally deviation has been regarded as such a serious default as to oust the protection of the Rules and thus expose the carrier to unlimited liability whatever the circumstances. By including deviation within the new COGSA regime the amendments go beyond the practice under the Hague/Visby Rules, but then subjects the question of the extent of liability to the test of deliberate or reckless misconduct found within those Rules.

Third, the limits of liability are regarded as a minimum level of compensation[17] because the Hague/Visby Rules prohibit any lessening of responsibility.[18] But they do not prevent a carrier from increasing its liability,[19] and the revisions to US COGSA would confirm this approach by expressly permitting a cargo claimant to take advantage of a defaulting performing carrier's contract or tariff if it should happen to afford a higher level of compensation than the COGSA regime governing the carriage agreement with the contracting carrier.[20] However, the COGSA amendments would also permit a lessening of liability in one situation, namely in the presence of a service contract to which the particular contracting carrier and shipper happen to be parties.[21] Such possible lowering of the limits of liability has never been entertained under the Hague or the Hague/Visby Rules. In short, though the proposed revisions to US COGSA invoke the limits of liability under the Hague/Visby Rules, they also complicate their application by many American made exceptions and variations.

In a number of ways the revision of US COGSA would also go beyond the Hague/Visby Rules. Undoubtedly the most important change to the substantive regime of risk allocation between carriers and shippers would be the abolition of the defence of navigational error. No longer would the defendant carrier be

able to excuse itself from liability by the unseemly argument that its crew members were negligent in the navigation and management of the ship. Henceforth the carrier would be vicariously liable for the acts of its employees, as in all other businesses, but the cargo claimant would have the burden of proving their conduct was negligent.[22] This may be no light task since all the evidence would be in the hands of the carrier and the impugned master and crew.

Other amendments would include deck carriage within US COGSA's regime,[23] create a contributory fault rule[24] so as to permit the apportionment of liability when a loss results from more than one cause, and make a small change to the exemption from liability for loss due to fire so as to allow the contracting carrier, as well as the performing carrier, to assert this defence.[25] The revisions of US COGSA would also clarify the carrier's position in giving a receipt for goods received from a shipper which it is unable to check or tally, such as cargo stuffed and sealed by the shipper in a container,[26] and also to restrain carriers' choices of foreign judicial fora or arbitration sites in protection of US cargo claimants' procedural rights.[27]

Intermodal Impact of a Revised US COGSA

The intention that the new US COGSA should regulate more than sea carriage, indeed that it shall reach out to govern every part of an intermodal movement touching the United States, is evident from the enacting clause at the beginning of the draft Act prepared by the US MLA Committee. It states that US COGSA will apply "to every contract that includes the carriage of goods by sea covering transportation to or from the United States."[28] Thus the involvement of sea carriage would be a requisite for the invocation of US COGSA but not a limit; it will apply to all parts of every such contract covering transportation to or from the United States. This wide scope of intermodal application is, even so, a reduced version in the Revised Report of the original proposal. The enacting clause of the Act as first drafted captured all contracts for transportation "to, from, or through" the United States.

This assertion of multimodal reach is reinforced in the enacting clause by the statement that US COGSA shall have the force of law in any action against a carrier or a ship, without regard to its legal basis, and shall constitute the complete regime of remedies available to the cargo claimant. Recalling the expansive definition of "carrier" to include the contracting carrier and all varieties of performing carriers and cargo handlers, it is evident that US COGSA is intended to become the exclusive regime for the uniform regulation of sea based intermodal carriage.

American truckers and railway companies that would be caught up in COGSA's regime showed enough resistance to the proposal to secure their partial exclusion in the Revised Report. Believing they would be better protected under their own service arrangements, the definition of "carrier" was altered to exclude interstate trucking and rail carriers when fulfilling the responsibilities of a performing carrier.[29] But if they should choose to act as a contracting carrier for an intermodal movement that includes a sea leg, they would still attract the regulation of the new US COGSA. Apart from this limitation the revised US COGSA would be a comprehensive regime that would apply to the contracting carrier for the whole period of the intermodal movement, and to each performing carrier during the time it has custody and control of the cargo or is otherwise performing the contracted carriage.[30]

In general the impact of the revised US COGSA displays the right sort of solution to the substantive problems of risk allocation in intermodal transport, but the wrong way procedurally to go about resolving them. The proposed Act represents the right solution in the sense that, in principle, it would establish an uniform legal regime for all participants in all modes of carriage.[31] It is, however, flawed by the exclusion of interstate rail and trucking companies as performing carriers. It is also limited by the necessary involvement of a sea stage to invoke its application. Thus the new US COGSA would not govern the rail-truck intermodal traffic that has recently grown so rapidly in North America, nor any combination of air-truck-rail movements in the absence of a sea stage. Such a partial solution to the demands for multimodal uniformity is inadequate and may result only in adding another layer of complexity to intermodal relationships.

In any case, this unilateral American approach is an unsuitable way to go about resolving a matter of international concern. Since US COGSA is intended to regulate trans-national transport, it cannot succeed without the cooperation of foreign participants and authorities. An uniform regime for intermodal transport can probably only be effectively achieved, if at all, by general international agreement. As a major trading partner of the United States, Canada may have much to say about the particular form of the proposed revisions to US COGSA but little opportunity to do so in the US Congress' process of their enactment.

An example of the international dimensions and difficulties that the revised US COGSA would precipitate is provided by an intermodal movement involving air transport.[32] Carriage by air is regulated, even in and out of the United States, by the Warsaw Convention,[33] which is an international convention having the force of law amongst all participating countries. One nation cannot legislatively override an international convention agreed with many other nations. Thus US COGSA cannot be extended to cover the whole of an

intermodal movement that includes an air transport stage in contradiction of the Warsaw Convention. And such contradiction is practically inevitable since the proposed US COGSA's limit of liability would be 2 SDR per kilo, which is very substantially less than the minimum limit of liability of 17 SDR per kilo mandated by the Warsaw Convention.[34]

Similar collisions of international transport conventions and the proposed US COGSA would occur in intermodal movements by land and sea between the United States and Europe. In that continent, international agreements mandatorily allocate the risks and liabilities for both rail and road transportation.[35] Since the United States is not a party to these conventions, its enactment of the proposed revisions to COGSA would not of itself cause conflict. But an intermodal movement between the United States and, say, the interior of Germany, would attract the compulsory application of either or both of the land transport conventions to the surface leg or legs of carriage in Europe. Hence an attempt to apply the revised US COGSA to the whole movement would directly conflict with these international regimes.[36] In short the attempt to project an uniform intermodal regime through US COGSA is likely to be a blighted step, which will only increase the conflicts of laws over transport services between trading partners, like the United States and Canada.

Effects of a Revised US COGSA on Canada

Canadian concerns about the proposed revisions to US COGSA stem from the well known fact that Canada's economy is dependent on foreign trade, most of which is with the United States. It may be noted immediately, however, that transport services used to move most of this north-south trade are provided by road and rail carriers. Unless some sea carriage is involved, US COGSA would not operate upon this traffic.[37]

The principal area of practical impact of a revised US COGSA is likely to be on Canadian trade through the United States to or from other destinations or origins, such as from Pacific rim countries and Europe. Under the original proposal to apply US COGSA even to traffic "through" the United States,[38] such Canadian trade would obviously have been affected. However, the revised proposal to limit the reach of US COGSA to carriage to or from the United States will not leave Canadian trade through the United States unaffected. By way of example, consider two typical movements, one from Montreal to Europe, the other from Japan to Toronto.

Suppose the export cargo from Montreal is moved by CP Rail to New York for sea carriage to Rotterdam and inland delivery. The revised US COGSA would not apply as compulsory law because the movement is neither to or from the

United States and US legislation cannot have statutory force in a foreign country like Canada. Canadian laws would apply to the combined transport document issued by CP Rail as the contracting carrier for this movement. Since Canada does not have legislation fixing the rights and liabilities of multimodal operators or rail carriers, CP Rail would be free to contract with the cargo owner as they may agree. This agreement could include the risks that CP accepts for itself and its subcontracted service providers, the limits of their liability, and the choice of governing law.

If an accident occurs and the cargo is lost or damaged, the claimant might pursue several possible defendants. CP Rail would always have to answer a claim for loss anywhere in the movement because it was the contracting carrier. The terms of its (Canadian-made) contract, evidenced by its combined transport document, would govern. Similarly a claim against CP Rail as the performing carrier on the initial carriage by rail for a loss on that stage would be determined by the same regime. But a claim against the shipping line for a loss on the sea leg of the movement would fall under the revised US COGSA because that subcontracted stage of carriage commenced from New York. And US COGSA would fix the liability of the shipping line, if any, even though the Canadian shipper had no commercial or contractual association with it because the proposed revisions would apply "with the force of law" to determine "the complete and exclusive remedy"[39] against such a performing carrier.

By contrast, consider the importation of goods to Canada from the Pacific rim. Suppose the cargo is carried by sea from Japan by American President Lines (APL) to the US west coast and then forwarded by an American railroad to Chicago and thence on a Canadian railway to Toronto for delivery. The revised US COGSA would not apply as compulsory legislation to this movement either, but APL would be quite likely to incorporate it by reference in its contractual documents. Such a choice of governing law would likely be respected by the courts.[40]

If a claim arose, the Canadian importer could always sue APL as the contracting carrier,[41] whose liability under the terms of the intermodal contract would be determined according to the revised US COGSA.[42] Equally, if APL were to be sued for loss during the initial sea voyage to the United States, the choice of US COGSA as the governing law would also apply to its conduct as the performing carrier. But direct claims against the rail carriers for damage to the cargo while in their charge might produce some unexpected results. The transit by the American rail carrier would not fall within the scope of application of the revised US COGSA. Its intermodal ambit excludes US interstate railway companies when, as here, they act as performing carriers.[43] Presumably the American rail carrier would try and rely on the terms of its

service contract with APL and its published tariff, subject to other US federal law about railways. The Canadian rail carrier, however, is not obviously excluded from the operation of the proposed US COGSA, which otherwise continues to apply as the chosen law to govern the intermodal delivery to Toronto.[44] Absent any mandatory federal Canadian law that would override this choice of governing law, once the Canadian importer becomes a party to the multimodal contract with APL, a Canadian court is likely to honour the application of the revised US COGSA. This may surprise the Canadian importer and may alarm the Canadian rail carrier. They might assume that a claim by a Canadian cargo owner against a Canadian rail carrier in a Canadian court would be decided according to Canadian law. In fact, the defendant rail carrier might find itself trying to argue its way out of liability as a performing carrier under the revised US COGSA.[45]

Conclusion

The impacts of the proposed revisions to US COGSA would not create a happy state of affairs in intermodal services. As laudable as the broad conception of a single uniform regime that covers all the players in any intermodal movement is,[46] this scheme will not achieve that objective. Meddling unilaterally with the alternative choices of the Hague/Visby or the Hamburg Rules on sea carriage will only compound the diversity that these internationally agreed sets of rules seek to eliminate. Espousing a regime for apportioning the risks in intermodal transport that has to include a sea passage to operate and may not reach road and rail carriage in some circumstances is too incomplete to be supported. Confronting existing internationally agreed conventions, which establish mandatory unimodal regimes, with an unilaterally created intermodal solution is an impractical way to proceed and cannot succeed. Projecting American solutions to intermodal problems beyond the United States into the trade and transport of other countries, such as Canada, is unacceptable. In short, the proposed revisions to US COGSA are more likely to irritate than to resolve the continuing problems of apportioning the risks of loss and liability in intermodal transportation.

ENDNOTES

* This paper and its presentation is made possible by the generous support of the SSHRCC as part of a larger study of modern shipping practices being undertaken by the author.

1. See H.M. Kindred, "Muddling Through the Maze of Liability for Multimodal Mishaps", Proceedings of 30th Annual Meeting, CTRF (1995) at 315.

2. See US MLA, Revising the Carriage of Goods by Sea Act: Final Report of the Ad Hoc Liability Rules Study Group (Feb. 15, 1995), as Revised by the Ad Hoc Review Committee (Feb. 9, 1996). The Rev'd. Rep. contains a 45-page discussion of the issues, a draft Bill for enactment in place of the current COGSA, and a section-by-section analysis of the draft, as well as an appendix displaying the changes to existing law. If the Report is adopted by the US MLA at its annual meeting this year, it will be presented to Congress with a resolution urging enactment of the proposed Bill.

3. 46 U.S.C. App. § 1300-1315.

4. See the Carriage of Goods by Water Act, S.C. 1993, c. C-21.

5. *Supra* note 2, Rev'd. Rep. at 3.

6. Draft Bill, *supra* note 2, s. 1(e).

7. *Ibid.*, s. 1(a). And see text *infra* at note 29.

8. Section 1(b).

9. But not charter parties.

10. Draft Bill, *supra* note 2, s. 1(g).

11. SDR are the Special Drawing Rights of the International Monetary Fund whose fluctuating value is calculated daily from a basket of real currencies. An exchange rate of $1.45 US per SDR is assumed in this paper.

12. Draft Bill, *supra* note 2, s. 4(5).

13. *Ibid.*, s. 3(6)(d)(ii). *Cp.* Hague/Visby Rules art. III (6 bis).

14. Section 4(5)(e).

15. Hague/Visby Rules art. IV(5).

16. *Supra* note 14.

17. Unless the actual cargo loss is less than the applicable limit of liability.

18. Hague/Visby Rules art. III(8).

19. Hague/Visby Rules art V.

20. Draft Bill, *supra* note 2, s. 5.

21. *Ibid.*, ss. 3(8)(a) and 4(5)(2). As regulated by US Shipping Act of 1984, s. 3(21).

22. Section 4(2) first proviso.

23. Section 1(c), unlike the Hague/Visby Rules, but like the Hamburg Rules.

24. Section 4(2) second proviso.

25. Section 4(2)(b).

26. Subsection 3(3) and 10.

27. Section 3(8)(b).

28. Unnumbered opening section.

29. Section 1(a)(v).

30. Section 2.

31. As suggested at note 1.

32. Though US COGSA would only be invoked if a sea leg is involved in an intermodal delivery, there are multimodal movements of goods by sea across the Pacific Ocean to the North American west coast and thence by air to Europe.

33. Convention for the Unification of Certain Rules relating to International Carriage by Air, 1929 as amended.

34. *Ibid.*, art. 22 as amended. Always assuming the value of cargo lost or injured is greater than the limit of liability that would be established by the revised US COGSA.

35. Convention on the Contract for the International Carriage of Goods by Road, 1956 (CMR) as amended, and Convention concerning International Carriage by Rail, 1980, Appendix B (CIM).

36. The partial exclusion of road and rail carriers from the operation of US COGSA proposed in s. 1(a)(v) will not resolve this conflict of laws because that section does not exclude road haulers and rail companies when acting as contracting carriers and, moreover, appears by its citation of "interstate trucking or rail carrier" to be aimed at these service providers in the United States alone.

37. But coastal deliveries with intermodal land links could be affected.

38. *Supra* at note 28.

39. Draft Bill, *supra* note 2, opening unnumbered enacting clause.

40. Unless it happened to be contrary to the law of Japan, where the intermodal contract was concluded and the goods were received by APL, the contracting carrier.

41. In practice, the importer would call upon its cargo insurance for full recompense, and the insurer would bring the action against APL to recover as much as possible of its insured's claim.

42. This conclusion reasonably assumes that by the time the claim arose the Canadian importer had become a party to the multimodal contract with APL.

43. *Supra* at note 29.

44. See note 36, *supra*.

45. Unless the extent of its liability under US COGSA turns out to be less than under the terms and conditions of its service, as regulated by Canadian law.

46. And so also does away with all the legal awkwardness of Himalaya clauses and circular indemnities.

CHEMICAL TRANSPORTATION RISK MANAGEMENT

by

John Cowan

KPMG Management Consulting

This paper explores the risks associated with the transportation, storage and distribution of bulk chemicals and chemical products. By identifying and measuring the risks in a systematic fashion, a comprehensive and effective strategy can be developed for minimizing the overall risks. This paper outlines the components of an effective risk management program for companies associated with the transportation and distribution of chemicals in Canada.

Chemical industry today

There are not many products reaching the consumer market that do not contain chemical products or which are not being treated or packed in products derived from the chemical industry. A whole range of items such as clothing, food, household articles, transportation equipment, cars, agricultural products etc. which are an integral part of our day-to-day life and economy rely on the resources and outputs of the chemical industry.

Chemicals are part of numerous industrial processes. There are three major levels of chemical processing:

i Raw materials: includes minerals, ores, plants, petroleum based feedstocks, oil, natural gas and coal.

ii Basic chemicals and derivatives:

- inorganic—e.g. sodium hydroxide and sulfuric acid
- organic—such as ethers and enzymes
- petrochemicals—e.g., methane, ethylene, methanol, styrene and benzene

iii End uses: e.g. bleaching and cleaning compounds, paints, medicines, plastics, textiles, polyesters etc.

The chemical industry in Canada produces almost $14 billion of product per annum and is responsible for a significant portion of Canadian exports and balance of trade. In 1994, exports totaled $9.4 billion and imports $13.1 billion.

In the east, a majority of chemical shipments originate in Sarnia, Amherstburg and Hamilton, destined for Quebec and the Maritimes. Other significant eastern load centres are Belledune and Becancour. Major receiving ports include Port Hawksbury, Montreal, Charlottetown and Halifax.

On the west coast, Vancouver is the leading port for-domestic loadings in Canada, handling over 62% of Canadian domestic shipments in 1994. The port handled 2.7 million tonnes in 1994 with Kitimat following with 1.3 million tonnes. In all, Canadian ports handled 7.8 million tonnes of chemicals and chemical products in 1994. The industry plays a crucial role in Canadian international trade and as the volumes indicate, is responsible for a considerable portion of the overall demand for transportation services.

In spite of the fact that nearly all consumer products contain or have been treated with chemicals, the public is wary of the chemical industry and its products, mostly through ignorance of the products and how they are produced and distributed.

The industry has been plagued internationally by serious accidents like those at Saves, Bhopal, Sandoz etc. Further more, certain chemicals and product groups have created huge waste and recycling problems. The problems are normally viewed at the local or municipal level, however the impact is felt throughout the country. There is a very real concern about the dangers associated with handling and transporting chemicals, and it is imperative that companies address these dangers by adopting effective risk management programs.

"Responsible Care"

Responsible Care is the global chemical industry's initiative for the pursuit of excellence in protecting human health, safety and the environment with respect to all their activities. These activities include manufacturing, design and engineering, research and development and transportation and distribution of raw and refined chemical products.

This initiative is a proactive effort by the chemical industry to reposition itself with respect to the public and authorities through substantial and measurable actions. By acknowledging its past failures, the industry is aiming to communicate with its industry members and the public on constructive and forward thinking actions to be taken. This program was established by the chemical industry associations to help their member companies achieve continuous improved performance in the fields of human health, safety and the environment. It hopes that a new attitude and way of thinking on the part of the industry will result so that it will be perceived as a partner with society as opposed a problem for society.

The *Responsible Care* initiative was started in Canada in 1984. It was adopted by the US Chemical Manufacturers Association in 1988. In 1989 the UK industry association launched its program and in 1990 the European Chemical Federation endorsed the international guiding principles, while leaving the commitment and implementation to the national associations and individual companies. Today Japan, Australia and New Zealand are operating programs committed to responsible care and action. This internal appreciation of the problems beleaguering the international chemical industry

is important for the successful development of programs and initiatives to reduce the risks associated with transporting and distributing chemical products today.

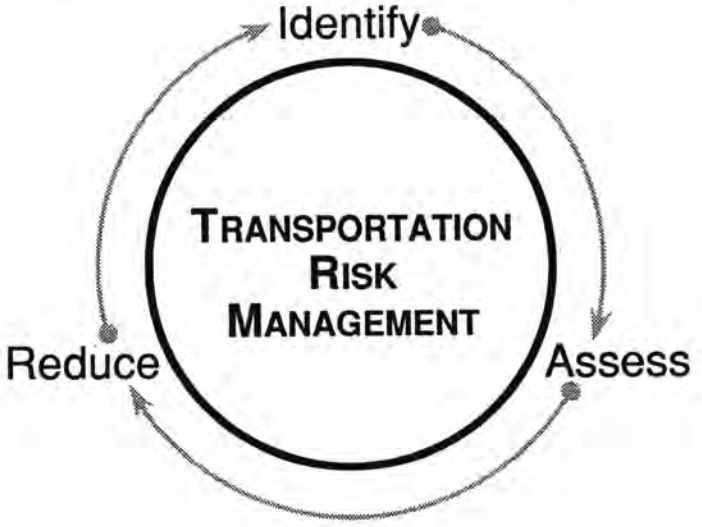
Transportation Risk Management

In order to market its products, the chemical industry relies on safe and effective transportation and distribution systems. These systems utilize all forms and transport modes—rail, truck—both bulk and van, marine, pipeline and even air. Although some chemical companies operate private transportation fleets and pipelines, most chemical transportation services in Canada are purchased from third party carriers. Given the magnitude of the chemical transportation activity and the nature of the products being transported, it is imperative that transportation services be undertaken in as safe a manner as possible.

The transportation and use of hazardous materials including chemicals poses threats that are of concern to individuals and industry, but which are not fully understood in terms of their likelihood of occurrence or the severity of the impact of an accident. The purpose of risk management is to provide basic data and procedures necessary to refine and evaluate individual hazardous material accident scenarios in terms of their probability or frequency of occurrence.

An effective risk management system allows companies to:

- ***identify*** the risks associated with transport
- ***assess*** the level of risk associated with the product, and
- ***reduce*** the overall chance of a dangerous accident occurring.



The nature of bulk and containerized transportation of dangerous and hazardous chemical material entails that unforeseen and varied emergency situations can occur. These emergencies can happen quickly, unexpectedly and usually require immediate response. The fact that hazardous materials like chemicals pose a threat to public safety and the environment is of vital concern to the government and the chemical industry. So should this concern be shared by those involved in the storage, distribution and transportation of chemical products.

Currently there is a clearly urgent need to ensure that in the event of an accident, those parties closest to the accident are able to respond quickly, efficiently, and effectively to reduce or prevent adverse impacts on public safety. Timing is critical in the first moments of an accident. A mismanaged response due to the lack of preplanning can contribute to fatalities and injuries, as well as an increase in damage to property and the environment. The cost and expense of these accidents must be borne by the parties involved. Hence the potential magnitude of claims and remedial treatment should be acknowledged by the transportation industry.

Quantifying the risks associated with the transportation of chemicals enables the prediction of high risk activities and the corridors or areas where events are likely to take place. This allows for effective preplanning, including the establishment of contingency plans and equipment to cope with hazardous situations.

An Effective Risk Management Program

Transportation risk management programs evaluate chemical transportation risks (risk assessment) and develop and implement risk reduction measures (risk reduction) that are appropriate to the level of risk.

Risk assessment should include the analysis of the following:

- hazards associated with the chemicals being distributed,
- likelihood of accident/incidents and resulting releases; and
- impact or consequences of these releases on the public and the environment.

A qualitative risk management process is the basis for implementation. In transportation, qualitative risk management generally uses the experience and judgment of knowledgeable people to

- analyze the causes and effects of transportation accidents and resultant releases; and
- provide appropriate ways to reduce their likelihood, magnitude and/or consequences.

When applying the process, the shipper determines which chemicals receive the initial focus by prioritizing the movements in a logical manner. This process requires the shipper to utilize all available commercial and distribution information. The balance of the process should be carried out in conjunction with the carrier to ensure that the entire

transport process, including loading, handling, discharging and storage is carried out in a safe and environmentally sensitive manner.

The development of a responsive risk management program for chemicals involves the following six steps:

1. Chemical listing

Companies should list the types of chemicals which they transport, handle or distribute. Chemicals may include products, raw materials and associated wastes. This should include the identification of the current modes of transport employed along with the annual volumes. Potential changes in transport mode should also be included, to account for a modal shift brought about by either market forces or government regulations.

2. Chemical hazard ranking

The purpose of this step is to develop a list of prioritized materials being transported based on inherent risk. Once this inventory is completed, a simple hazard ranking process can be used to prioritize those chemicals that may have the potential, due to their inherent nature, to adversely affect people and the environment.

One approach to hazard ranking may be to use existing regulatory classification systems, such as those established by Transport Canada's Transportation of Dangerous Goods regulations.

Other ranking systems may be utilized, with the emphasis on prioritizing the chemicals based on the inherent risks associated with transport. Risk factors include:

- flammability
- vapour pressure
- explosivity
- reactivity
- toxicity
- corrosivity
- overall environmental sensitivity

Ranking systems would include the assignment of degree of risk for each component. These would be tallied and then analyzed to determine a realistic weighting for each risk category. This process is a critical step in the management process, as it sets the stage for further analysis. Most importantly, it identifies the issues for an effective risk reduction program to address.

3. Movement Exposure Ranking

This step measures and ranks the exposure potential of individual product movements to the public and the environment. It involves the assessment of risks associated with the actual movement of the products. The applicable information for risk weighting is determined by load/discharge routings and patterns combined with product volume. The actual nature and size of the shipment type should be assessed so that like shipments are compared with like shipments. i.e. the same product type, size, packaging etc. should be identified.

The following factors are to be considered:

- annual volumes per destination—number of trips anticipated per annum per mode
- trip distance relative to the established or anticipated accident rate for the mode used.
- population densities of the regions and corridors where the products are transported
- nature of the transportation corridor utilized, i.e. road, railway, canal etc.
- volumes and seasonality of traffic flows in those corridors
- environmental considerations.

4. Prioritizing For Risk Reviews

This involves analyzing the previous three steps to establish the potential for accidents to occur and the priority in which they should be treated The priorities will allow effective weighting and facilitate the decision making process for those customer routings which should be more closely reviewed in Steps 5 and 6.

In setting priorities for risk reviews, the frequency of future reviews should also be considered with the higher priority product routings receiving more frequent reviews. Shippers and carriers should use these guidelines to develop and implement a program of risk reviews that are performed on a regular basis.

5. Conducting Risk Reviews

This step is specific to each individual shipper or carrier. A program that applies and works well for one may not be applicable or effective for another. The goal is to anticipate and determine possible actions to help prevent incidents that may cause a loss of containment of a chemical throughout the transportation chain or during storage periods.

In addition to commercial and routing research, visual information gathering and samplings along the transportation corridor should be carried out. The risk

reviews themselves involve a more detailed consideration of the hazards and the exposure assessment, coupled with an assessment of the potential impact of both accident and chemical release.

The process would normally focus on current activities and practices. Various "what-if" scenarios of accident occurrences and probabilities may be helpful in assessing the acceptability of current practices and procedures.

Risk reviews should commence by asking the following questions:

- What are we doing now ?
- What could go wrong ?
- What could we do differently and how?

Once these questions have been addressed, the following risk review activities can take place:

Hazard Assessment:

This determines the degree and nature of the hazards of a particular chemical relative to those of other chemicals. It is measured relative to releases occurring as a result of transportation accidents. Examples of such issues include:

- acute inhalation hazard to humans
- explosivity and flammability
- vapour cloud potential
- reactivity
- environmental damage potential
- any unique hazards inherent to the particular product
- regulations applicable—either by industry or government, including both domestic an international
- container volume
- container pressure and temperature.

Exposure Assessment:

This determines the degree of exposure to accidents and the resultant releases for a chemical movement, relative to other chemical movements. Pertinent issues include:

- annual volumes transported
- trip length, both in time and distance
- number of trips in given time period
- average and specific population densities along the route
- size of metropolitan areas along the route
- nature and number of waterways along the route.

Accident Likelihood Assessment:

This step determines the likelihood of accidents occurring within the transportation mode employed. It includes an assessment of the carriers and their practices currently utilized for this movement. Issues to focus on include:

- relative accident rate for transportation mode employed
- relative accident rate for the carrier used
- effect of any route specific accident rates
- special procedures or controls employed
- specialized training
- effect of specialized equipment and technology
- age and condition of equipment and facilities.

Release Likelihood Assessment:

This step determines the likelihood of an accident causing a release from the container or vehicle used for the movement, relative to other container choices. Issues include:

- types of release / spill scenarios, often determined by the product state
- impact resistance of the container or fittings to potential damage
- special protective features e.g. shields, double tanks, containment areas
- container/vehicle preparation e.g. preload inspection
- container/vehicle age and maintenance history
- container/vehicle accident or incident history.

6. Developing and Implementing Risk Reduction Alternatives

Once the review is complete, the key procedure is to consider how to reduce the risks. The following are components of the transportation chain where some risk reduction alternatives could be employed.

Container selection and design

Containers come in a wide variety of sizes, type, configurations, and materials for construction. To reduce risk in transportation, the following should be considered when determining container design or modifications:

- materials of construction, design strengths and product compatibility
- pressurized versus non-pressurized designs
- type of loading/discharge fittings, their location on the container (top versus bottom), excess flow valves and physical protection from incidents and vandalism
- owned versus leased or carrier supplied containers
- frequency of container testing, inspection and maintenance
- pressure relief valves versus rupture disc vents, their operating settings and test frequency of relief valves
- insulated versus uninsulated and thermally protected containers
- volume/weight selection which relates to the amount of material allowable by mode
- degree of adequate securement before shipment to reduce the likelihood of spillage during normal distribution handling
- special design features such as double-walled containment, multiple compartments and protective shields.

Mode selection:

Risk reduction alternatives may be developed by considering the various modes, modal combinations and options within a particular mode. The primary modes are rails, truck, air, marine and pipeline.

The following options exist within each mode:

- common versus contract and private carrier
- truckload (TL) versus less-than-truckload (LTL)

- bulk versus packaged or containerized cargoes
- ship versus barge
- intermodal versus single mode.

Important items to considered with regard to modal choice include:

- number of trips
- modal incident history
- maintenance procedures and history
- in-transit storage delays
- time in transit
- product handling techniques.

Carrier selection:

An important aspect for risk reduction involves the considerations given when selecting a carrier. One way is to select a carrier based on its safety performance record, safety programs in place, and the quality of its equipment, facilities and personnel. The type of training programs for all personnel, including ongoing and future, should also be reviewed.

While there is always a comfort level associated with using carriers which have previously handled a company's transportation programs, their performance should be measured and benchmarked against others in the industry. It is important that shippers gain an understanding the needs and requirements of the carriers so that an effective partnership can be developed.

Route selection

In order to minimize overall risk, the following should be considered when selecting a route.

- accident frequencies along the route
- length of the route
- physical condition and maintenance of the route
- the type of terrain to be encountered
- the population densities and location of towns and population areas along the route

- environmentally sensitive areas along the route.

Operating procedures and practices:

Changes in operating procedures and practices influence the degree and level of risk and should be factored into the assessment of risk reduction alternatives. The following management procedures should be reviewed:

- frequency and intensity of carrier evaluation
- details of loading and securement
- inspection procedures during pre-loading, post loading and in-transit
- speed control and weight limit programs
- specified routings through metropolitan and high density areas
- unloading procedures
- maintenance schedules—including the level of comprehensiveness and frequency
- specific policies affecting transport operations such as operating during daylight hours or weekdays only.

Emergency response capabilities

Emergency response and public preparedness are an integral component of risk management because they can reduce the consequences resulting from an accident. Within Canada, the Transportation of Dangerous Goods Act and Regulations have mandated the required Emergency Response capability on certain high risk chemicals transported in Canada. Shared knowledge of emergency response plans within corridor communities on high volume routes can assist those communities in preparing their own emergency response plans.

Training

Personnel training is a vital component of a risk reduction strategy and program. Detailed training in loading, unloading and inspection procedures are important for reducing the chance of an incident occurring. Training in safe handling, hazard identification, inherent dangers and chemical properties associated with the chemicals being carried should be included. Thorough training in emergency response procedures is necessary to quickly and effectively minimize any incident-related harm to the general public, transportation personnel and the environment.

Chemical and physical properties

In certain instances, options may be available to alter the physical properties or state of the chemicals being stored or transported. These options could reduce the volatility, explosivity, reactivity or toxicity of the products. Examples of this include

- further processing to afford a more inert state for transport
- mixing or diluting to decrease toxicity
- lowering temperature
- adding inhibitors.

Once all of these steps and components of risk reduction alternatives have been addressed and analyzed, the risk profile of the product can be adequately developed. What may have been considered a high risk move initially, could become a lower risk priority once the assessment process has been completed. Overall, it is most important to optimize the level of operating safety on the route, even if this means changing the initial risk assessment profiles for each product.

7. Updating the risk management process

Effective transportation risk management must be considered as a continuous process of risk reduction. Since many of the factors involved in the overall risk management are dynamic, this process should be repeated at regular intervals. In addition, as more pertinent information becomes available, it should be used to improve this process. Higher risk movements should receive more frequent and thorough reviews.

Conclusion

Risk management programs are becoming increasingly important, especially in light of the public sensitivity to transport safety and environmental concerns. Shippers, carriers distributor and all persons and companies associated with the distribution of chemicals must further develop their appreciation and utilization of risk management and reduction programs. By employing an effective transportation risk management regime, those companies associated with the transport and distribution of chemicals will enjoy a safer working environment and lower their overall costs of conducting business.

Sources

The Canadian Chemical Producers' Association Transportation Code of Practice. Reprinted from PLANT/OPERATIONS PROGRESS, October 1990.

Chemical Trade and Waterways: The Canadian Chemical Industry and the Use of Marine Transport. Canadian Coast Guard. Draft report. 1995.

Finlayson, David. Directions in Chemical Marine Emergency Response—An Industry Perspective. Canadian Chemical Producers' Association.

Marine Emergency Response Contingency Plans. Chemical Manufacturers Association 1993.

Summary Report: BC. Marine Terminal Emergency Response Planning Review 1994. The Canadian Chemical Producers' Association.

Transportation Code of Practice—Risk Management. The Canadian Chemical Producers' Association.

Establishing Insurance Requirements for an Evolving Shortline Railway Industry: The Ontario Experience

Mike Ronson and Ann Misetich
Ministry of Transportation, Ontario*

INTRODUCTION

Ontario is in the midst of establishing new policies on insurance for shortlines under its jurisdiction. To set the context, the first part of this paper will provide an overview of the changing shortline environment that has prompted Ontario to develop new policies. Initially, discussion will focus on the opportunities for shortline growth offered by impending changes to the regulatory framework governing federal railways. This will be followed by a review of opportunities presented by the current business climate -- in particular, the financial problems of the eastern railway networks. Finally, Ontario's response to these changes, its *Shortline Railways Act*, will be noted.

The second part of this paper will discuss the issues facing Ontario as it develops its policies on insurance. The first issue to be addressed is the extent to which governments are responsible and liable for ensuring that railways have sufficient insurance coverage. This will be followed by an examination of ways in which government could ensure minimum coverage levels. Included here will be a description of the case-by-case approach and the use of published levels. Finally, issues surrounding the deductible component of coverage will be looked at.

CANADA'S CHANGING RAILWAY INDUSTRY

Regulatory Environment

Since the passage of the *National Transportation Act (NTA)*, 1987, Canadian law has allowed for third parties to acquire segments of railway land and infrastructure slated for abandonment, or by way of a transfer independent of abandonment, and maintain rail services through shortline operations. Shortline carriers, with lower operating costs, can offer tailored services to shippers on feeder lines, while simultaneously providing traffic and revenues to Class I railways. Shippers, carriers, communities and the economy all stand to benefit. Despite this, there has not been a proliferation of shortline railways in Canada since the *NTA* came into effect.

* The views expressed are those of the authors and not necessarily those of the Ministry, nor their colleagues, whose constructive criticism is gratefully acknowledged.

The proposed *Canada Transportation Act (Bill C-101)* is expected to change this --one of its key objectives is to encourage shortline growth. The *Act* will make it easier for Canada's railways to sell or discontinue the operation of unprofitable branch lines. The current process is slow and cumbersome because federal regulations require extensive documentation and public input. The new *Act* will only require railways to file rolling three-year plans identifying what they will retain, shortline or abandon, and once this is done, the abandonment of individual lines could only take a few months. Both railways are expected to start selling and abandoning track aggressively after the *Act* is passed in early 1996. Under the new legislation, Transport Canada has estimated that 20% of Canada's 71,000 kilometres of track will eventually be abandoned.[1]

Deregulation of the U.S. rail industry in 1980 has had a tremendous impact on American shortline development. More than 300 new companies have started up shortlines since this time. The peak year for shortline creation in the 1980's was 1987, when 46 shortlines were formed and 6,674 miles (10,740 kilometres) of rail line were transferred to shortline operation.[2] Today, the industry employs 10,500 people and runs over 59,000 kilometres of track.[3]

The Business Environment

Intense trucking competition, the loss of rail dependent industries and declines in bulk commodity traffic have put both Canadian National (CN) and CP Rail (CP) in a position where they have excess capacity. To be competitive, these carriers are shedding uneconomic feeder line operations. Further paring of their money losing eastern networks and joint use of viable rail segments are essential to restoring their financial viability. CN and CP maintain that they have lost over $2 billion in eastern Canada between 1988 and 1993, partly because they are forced to operate uneconomical branch lines. For instance, more than half of CP Rail's network carries just 5% of its total business.[4]

CP Rail System is reviewing the viability of 7,600 kilometres of its track -- 25% of its North American network.[5] Identified priorities include grain dependent lines in western Canada, lines in the U.S. Midwest, the Quebec line from Montreal to Quebec City, and 480 kilometres in the Ottawa Valley. CP's November 1995 restructuring, driven by management's pledge to increase rail earnings and overall corporate profitability for investors, specifically targeted CP's eastern network for remedial measures.

To date, CN has identified 6,400 kilometres of low density track for rationalization by 1999.[6] Reports indicate that CN will sell or abandon some 3,200 kilometres of this track in 1996 alone and that half of that will be in Eastern Canada.[7] The company has indicated that it wants to cut 50% of its network east of Winnipeg, abandoning 30%, and selling the rest to shortline operators.[8]

The fundamentally changing railway industry, and regulatory environment in which it operates in Canada, signal the potential for unprecedented shortline growth. As many as 20 portions of CN and CP are estimated to have enough

potential to be taken over by shortline railways.[9] The president of Railtex, Inc., the Texas based holding company that owns 25 shortline properties in North America, estimates that CN and CP will put some 50 lines up for sale, and that Railtex will probably pick up ten of those for about $10 million each.[10] Lines with annual traffic of at least 2,000 carloads for each two-person crew are considered viable candidates.[11] In Ontario, Railtex may try to buy lines between Toronto and London via Stratford, in the Bruce Peninsula and in the Ottawa Valley.

ONTARIO'S NEW SHORTLINE RAILWAYS ACT, 1995

Prior to the passing of Ontario's *Shortline Railways Act* in 1995, provincial shortlines were governed by the outmoded *Railways Act* -- legislation last consolidated in 1950 and not amended since. The old legislation did not adequately address the formation, licensing, regulation, safety enforcement, and market entry/exit of shortlines. One of these deficiencies was railway insurance.

The guiding principles for the new legislation were: keep it simple and workable, while minimizing the administrative burden on the province. Generally speaking, the new *Shortline Railways Act* focuses on safety, and not economic regulation, and cost-effectively employs the existing safety expertise within the federal government. It eliminates the requirement for a special Act of the Legislature to incorporate a company wishing to operate a shortline railway, and simplifies the approvals process -- it will make starting shortline rail services in Ontario easier.

Ontario's new *Shortline Railways Act*, and its associated insurance requirements, apply only to shortline railways, as carefully defined, and not to all railways operating within provincial boundaries. The *Act* defines shortlines in the following way:

"Shortline railway means a railway within the legislative jurisdiction of the Province of Ontario...that has been the subject of an abandonment order, discontinuance or conveyance, has been discontinued or conveyed under this *Act*, or...that was acquired before the coming into force of this *Act* by a corporation incorporated by a special Act in accordance with the *Railways Act*".

ESTABLISHING INSURANCE POLICIES

The Government of Ontario has consulted with numerous insurers, railways, governments, railway associations and agencies on issues surrounding shortline insurance. Internally, it has drawn heavily upon the expertise of those within its Office of Insurance and Risk Management -- the people responsible for insuring government operations, including the Province's rail services. As well, its legal counsel has been fully integrated into the project.

Government Responsibility and Liability

Responsibility for shortlines falls under either federal or provincial jurisdiction. The *Constitution Act*, 1867, allows the Parliament of Canada to declare some shortlines as being to the "general advantage of Canada". As well, federal jurisdiction extends to "...*railways... connecting the Province with any or others of the Provinces, or Extending beyond the limits of the Province*". Other shortlines normally would fall under provincial jurisdiction.

Currently, the federal government addresses its responsibilities for shortlines under its jurisdiction through the *National Transportation Act*. However, this Act is expected to be superseded by the proposed *Canada Transportation Act* in early 1996. The proposed *Act* states that "The Agency (Canada Transportation Agency) shall issue a certificate of fitness ... if the Agency is satisfied that there will be adequate liability insurance coverage...as determined in accordance with the regulations".

Ontario recently passed its *Shortline Railways Act* to deal with the shortlines under its jurisdiction. This *Act* states that "The Registrar may issue a license to operate a shortline railway if the registrar is satisfied that there is adequate liability insurance coverage for the operation of the shortline railway".

It is generally accepted that governments, as a result of their jurisdiction over rail matters, would find themselves targeted in situations where a railway becomes insolvent and third party liability claims exceed the railway's insurance coverage. Third partly liability in this instance is understood to include some damages related to the environment and to those involving passengers. Government concerns tend not to extend to coverage of cargo as this is understood to be a matter between shippers and carriers.

While there are different views as to the extent governments may be found technically liable for shortlines having insufficient insurance, this may not be overly important in practice. Governments may find it cheaper to pay claims rather than fight them in court and/or find themselves compelled by public duty to cover them.

ENSURING ADEQUATE COVERAGE LEVELS

The Rationale

Each railway operation presents certain risks to the public. If these risks materialize, and cannot be covered by a railway's insurance policy or assets, they can end up being forced onto others. This could be through government compensation, personal losses, damage to the environment and so on. In essence, operators whose risks are not covered adequately by insurance are being subsidized by somebody. The dollar amount of this subsidy would be the cost of bringing a railway insurance policy up to a level that the public regards as sufficient.

Failing to ensure that all operators carry adequate insurance could also be considered unfair to those that do. An example of this would be a situation where two tourist shortlines set up in close proximity to each other and one opted to maintain insufficient insurance. The negligent operator could conceivably undercut the ticket prices of the other because of the savings it achieved on insurance premiums. A more common example of this fairness issue can be borrowed from the truck mode where it is often argued that failure to crack down on unsafe trucks puts companies who invest in proper maintenance at a competitive disadvantage.

Setting Levels by Individual Review

Reviewing the risks posed by applicants on a case-by-case basis, and consequently setting a level for each, is one method of ensuring that railways carry sufficient coverage. It is the method being proposed by the federal government for shortline railways which fall under the *Canada Transportation Act*.

Under the federal government's proposed plan, a railway would file information such as ridership, mileage, volume of traffic, population density of areas served, number of level crossings, speed and safety record with the Canada Transportation Agency. From such data the Agency's experts would determine the risks posed by the shortline and an appropriate level of coverage. The Agency would also collect financial information on the railway and use it to determine an acceptable limit for the deductible portion of the coverage. The process being proposed by the federal government is essentially parallel to that performed by insurance companies when quoting rates for applicants.

Case-by-case analysis is very comprehensive -- it clearly provides for the individual risks posed by each applicant to be taken into account. As a result, government is provided with a high level of confidence that the public is adequately protected and that a fair decision was made. Being able to fall back on the findings of a comprehensive review may be particularly useful should a government find its determination questioned or find itself being sued for failing to ensure adequate coverage. Relevant to the above is the fact that both Ontario's *Shortline Railways Act* and the *Canada Transportation Act* have mechanisms whereby an applicant could appeal its insurance requirements. However, comprehensive approaches such as case-by-case review also have their drawbacks. There is great potential for this type of approach to generate red tape and costs.

Using Published Levels

The key alternative to the case-by-case approach is the use of published levels. The simplest of these would be to have a universal minimum level. Establishing a universal minimum would be advantageous in that it would eliminate uncertainty and generate few administrative costs. Shortline investors would know exactly, from a regulatory standpoint, what amount of insurance they

would need. Government would not require expertise to review each case -- expertise for which the cost would either have to be paid for by the government or recovered from the shortline.

A universal level, however, raises questions in regard to how different sets of risks posed by different shortlines would be accommodated. One way of addressing this concern is to establish a number of different minimums. Simple distinctions could be made such as freight and passenger operations. Alternatively, complex ones aimed at differences between similar types of operations could be established. The problem with having multiple levels is that creating more than a very simple set of distinctions diminishes the benefits of this approach. If applicants could no longer be sure what class they would fall into, certainty and administrative ease would be lost. More significantly, governments would have to obtain the expertise to assess which category an applicant fit into. At the highest level, this approach would not be much different than evaluating cases individually.

Ontario's Situation

Evaluating the insurance requirements of shortlines on a case-by-case basis, or setting detailed published levels, would require the Province to either maintain in-house expertise or contract out the task. This would not fit very well with the guiding principles for the *Shortline Railways Act* which are to keep it simple and workable, while minimizing the administrative burden on the Province. In addition, absorbing the cost of expertise would be very difficult given the Province's financial constraints. Passing the cost on to shortlines would not be much better -- the financial resources of some shortlines are also very limited.

Administratively, the universal minimum is much better suited to Ontario's situation. More importantly, circumstances in Ontario are such that it may be possible to use this approach yet still sufficiently accommodate the individual risks posed by each shortline. The Province has conducted research into the range of risks which shortlines operating in Ontario might pose. Findings to date illustrate that while different shortlines present different sets of risks, the absolute risk posed any of them, is relatively constant. Illustrative of this, but by no means definitive on its own, is a recent decision handed down by the National Transportation Agency (NTA) on insurance levels for the Waterloo St. Jacobs Railway. Once fully developed, this shortline company is to operate two very different businesses: a tourist excursion train and a local freight service. While the NTA reviewed the risks posed by each operation separately, it found that adequate and reasonable coverage for each would be $10 million.

The similarity in absolute risk greatly simplifies the Province's task of setting a level that is not excessive for any operation but that also ensures a sufficient level of protection to the public. Available information would appear to indicate that in no case should any shortline be allowed to operate with less than $10 million in insurance. It should be noted that this does not imply that throughout North America, and in Ontario itself, there are not significant differences in the coverages shortlines have purchased -- including some very

limited policies. However, what has been done in the past or is occurring in other jurisdictions is not necessarily what is appropriate for modern day Ontario.

To further enhance the protection offered to the public, the province could insert a clause stating that the shortline must have insurance which is usual and customary for similar operations in Canada but not less than the prescribed limit. This condition would make it clear that a railway's responsibility to obtain adequate insurance is not absolved by the minimum requirement.

DEDUCTIBLES

Views on government responsibility and liability for deductibles, should a shortline become insolvent, tend to be along the same lines as government liability for insufficient insurance --whether or not they are technically liable, they could find themselves covering the shortfall. Unless a government has elected to use the case-by-case approach to setting minimum coverage levels, this is a difficult problem to overcome.

Railway insurance policies typically require the purchaser to have a certain level of deductible. It would, therefore, not be practical to regulate against this practice. Placing a maximum limit on the amount of deductible is also not workable. If the limit was set high enough that companies who have the financial capabilities to cover high deductibles are not penalized, the limits would be so high that they would not protect against the possibility of insolvency from smaller operations.

Realistic maximum deductibles could be determined by reviewing the financial resources of each operator. This, however, has two potential problems. First, a financial review would require expertise and the funding to support it. If case-by-case review is being used to set coverage levels, this extra cost may not be overly significant. Second, while a company may be found to have the means to entertain a high deductible when evaluated, this could rapidly change.

Leaving deductibles as a matter between shortlines and their insurance companies may be a fitting arrangement for Ontario. Reviewing the financial resources of each operator would be very costly should the Province not choose to maintain insurance expertise for the purpose of evaluating cases individually. Justifying the cost would be especially difficult given the fact that the assessment could quickly become outdated, and that the money involved with deductibles is typically a fraction of that involved in total coverage.

SUMMARY

The passing of the *Canada Transportation Act* in early 1996 is expected spur shortline development that remained dormant under the *National Transportation Act.* This regulatory change will coincide with plans by the national railways to improve their financial viability by selling off branch lines -- especially on their eastern networks. Ontario has prepared for the opportunities these changes may bring with the introduction of its *Shortline Railways Act.*

The Province has a fundamental interest in ensuring that shortlines under its *Act* operate with sufficient liability insurance. To do this, it must choose between the costs and benefits of reviewing the needs of each shortline individually or setting published levels. For Ontario, adopting a universal insurance minimum would appear to be the best way to meet the needs of both the Province and the railways. While insurance policies vary greatly throughout North America, and in Ontario itself, research indicates that $10 million would be a reasonable minimum for any operation. While it may be desirable to control deductibles, acquiring the necessary expertise to evaluate them would not likely be cost-effective for the Province.

ENDNOTES

1. Jeff Heinrich, The Gazette, *Small is Beautiful*, (Montreal), July 8, 1995, p.C1.

2. Michael W. Babcock, Marvin Prater, John Morrill, *A Profile of Shortline Railroad Success*, Transportation Journal, Fall 1994, p.22.

3. Jeff Heinrich, The Gazette, *Small is Beautiful*, (Montreal), July 8, 1995, p.C1.

4. Mark Hallman, Financial Post, *U.S. Shortline Rail Firm Eyeing $100 Million Purchase of Canadian Track*, (Toronto), December 2, 1995, p.10.

5. Mark Hallman, Financial Post, *CP Rail Reviews Viability of a Quarter of Its Track*, (Toronto), January 11, 1996, p.8.

6. Mark Hallman, Financial Post, *Eastern Exposure*, (Toronto), November 25,1995, p.20.

7. Mark Hallman, Financial Post, *CN To Cut 1,500 More Jobs*, (Toronto), January 26, 1996, p.1

8. Mark Hallman, Financial Post, *U.S. Shortline Rail Firm Eyeing $100 Million Purchase of Canadian Track*, (Toronto), December 2, 1995, p.10.

9. Jeff Heinrich, The Gazette, *Small is Beautiful*, (Montreal), July 8, 1995, p.C1.

10. Mark Hallman, Financial Post, *U.S. Shortline Rail Firm Eyeing $100 Million Purchase of Canadian Track*, (Toronto), December 2, 1995, p.10.

11. Mark Hallman, Financial Post, *U.S. Shortline Rail Firm Eyeing $100 Million Purchase of Canadian Track*, (Toronto), December 2, 1995, p.10.

NAFTA AND TRUCK CONFIGURATIONS: A FRAMEWORK FOR ASSESSMENTS

by Amiy Varma[1] and Nabil Suleiman[2]

ABSTRACT

Transportation is the nerve center of all commerce and trade. Increased flow of commodities have taken place across the borders of the United States, Canada, and Mexico since North American Free Trade Agreement (NAFTA) went into effect. It is of considerable interest to evaluate the transportation implications of this agreement, whether direct or indirect. One of the key transportation-related concerns is if the establishment of compatible land transport technical and safety standards, and the harmonization of driver certification safety standards, and truck size and weight, will be beneficial to safety and commerce. This paper focuses on the issue of harmonization of truck size and weight and configurations. The paper discusses the issues involved in developing a framework for identifying and assessing the configuration that would be most appropriate in the post-NAFTA environment. The framework should include consideration of trade, economy, carriers' and shippers' costs and productivity, transportation agency costs, safety, pavement damage, bridge impacts, cost recovery, intermodal shifts and arrangements, and other concerns. Such a framework could be used to identify the conflicting concerns and the tradeoffs that may exist.

INTRODUCTION

The North American Free Trade Agreement (NAFTA) has led to considerable increase in cross-border commerce that is unlikely to slow down before the end of the century. Annual volume of foreign trade to Mexico is expected to total $ 100 billion by 2000 (Valdez 1995). Almost half the tariffs on the U.S. and Canadian exports to Mexico were eliminated under NAFTA and most of those remaining will be phased out by the year 2005. Such an increase in commodity flow would relate into a sizeable increase in freight movement, considerable portion of which is going to be carried by trucks. Thus there

1) Assistant Professor, Department of Civil Engineering, North Dakota State University, Fargo, ND 58105
2) Ph.D Student & Graduate Research Assistant, Department of Civil Engineering, North Dakota State University, Fargo, ND

is distinct need to address the issues and concerns related to the tremendous increase in the cross-border truck traffic and related truck movements on the North American highway facilities. There are varied truck configurations in use in the United States, Canada, and Mexico. One of the key transportation-related goal is the establishment of compatible land transport technical and safety standards, and the harmonization of driver certification safety standards, and truck size and weight, which will be beneficial to safety and commerce. Towards this goal, an examination or evaluation of truck configurations from various viewpoints is necessary. A framework for such examination and evaluation would aid in identification of the conflicting and complementing concerns and help highlight the tradeoffs that might exist. This paper disucsses the issues related to the design of such a framework.

BACKGROUND ON NAFTA

The pressures of globalization of economy and competitiveness led to the integration of the European Community countries in 1992. To improve their competitiveness on global scale, the North American countries ventured on a trade agreement, which would lower the tariffs and increase trade and jobs in this part of the hemisphere. Canada, Mexico, and the United States signed the North American Free Trade Agreement, more commonly known as the NAFTA, on December 17, 1992 (USGPO 1993). The U.S. Congress and Senate finally passed it in December of 1993.

Some of the surface transportation provisions of the NAFTA included (Goodwin 1993; USGPO 1993):

* phase-out of mutual restrictions on truck and bus operations between Mexico and the United States;
* Canada imposed no restriction against U.S. and Mexican truckers, but its operators were barred from Mexico;
* creation of a time-table for removal of barriers to the provisions of motor carrier services among the NAFTA countries for the carriage of international cargo and of passengers;
* establishment of a work program for making land transport technical and safety standards for motor carrier and rail operations more compatible;
* assurance that U.S. railroads and intermodal companies will be able to take advantage of gains made through informal agreements with Mexico to market services, operate unit trains, construct and own terminals, and finance rail infrastructure;
* permission for full investment and operating rights to U.S. and Canadian companies in Mexican port facilities; and

* commitment from all three countries to work toward compatible safety standards, with establishment of operational commercial vehicle and vehicle safety standards by 2000.

Some of the concerns raised by American Trucking Association (ATA) from the equity and fairness standpoint were:

* the inequity of trucking investment provisions;
* insuring the competition for small package delivery;
* harmonization of the use of 53-ft trailers in North America;
* harmonization of truck weight standards;
* removal of the barriers surrounding the use of used trucking equipment; and
* enhancement of environmental and labor protections

Some of the concerns raised by Association of American Railroads (AAR) were:

* land harmonization process may result in proposals to increase truck size and weight and in relaxation of ISTEA 1991 freeze on Longer Combination Vehicles (LCVs);
* increased size and weight of trucks would result in considerable intermodal shift, which would lead to more congestion, emissions, and accidents on road network and on border crossings;
* all this can be avoided and the demand for freight movement could still be met by railroads, as they have revitalized their N-S corridor and Mexico has done the same; and
* pattern of trade flows has been wrongly assumed.

Key immediate concerns were inadequate physical and service capacity at border crossings to handle the increase in freight traffic and the institutional and regulatory barriers inhibiting cross-border transportation.

According to the schedule of liberalization agreed in the NAFTA, now the U.S. and Mexico will allow access to each other's border states for the delivery and backhaul of cargo. Also, Mexico now allows Canadian and U.S. investment of up to 49 percent in carriers established in Mexico to transport international cargo and in bus companies. An investment of up to 51 percent will be permitted by 2001 and by 2004 Mexico will permit 100 percent foreign investment foreign investment in truck and bus companies. However, foreign investment will not be allowed in enterprises providing truck service for the movement of domestic cargo either in the U.S. or in Mexico. Mexican carriers are now permitted to establish Mexican-owned or -controlled subsidiaries in the U.S. to transport international cargo from one point in the U.S. to another point. The 53-ft trailers are now allowed in Mexico. Some actions have been taken to improve the physical capacity of the border crossings with increased infrastructure

investment. Some considerable investments have been made by Mexico not only for the border crossing facilities but also for internal highway system. Over $ 300 million was spent on road construction in Mexico (Rhoades 1994). Also, the service capacity of border crossing has increased with different degrees of automation for customs check, which has led to increased productivity of both the carriers and the custom officials (Carr 1995).

But looking ahead five years from now, the most important issue is the establishment of compatible land transport technical and safety standards, and the harmonization of driver certification safety standards, and truck size and weight, which will be beneficial to safety and commerce. This will be increasingly important when U.S. trucks will start traveling on the Mexican and Canadian highways and similarly Mexican and Canadian trucks would start moving on highway facilities in the NAFTA countries. This also happens to be the most complex and controversial topic to deal with. The U.S. truck size and weight regulations have evolved over last 30 years and there is a lot to be learned from the process that was involved and the factors that were considered. The process was driven by a concern for national uniformity, whereby an appropriate match occurs between the size and weight of the vehicles and capacity, alignment, safety, and durability of the highway infrastructures, which are primarily publicly owned. Mechanisms of cost recovery were also used in devising some of these regulations in the United States (FHWA 1995). Recently a comprehensive study was done in Canada looking into impacts of Canada's Heavy Vehicle Weights and Dimension (TAC 1994). However, the considerations for exploring alternatives for harmonizing truck size and weight for the entire North American continent are much more. The complexity would increase as the North American economies become more interdependent because policies set forth to address this concern would directly impact the producers, carriers, consumers, and transportation agencies in all three countries. Thus it is important to investigate the different issues involved in addressing this concern.

IMPACT ON TRUCK TRAFFIC

The truck traffic at various border crossings have increased considerably since the enactment of NAFTA. For example, between 1991 and 1994, there has been almost 55 percent increase in both the empty and loaded truck movements in the Pembina port of entry in Emerson on I-29 forming a border between Manitoba in Canada and North Dakota in the United States. The five major U.S.-Canada border crossings in the west based on dollar values of the commodity flows are Pacific Highway border on I-5 between British Columbia and Washington, Emerson border on I-29 between Manitoba and North

Dakota, Coutts border on I-15 between Alberta and Montana, North Portal border on H-2 between Saskatchwan and North Dakoata, Kingsgate border on H-95 between British Columbia and Idaho (Swanson 1993). The U.S.-Mexico border extends for almost 2000 miles and has several crossings: 6 between California and Baja California; 6 between Arizona and Sonora; 3 between New Mexico and Chihuahua; 5 between Texas and Chihuahua; 4 between Texas and Coahuila; and 9 between Texas and Nuevo Leon and Tamaulipas. Approximately 70 percent of the total United States/Mexican freight trade crosses the Texas/Mexico border (Pezo & Cook 1993). A more recent study (GAO 1996) found that trucks move about 80 percent of the freight transported between the United States and Mexico. In 1995, about 11,000 truck crossings from Mexico into the U.S. occurred daily. Truck traffic from Mexico into the U.S. increased about 27 percent from 1992 through 1995 and about 66 percent of this traffic entered Texas, while about 24 percent entered California and about 10 percent, Arizona. The truck traffic from Mexico into New Mexico is negligible.

However, it is not clear how much of an increase has taken place on the other links of each of the nation's highway network due to NAFTA. It is difficult to clearly identify the increase associated with NAFTA enactment even in the border states. There have been some perceptible increases observed on Texas highway facilities. Examination of the classification count data of North Dakota highway facilities for the period between 1991 to 1994, do indicate some increase in truck traffic. Most perceptible increase in truck traffic in urban environment has been that on interstate and minor arterials. There has been a consistent and significant increase in truck traffic on all highway facilities--interstate, principal arterial, minor arterial, and collectors--in rural environment. The increase in traffic do signal increase in commodity movements and flows by trucks in the state of North Dakota. However, it is not clear whether this has been due to increase in domestic shipments or international shipments associated with NAFTA. One of the reasons it is difficult to ascertain this is because at present trucks are not allowed move on other nation's highway facilities, which will change in another five years. Another point to be made, is that due to NAFTA there would be lot more increase in North-South freight movements. There is every indication that the states that fall in this corridor need to invest time and resources to understand the implications of such movements from different perspectives--nation, agency, carrier, shipper, consumer, and others. Not all these stakeholders would have the same goal or objective and it is important to reconcile these diverse needs, to the extent possible, through an approrpriate harmonization of truck weight and size standards.

TRUCK WEIGHT AND SIZE LIMITS

Truck size and weight regulations play a very important role in a carrier's choice a particular truck configuration. These regulations define the envelope (maximum length, width, and height) and within which trucks are to fit and the maximum weight they are to observe. The U.S. Federal government establishes truck weight law which applies to the Interstate system and vehicle size laws which applies to the designated National Network, which includes the Interstate system. In Canada there are no Federal law and individual provinces and

Table 1: Comparison of Truck Size & Weight Limits

U.S. Federal Limits:
* 20,000 lbs for single axles
* 34,000 lbs for tandem axles
* Use of Bridge Formula B for other axle groups, up to 80000 lb GVW
* 102 inches for vehicle width
* 48 ft (minimum) for semitrailers in a semitrailer combination
* 28 ft (minimum) for trailers in a twin-trailer combination
* Weight law applies to the Interstate System
* Size limits applies to designated network, which includes interstate

---

Canadian Regulations:
* 12,125 lb for steering axle
* 37,479 lb for tandem axles
* 46,297 to 52,911 lb for tridem axle, depending on axle spacing
* Up to 137,500 lb GVW, which vary with truck configuration
* 75.5 ft overall vehicle length
* 53 ft maximum semitrailer length
* Wheelbase, kingpin and overhang limits to control off-tracking
* Minimum interaxle spacing to control bridge loadings

Mexican Regulations on their National Highways:
* 14,330 lb for steering axle
* 42,990 lb for tandem-axle
* 49,604 for tridem axle
* Up to 136,600 GVM, which vary by truck configuration
* 68.2 ft for overall length of semitrailer combinations
* 102,7 ft overall vehicle length for double cargo unit combinations
* Minimum interaxle spacings to control bridge loadings

Source: Adapted from (FHWA, 1995)

territories are responsible for truck size and weight laws. In Mexico, the regulations applicable on national highways are established by Federal government. Table 1 shows the comparative truck weight and size regulations in the U.S., Canada, and Mexico.

TRUCK CONFIGURATION

Several studies have been conducted in the U.S. looking into the feasiblity, effectiveness, and productivity of various truck size and weight limits of different truck configurations (TRB 1986, 1990a, 1990b, 1990c). A detailed study on the feasibility of a nationwide network for Longer Combination Vehicles (LCVs) was pursued by Congress in early and middle 1980s (USGPO 1985). Some recent studies have been devoted on the issue of the designated network for LCVs and the impacts of LCVs (Maze et al. 1994; Stoner et al. 1992; Stoner & Bhatti 1994; GAO 1994).

The network over which Long Combination Vehicles (LCVs) could be used was frozen by Intermodal Surface Trasportation Efficiency Act (ISTEA) of 1991. Since the NAFTA passage, discussion about longer combination vehicles (LCVs) has intensified. Prior to NAFTA agreement, LCVs were used at a limited scale in some Eastern and Northwestern states of the U.S. under the grandfather clause.

Among the different trucks configurations in use in the U.S. are 5 and 6 axles tractor semi-trailer, 5, 7, or 9 axles turnpike double, 7 axles rocky mountain, 7 and 8 axle triple. Trucks used on Canadian highways include 3-6 axles tractor semitrailer, 5-8 axles A Train Double, 5-8 axles B-Train Double, and 5-8 axles C-train double (TAC 1994). The most widely used configuration is the 5-axle tractor semitrailer both on the U.S. and Canadian highways. It is important to note that 80 percent of North American medium and heavy trucks are manufactured and sold in the United States and the remaining 20 percent are split fairly even between Canada and Mexico (TRB 1994). But the populations and growth potential may cause this to shift as all North America develops together. Thus, the U.S. leaders from manufacturing, government, truck carrier industry, and other sectors can play a lead role in devising an appropriate harmonizing of the truck size and weight standards across the North America.

The truck characteristics important for comparison purposes include length, width, height, gross vehicle weight, number of axles, axle weights and spacings, volume, payload, optimal density, articulation, braking system, weight/power ratio, offtracking (low and high-speed), turning radii, hitch type, acceleration and deceleration characteristics, track width, center of gravity height, suspension properties, braking

sideway movement under loaded and unloaded conditions, type of fuel and emission rates, and others. These characteristics affect (a) rate of the infrastructure (pavement and bridge) damage, (b) geometrics, capacity, congestion and safety of the highway and intermodal access, (c) environmental pollution, (c) costs and productivity of the carriers and shippers, (d) efficiency of intermodal transfers, and (e) different aspects of trade and economy.

There have been several advancements in truck design with considerable benefits to the trucking industry and the public at large. There has been approximately 40 percent improvement in fuel economy in the last 12 to 15 years and 70 percent improvement in the ton-miles traveled per unit of fuel primarily because trucks are more aerodynomic (TRB 1994). Today's heavy truck engines represent a 75 percent reduction in particulates, and a 67 percent reduction in nitrous oxide from pre-1988 levels and the future engines are anticipated to be almost smokeless. The installation of three-point shoulder belts and antilock brakes have all contributed to enhancements in safety. The use of belts by drivers have increased and the anitlock brakes minimizes jacknifing and increases the stability of the tractor-trailer.

INFRASTRUCTURE IMPACTS

Various truck related factors affect pavement design and rate of pavement deterioration. Among these factors are changes in tire types, pressures, sizes, and loading characteristics; different axle configurations and loads; truck-pavement interaction, vehicle dynamics and suspension systems. Pavement and productivity benefits could realized by increasing tandem axle loads. Increasing the tandem load to 35,000 lbs within an 80,000 lbs maximum GVW, could reduce the pavement damage caused by a fully loaded 5-axle tractor semitrailer by 10 percent (FHWA 1995). The tridem axle load limits are controlled by Bridge Formula B. Tridem axles could be an effective means to increase truck load capacity while reducing pavement damage. Mexico and Canada permit substantially higher weights on tridems than is permitted by the U.S. Bridge Formula B. It is in this area where some innovative designs can be found and implemented. The maximum dimensions of the ISO container dimensions are compatible with U.S. truck size laws but their weights are not compatible with U.S. weight laws. This is another area which needs to be explored to bring about more efficient standards and harmonization. Tire pressures are not well regulated and seem to have largest impact on the fatigue of flexible highway pavements. Some states in the U.S. do have some laws governing the tire pressures. The impact of suspension

systems on the final load by which pavement and a bridge is impacted is not well known and are subject of research at present.

Bridge design and rate of damage of bridges are highly dependent on the acceptable loading requirements used by the designer. The bridge costs are very much sensitive to the assumptions regarding the acceptable levels of stress in bridges. Majority of states use operating rating approach, while some use inventory rating approach (which is more restrictive). Bridge Formula B is used to prescribe the maximum gross vehicle weight but because of the arbitrary nature of the relationship it uses has been criticized. Bridges on Interstate highways can generally carry higher weights than those allowed under the current formula without being overly stressed. It is also overly restrictive for shorter trucks (dump trucks, trash trucks, construction vehicles) and overly permissive for short six-axle trucks and all trucks with seven or more axles.

Removal of 80,000 lb GVW limit, would lead to overstressing of H-20 Interstate system bridges by 9-axle combination trucks. The basic criteria is the acceptable amount by which you can overstress these and other bridges.

It is important to note that there are various aspects of road and birdge behavior under dynamic loadings of trucks with different configuration that needs to be understood better. As we contemplate harmonization of truck size and weight standards across the North America, we also need to address the harmonization of highway and bridge design criteria, controls, and standards. Mexican highways are not built to the same standards as the U.S. and in some instances Canadian highways are designed to a higher standard than that in the United States.

SAFETY IMPACTS

There are different issues to look at when determining the safety impacts: (a) driver related, (b) vehicle-related, and (c) infrastructure-related.

Lack of driver training, hours of driving at a stretch, and driving under influence of alcohol do jeopardize the safe movement of the truck and well-being of the other traveling public on the highway. Harmonization of driver certification is an important issue in the NAFTA. Mexico has made considerable improvement in this regard. There is still a lot to be done from enforcement standpoint. Driver fatigue remains an important concern and there is a difference among the U.S., Canadian, and Mexico standards in this regard.

Changes is truck size and weight policies can affect characteristics of combination trucks because of changes in vehicle design features such as wheelbase, track width, center of gravity height, suspension properties, and axle weights. These design changes can in turn affect vehicle braking, handling, stability, and maneuverability properties.

The static rollover threshold is the level of lateral (sideward) acceleration that a truck can achieve during turning, without rolling over. Vehicles with low rollover threholds are prone to rolling over when negotiating exit ramps from freeways, when making severe accident avoidance lane change maneuvers, or when they run off the road. The principal determinant of rollover threshold is the ratio of the center of gravity height of the vehicle's mass and cargo to one-half the vehicle's track width. Suspension and tire properties also influence this property but to a lesser degree.

Harmonization of motor carrier safety standards between the U.S. and Canada and the U.S. and Mexico is an ongoing process and was started much before NAFTA. The U.S. and Canada have worked through the U.S.-Canada Motor Carrier Consultative Mechanism established in 1982, and the U.S. and Mexico have cooperated through the U.S.-Mexico Transportation Working Group, established in 1989 (USGPO 1993).

Safety related infrastructure design changes may result if changes in truck size and weight take place. The stopping and passing sight distances, intersection curb radii, the curvature and pavement widening for the interchange ramps and loops, intermodal accesses and the curved highways may have to be altered to accommodate safely the LCVs and other trucks with changes in configurations. The passenger car equivalencies of these truck configurations and the resulting impact on the capacity needs to be determined. There might be need for climbing lanes in certain settings, especially when there is considerable grades. In this regard, the harmonization truck size and weight standards for North America should be accompanied by some changes in facility design standards used in the different countries.

AGENCY COSTS

The highway transportation infrastructure for the most part is publicly owned. In other instances, toll authorities are responsible for the upkeep for the built facilities. These agencies would have to incur costs for the redesigns, or increased level of maintenance, or quicker rehabilitation needs from functional, durability, capacity, congestion alleviation, and safety standpoint. The major issue within the agency is regarding how to improve the cost-effectiveness

of the enforcement of such vehicle and safety standards. A related issue is development of a credible truck traffic database, which could be used to estimate the origin and destination of truck traffic by classification. This is a pressing need in the North Dakota Department of Transportation. Much of the evaluation of current size and weight standards would require such information. Moreover, it is a necessary input for development and effective use of the management systems. The agency is also responsible for evaluating the equity in cost-responsibility among the various trucks using the highway facilities. The evaluation of revenues and truck costs require truck traffic information.

SHIPPER'S PERSPECTIVE

Shippers are most concerned about optimizing their shipments to reduce total costs by balancing inventory carrying costs against transportation costs. The producer, retailers, and wholesalers are very much interested in just-in-time delivery and receipt of goods they purchase or sell. The effectiveness of such logistic would depend to a great extent on the ability of the transportation mode and the associated infrastructure to support that movement. However, the shippers most likely to benefit from changes and harmonization of truck size and weight regulations are those shipping goods that are low-value and non-time-sensitive where transportation costs makes up a significant portion of the total final product cost.

CARRIER'S PERSPECTIVE

The motor carrier buyer exercises a much higher degree of authority and latitude in selecting the type of truck equipment. The buyer may choose from a variety of options pertaining to--frames, axles, springs, brakes, engines, transmissions, instruments, seating, fuel tanks, batteries, hardware and software for the powertrain. The choice of the truck configuration is made by the carrier based on various factors such as service life or durability of the vehicle, efficiency and productivity gains; operating cost, including labor and fuel; infrastructure and safety impacts and the associated cost-responsibility; safety features of the truck; and others. Single most valued quality in any piece of truck equipment is durability and service life. In Mexico, the truckers use a lot older set of truck equipment than are used in the U.S. and Canada. This is cause of concern, especially if this trend continues five years from now. The largest segment of heavy hauling both in the U.S. and Canada is accomplished in five-axle tractor semitrailer and seem to

have a typical service life of 17 years (TRB 1994). Carriers would also try to get a truck equipment which might make the intermodal transfers more efficient.

INTERMODAL SHIFTS

Innovations in manufacturing rely on transportation infrastructure. In each mode carriers are working to establish new services and linkages to meet increasing demand for the movement of goods. Intermodalism is on the rise: trucking companies and major railroads are forming new alliances; airports are negotiating with land carriers to reach new markets that do not have access to air service; and railroads and barge companies are cooperating to provide seamless grain transportation from the Midwest to Mexico. The changes in truck size and weight standards would result in intermodal shift from rail to truck for certain freight movement. This may result because of the "flexibility" of the motor carrier. The evaluation should incorporate the impacts that might arise due to this diversion . The primary interest is on impact it may have on roads, intersections, and interchanges that may not be used by the trucks in the absence of such diversions.

FRAMEWORK

Examination of current or new truck configurations and size and weight standards is multifaceted. There are various stakeholders that need to be identified and whose input and perspective need to be incorporated in the examination. Such stakeholders could be the manufacturers of the trucking equipment; other manufactures, producers, retailers, and consumers who generate need for truck movements; traveling public; transportation agency; motor carriers; enforcing personnel and agencies; and others.

A thorough examination of truck size and weight standards and move towards any kind of harmonization should consider all of the following goals and objectives:

* conserve energy
* protect environment
* recover cost
* maintain and enhance safety
* optimize system performance
* promote innovative strategies
* effect uniformity
* facilitate trade
* ensure compliance
* enhance logistics
* reduce paperwork by streamlining procedures

* protect and preserve infrastructure (highways and bridges)
* minimize expenditures
* enhance competitiveness
* reduce truck costs, including cost responsibility
* foster equity across countries and modes of transportation
* expand markets for trucks and trailers
* maintain equity among the for-hire trucks, private trucks, intercity buses, and governmental trucks and buses
* improve efficiency in transportation and motor carrier fleet design and operation

It is evident that not all these goals are mutually supportive. Added to this are the differing political and cultural realities of the three countries which might affect decision-making on such policy issues. A knowledge-base needs to be developed about data requirements; and acceptable performance standards for vehicle, driver, safety, infrastructure, enforcement, and intermodal transfers. The relationship between truck configurations and highway and bridge designs, including safety elements should be better understood. Truck traffic data regarding origin and destination volumes, volumes by classification types, carrier types, trade corridors and routes has to be developed. In many instances, deficiencies in this regard would reflect in the various approximations and uncertainties that might be inherent in the examination and evaluations.

LIST OF REFERENCES

Carr, R.B. (March 1995). "Don't Let Borders be Barriers," Transportation & Distribution, pp. 65-67.

FHWA (1995). Comprehensive Truck Size and Weight Study-Summary Report for Phase 1--Synthesis of Truck Size and Weight (TS&W) Studies and Issues, Federal Highway Administration (FHWA), WAshington, D.C.

GAO (1994). Longer Combination Trucks: Potential Infrastructure Impacts, Productivity Benefits, and Safety Concerns, United States General Accounting Office (GAO), Washington, D.C.

GAO (1996). Commercial Trucking: Safety and Infrastructure Issues Under the North American Free Trade Agreement, RCED-96-61, United States General Accounting Office (GAO), Washington, D.C.

Goodwin, S.R. (1993). "Implications for Transportation of the North American Free Trade Agreement," TR News, No. 164, pp. 6-10, TRB, Washington D.C.

Maze,Tom, Walter, Clyde K. and Smadi, Ayman (1994). Policy Issues of an Iowa Longer Combination Vehicle Network, Midwest Transportation Center, Iowa State University, Ames, Iowa.

Pezo,R.F., and Cook, G. (1993). Impacts of Increased Trade on Highway Safety Along the Texas/Mexico Boarder Region, Research Project 1984-1F, Center for Geotechnical & Highway Materials Research, The University of Texas at El Paso.

Rhoades, M.P. (1994). "NAFTA's Implications for the Transportation Industry," Transportation Quarterly, Vol. 48, No. 2, pp. 135-148.

Stoner, J., Bhatti, M. (1994). Estimating Pavement Damage from Longer and Heavier Combination Vehicles. Prepared by the University of Iowa Public Policy Center in conjunction with the Midwest Transportation Center, Ames, Iowa.

Stoner, J., Bhatti, M., and Soster, N. (1992). The Economic, Operating, and Infrastructure Impacts of Concentrated Truck Transport Service and Designated Commercial Highway Networks. Prepared by the University of Iowa Public Policy Center in conjunction with the Midwest Transportation Center, Ames, Iowa.

Swanson, L (1993). "Rocky Mountain Trade Corridor: Implications for Transportation Planning in Montana and the Rocky Mountain West," Montana Business Quarterly, Spring 1993.

Transportation Association of Canada (TAC) (1994). Impacts of Canada's Heavy Vehicle Weights and Dimensions Research and Interprovencial Agreement. 2323 St Laurent Blvd., Ottawa, Canada K1G4K6.

TRB (1994). Conference Proceedings of International Symposium on Motor Carrier Transportation at Williamsburg, Virginia, TRB, Washington, D.C.

TRB (1990a). Large Trucks, Special Report No. 223, Transportation Research Board, National Research Council, Washington, D.C.

TRB (1990b). New Trucks for Greater Productivity and Less Road Wear: An Evaluation of the Turner Proposal, Special Report No. 227, Transportation Research Board, National Research Council, Washington, D.C.

TRB (1990c). Truck Weight Limits: Issues and Options. Special Report No. 225, Transportation Research Board, National Research Council, Washington, D.C.

TRB (1986). Twin Trailer Trucks. Special Report No. 211, Transportation Research Board, National Research Council, Washington, D.C.

USGPO (1993). Surface Transportation Implications of NAFTA. Hearing of the Committee on Commerce, Science, and Transportation, US Senate, One Hundred Third Congress, U.S. Government Printing Office (USGPO), Washington, D.C.

USGPO (1985). The Feasibility of a Nationwide Network for Longer Combination Vehicles, Report of the Secretary of Transportation to the United States Congress Pursuant to Sections 138 and 415 Surface Transportation Assistance Act of 1982, FHWA, Washington, D.C.

Valdez, D. (January 1995). "Switch to Free Trade," Transportation & Distribution, pp. 53-57.

Identifying Needed Transportation Investment, by Corridor, to Support NAFTA Trade Flows in Washington State

Ken A. Eriksen and Dr. Ken L. Casavant*

Introduction

Washington State provides critical gateways and corridors for North American trade between Canada and the United States. In 1992, Western U.S.-Canada commercial truck traffic made 848,000 or 53.0 percent of the 2.4 million cross-border trips at Washington-British Columbia ports of entries.[1] Canadian commodities represent 28.3 percent of the 90.1 million tons of commodities on Washington highways, where in 1994, over 1.8 million truck trips transported nearly 25.5 million tons of Canadian commodities valued at approximately $33.3 billion.[2,3] Four corridors, three north-south: Interstate 5; U.S. 97; and U.S. 395, and one east-west: Interstate 90, are used for the transportation of Canadian commodities on Washington highways.

On January 1, 1994, Canada, Mexico and the United States enacted the North American Free Trade Agreement (NAFTA) creating the world's largest free trade agreement encompassing more than 376 million consumers and $7 trillion in gross domestic product.[4] The NAFTA purposes to create a free trade environment where producers, processors and transportation firms move goods safely, timely and cost effectively in a continuous flow, without complications or delays, from a packing house or processing plant directly to the buyer. The NAFTA is an extension of the Canada-United States Free Trade Agreement (CUSTA) which was enacted on January 1, 1989. The CUSTA's goals were to reduce and remove trade barriers, in particular tariffs, by 1999, to increase trade between Canada and the United States. The CUSTA was subsumed into, and superseded by the NAFTA. The NAFTA expands the CUSTA objectives by adding protection for intellectual property rights, ensuring land transportation market access, increasing investment opportunities and upholding environmental standards and agreements while promoting sustainable development.[5]

Free trade agreements attempt to let international markets perform on a level playing field between trading partners. Under a free trade agreement,

---

*Transportation Economist and Professor, respectively, in the Department of Agricultural Economics, Washington State University

barriers that inhibit trade and keep out competition are removed. Commonly used trade barriers between countries are tariffs and non-tariff barriers (quotas, sanitary and phytosanitary requirements, licences, and etc.) on imported goods. As free trade agreements are enacted and implemented, the type and flow of trade changes as firms pursue new markets, and manufacturing and logistics strategies under cost competitive advantages. Trade flows are dependent upon the transportation sector to transport commodities safely, timely and cost effectively from manufacturer, processor and producer to consumer. However, problems exist when transportation options and corridors are inadequate to support increased trade volumes, or price signals do not reflect all costs, which then impedes the movement of trade and decreases the economic returns to production and/or the utility of the consumer.

This paper evaluates free trade agreements, identifies NAFTA trade flows, and the needed highway transportation investment, by corridor, to support those NAFTA trade flows in Washington. The focal point of the paper will be the direct analysis of the impacts on the transportation infrastructure and the need of industry capacity to achieve the requirements of free trade. Analysis will be accomplished by identifying the specific North American, in particular Canada's, commodity flows into, out of and through Washington, by Standard Industrial Classification (SIC) codes. Highway damage coefficients will be used to determine corridor investment requirements to sustain industry capacity utilization of the infrastructure. Potential policies and implementation measures conclude the paper.

Free Trade and the NAFTA

Free Trade

Free trade agreements result when countries purpose to eliminate trade barriers between each other.[5] Barriers to free trade include tariff and non-tariff barriers. A tariff levies a tax on imports (goods and services) which raises the price of a good sold in the domestic market and generates government revenues for the importing country. Non-tariff barriers include import quotas (a maximum limit of goods entering a domestic market); tariff-rate quota (a low duty on a set number of imports, where imports over the set limit are levied an increased tariff); import license (controlling the number of licenses which allow imports); and other non-tariff barriers (sanitary/phytosanitary regulations, nationalism, labeling requirements and export subsidies).[5]

Free trade implies that a country will allow open access to foreign competition in the domestic goods and services markets. Tariff reduction and removal decrease the price of imports by the amount of the tariff; thus, competition in the domestic market increases as foreign competitors enter the market. Foreign competitors export goods and services when a greater comparative cost advantage in a competing country exists, and will import

goods and services when the comparative cost advantage is less than the competing country. Increased competition from foreign imports will cause inefficient domestic firms to exit the market, while leaving the most efficient domestic firms to compete with the foreign competition. Under free trade, global competition determines the equilibrium market price and quantity of goods and services available for consumption. The consumer ultimately encounters competitive prices and better products. Such competition, based on complete resource reflective price signals, contributes directly to the sustainability of production, as well as the sustainability of the transportation function.

The NAFTA

The goal of the CUSTA was to remove all tariffs on all commodities by 1999. In 1990, one year after the enactment of the CUSTA, Mexico and the United States established discussions for a bilateral free trade agreement. By early 1991 Canada joined the talks to form a trilateral agreement. Two and one half years later, on January 1, 1994, the three countries assimilated the CUSTA into the NAFTA.

The NAFTA goals are to create an expanded and secure market for goods and services produced in the NAFTA region, reduce distortions to trade, ensure a predictable commercial framework for business planning and investment, enhance the competitiveness of firms in the NAFTA region to the global market, and promote sustainable development.[5] Canada, Mexico and the United States purpose to meet the goals of the NAFTA through six objectives:

- Eliminate barriers to trade in, and facilitate the cross border movement of, goods and services between the territories of Canada, Mexico and the U.S.;
- Promote conditions of fair competition in the free trade area;
- Increase substantially investment opportunities in their territories;
- Provide adequate and effective protection and enforcement of intellectual property rights in each Party's territory;
- Create effective procedures for the implementation and application of the NAFTA, and for its joint administration and the resolution of disputes; and
- Establish a framework for further trilateral, regional and multilateral cooperation to expand and enhance the benefits of the NAFTA.[6]

While the CUSTA was subsumed into and is superseded by the NAFTA, Canada and the United States use their original tariff reduction schedule as outlined in the CUSTA. Mexico and the United States will eliminate tariffs immediately or during a phase out period of up to 15 years for

some commodities. The NAFTA also addresses market access, trade rules, services, investments, intellectual property, and trade remedies.

Data Analysis Procedure

Identifying international trade corridors requires extensive data including: commodity type, volume and value; origin and destination information (state, city and facility type); and highways or routes utilized for transporting international commodities. However, difficulty obtaining such corridor information exists for highways due to the large number of trucking carriers and the numerous potential origins and destinations. Freight movement by rail and water can be tracked adequately through the Interstate Commerce Commission (now the Surface Transportation Board) waybill samples, the Corps of Engineers Waterborne Commerce data and other published sources.

Eastern Washington Intermodal Transportation Study. To identify international commodity corridors in Washington, an extensive state-wide truck survey was utilized. The Washington State Department of Transportation (WSDOT) initiated a state-wide truck survey as an element of the Eastern Washington Intermodal Transportation Study (EWITS). EWITS is a six-year Intermodal Surface Transportation Efficiency Act of 1991 (ISTEA) planning study to define the multimodal network necessary for the efficient movement of freight and people throughout Washington's eastern region. The WSDOT provided supplemental funding to expand EWITS to include the entire state.[3]

The Washington study, the first in the United States, collected state-wide freight truck origin and destination data through direct personal interviews of more than 28,000 truck drivers. To get a representative view of commodities on Washington's highways, 30 interview sites were selected throughout Washington (21 Washington State Patrol weigh stations, 3 Canadian border locations and the Oregon Port of Entry at Umatilla, some were different directions at the same location). Personal interviews were conducted four times over a one-year period, beginning in the Summer of 1993 and concluding in the Spring of 1994, to reflect seasonal differences in freight movements. Interviews at each location were conducted on Wednesdays over a 24-hour period to obtain median weekly truck flows for a complete one day truck movement profile. Interview questions included origin and destination information (city, state/province and facility type: trucking yard; railroad yard; river or ocean port; airport; factory; processing plant or mill; warehouse; farm or forest; retail store or gas station; and job or construction site) by Standard Industrial Classification (SIC) code, volume, value, truck and trailer configuration and number of axles, and route used while in Washington.[3]

Managing the survey data was accomplished using a computer software database, i.e, Borland's Paradox. The database software made data entry

efficient and performing data query operations of specific data characteristics on the entire data set flexible. For example, specific state and province origin and destination commodity data could be extracted to reflect cargo payload weight, value and route used in Washington. Data query results were exported to a spreadsheet, i.e., QuattroPro, for origin and destination analysis.

Corridor Investment Requirements. Corridor investment requirements are annual dollar estimates, based on current Canadian commodity volumes, to maintain highways from vehicle wear and tear. To determine corridor investment requirements, each truck shipment route was identified and distance in miles measured, by corridor and entire shipment route while in Washington. The miles were multiplied by commodity tons being transported, including the weight of the truck, assumed to weigh 12 tons, to arrive at a tonmile estimate for Canadian commodities transported on Washington highways.

The tonmile estimates were then multiplied by a highway damage coefficient of $.071 per tonmile for state and U.S. highways. The resulting value is the estimated highway damage associated with Canadian commodities on Washington highways, and therefore, investment needed to maintain highway service for those Canadian commodities. Highway damage coefficients were obtained from a previous research study by Ken Casavant and Jerry Lenzi titled "Procedure for Predicting and Estimating the Impact of Rail Line Abandonments on Washington Roads," which was updated and calibrated for 1995 by Denver Tolliver and Casavant for the WSDOT.[7,8]

North American Trade

In 1993, the NAFTA region had a population greater than 376 million and a combined gross domestic product of nearly $7.2 trillion in 1994.[4] However, the region has a negative global trade balance totaling $12.8 billion where in 1994 exports totaled $61.8 billion and imports nearly $74.7 billion dollars.[9] Trade among the NAFTA partners in 1994 totaled $347.6 billion, an increase of 17.2 percent between 1993 and 1994. Trade between Canada and the United States totaled $243.3 billion, 70.0 percent of the NAFTA trade, and between Mexico and the United States $100.3 billion or 28.9 percent, and between Mexico and Canada, $4.0 billion or 1.1 percent in 1994, as shown in Table 1.[10]

The level of trade among the NAFTA countries experienced substantial growth the year after the implementation of the NAFTA. The second phase of the NAFTA tariff rates reductions occurred on January 1, 1995 and the majority of the CUSTA's tariffs have been removed. NAFTA trade, by all accounts, will experience a minor set back given Mexico's looming peso devaluation in December 1994 through 1996. However, the U.S. Department of Commerce, NAFTA Office, expects 1995 NAFTA trade levels to be greater

Table 1. Population, gross domestic product, exports and imports to the world and the NAFTA region for Canada, Mexico and the United States, 1994.

	Canada	Mexico	U.S.	NAFTA Region
Population (thousands, 1993)[4]	28,782	90,029	258,000	376,811
GDP (US $millions)[4]	$574.3	$344.1	$6,275.9	$7,194.3
Exports (US $millions)[9]	$13,817	$5,292	$42,701	$61,810
Imports (US $millions)[9]	$12,319	$7,021	$55,319	$74,659
Trade Balance(US $millions)[9]	$1,498	($1,729)	($12,618)	($12,849)
NAFTA Trade (US $billions)[10]				
Canada (exports)	--	$0.7	$128.9	$129.6
Mexico (exports)	$3.3	--	$49.5	$52.8
Untied States (exports)	$114.4	$50.8	--	$165.2
Total trade imports	$117.7	$51.5	$178.4	$347.6

than 1993 levels (pre-NAFTA) but slightly less than 1994 levels.[10] Long term prospects ensure that trade among Canada, Mexico and the U.S. will expand dramatically.

Washington State's Transportation System

Washington's integrated transportation system provides a competitive and complementary multimodal environment to move people, goods and services. Highway, rail, water, and air transportation modes form the physical transportation infrastructure.

The Transportation Infrastructure

Highway Transportation. Washington highways, the focus of this initial analysis, function as the backbone for the transportation infrastructure system spanning, more than 7,000 miles (773 miles of Interstate highways, 794 miles of non-interstate urban highways and 5,481 miles of non-interstate rural highways).[11] The highway system interconnects major population centers, sea ports, airports, public and private intermodal facilities, interstate and interregional travel, and international borders.

Twelve international ports of entry allow travel between the United States and Canada on the Washington-British Columbia border. Four ports of entry are located in Western Washington and 8 in Eastern Washington. In 1992, 87.3 percent and 12.7 percent of all cross-border commercial traffic along the Washington-British Columbia border, crossed in Western and Eastern Washington respectively.[1]

Rail Transportation. In 1994 Washington's rail system, a complement and competitor to the highway system, was made up of 13 common carriers (10 line haul carriers ranging in size from short lines to Class I, one nonoperating rail line owner, and 2 switching/terminal companies) and several noncommon carriers (lines owned by timber companies and the U.S. government who move freight within the state). Total rail trackage amounts to 2,992 miles where Burlington Northern owns 67 percent or 2,005 route miles and Union Pacific owns 12 percent or 359 route miles, while short line railroads own 20 percent or 598 route miles and switching/terminal railroads own 1 percent or 30 route miles of trackage.[11]

The Washington freight rail system performs an important economic function transporting farm products, lumber/wood products, food/kindred, metallic ores and containers/trailers on flat cars to and from production, processing, distribution and intermodal (deep water marine ports, river ports, road terminals, and shipper connections at sidings, elevators, etc) facilities throughout the state. The river and marine ports depend upon on a highly efficient rail system to transport international cargo to destinations outside Washington to compete with other West Coast ports (Portland, Oregon; Los Angels/Long Beach, California, Vancouver, British Columbia). Washington's rail system has connections to destinations outside Washington including other states and Canada, at Blaine, Clarkston, Pullman, Spokane, Sumas, Vancouver, Wallula, and Wishram, Washington.

Marine Transportation. Washington's marine transport system includes more than 21 river ports on the Snake and Columbia river systems, and 15 deep draft marine ports on the Puget Sound, Lower Columbia and the Pacific coast. The ports provide access to world trade markets and compete for waterborne cargo with ports in Vancouver B.C., Oregon and California. Washington's deep draft ports handled more than 109.7 million tons of waterborne cargo in 1992 and the river ports nearly 7.0 million tons in 1993.[12] Washington's waterborne commerce was valued at $57 billion in 1990.[11]

Washington's ports rely on good highways, adequate rail and waterside access, and intermodal facilities to maintain a competitive transportation advantage for world trade. Washington ports generate approximately 50 percent of the state's total rail traffic. The Puget Sound ports have a competitive advantage delivering goods from Pacific Rim countries to the Midwest saving an estimated two-hours by truck.[11]

Study Findings

Commodities on Washington highways totaled 90.1 million tons in 1994 and Canadian commodities accounted for 28.3 percent of those tons.[2] Canadian

shipments on Washington highways were transported by more than 1.8 million truck trips, hauling nearly 25.5 million tons of Canadian commodities valued at approximately $33.3 billion.[3] Three movement characteristics exist for the transportation of Canadian commodities on Washington highways: trips destined for Washington with Canadian origins (imports); trips destined for Canada with Washington origins (exports); and Canadian movements that transit Washington where the movement either originates in Canada and has a destination outside Washington (southbound transit) or the movement has a destination in Canada from an origin outside Washington (northbound transit). Determination of commodity trips, tonnage, and value were established by corridor origin or destination, and transit movements by corridor(s) utilized. Transit totals by corridor may reflect double counting. For example, if a transit shipment entered Washington in the U.S. 395 corridor and exited Washington into Canada through the I-5 corridor, then its characteristics, tonnage and value, would be included in the U.S. 97 and I-5 corridor analysis as well. Investment requirements were determined by actual tonmiles in corridor of use, thus accurately reflecting corridor highway needs.

Canadian movements that transit Washington occur most frequently; in 1994, 861,000 truck trips transported 15.3 million tons, or 60.0 percent of the Canadian commodities on Washington highways. Movements that originated in Washington (exports) and destined for Canada totaled more than 6.4 million tons or 25.1 percent of the total Canadian tons and conversely, movements that were destined for Washington (imports) with Canada origins were about 3.8 million tons or 14.9 percent as shown in Table 2.[3]

Highway needs to sustain existing highway service levels for Canadian commodity flows in Washington, are estimated to be $736.7 million annually. Canadian commodities transiting Washington require $505.5 million or 68.6 percent of the state total, while movements originating in Washington require $136.8 million or 18.6 percent and those movements destined for Washington $94.3 million or 12.8 percent (see Table 2).

Investment requirements for Washington highways to support Canadian commodity flows are substantial and vary by movement type. Yet the incidence of impact does vary within Washington State. To identify the variation of impact on Washington's highways, four transportation corridors, three north-south: the Interstate 5; U.S. 97; and U.S. 395 corridors, and one east-west, the Interstate 90 corridor, have been identified as forming the major infrastructure for which Canadian commodities move into, out of and through Washington State (see Figure 1).

Interstate 5 Corridor. The Interstate 5 (I-5) corridor begins at Blaine, Washington, on the U.S.-Canada border on the north and proceeds 276 miles

Table 2. Canadian shipments, by percent of tons, truck trips (including empty truck shipments), commodity value, and highway needs, on Washington highways, by movement type, 1994.[3]

Characteristic	Transit (percent)	Origin (percent)	Destination (percent)	Total
Commodity Tons (millions)	60.0	25.1	14.9	25.5
Trips (thousands)	39.6	45.8	14.5	2,370
Commodity Value ($billions)	58.6	31.5	9.9	$33.3
Highway Needs ($millions)	68.6	18.6	12.8	$736.7

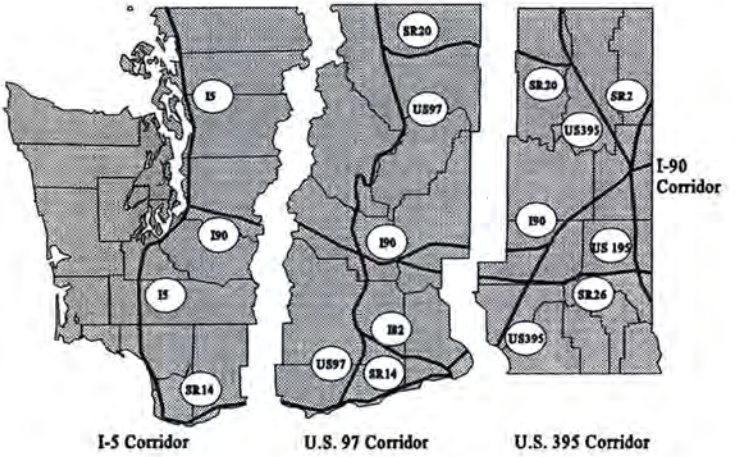


Figure 1. Major highway corridors that support the transportation of Canadian commodities on Washington highways.

south to the Washington-Oregon border at Vancouver, Washington. The I-5 corridor has a population of nearly 4.1 million or 77.8 percent of the state population and employs 1.7 million or 78.7 percent of the state work force.[13]

The distribution of Canadian commodities being transported on the I-5 corridor vary by movement type. For instance, 29 commodity groups have origins in the I-5 corridor, and 5 groups: food and kindred products; general freight; chemicals and allied products; paper and allied products; and wood and lumber products accounted for 57.7 percent of the commodities destined for Canada; 23 commodities were destined to the I-5 corridor, and 3 groups: wood and lumber products; food and kindred products; and chemicals and allied products accounted for 58.4 percent of the commodities with Canada origins.

Canadian commodities using I-5 corridor highways were valued at $30.1 billion and exceed 21.1 million tons, 83.1 percent of the state total. Those commodities originating in the I-5 corridor totaled 5.1 million tons or 24.1 percent of the corridor total and those destined to the corridor weighed 2.6 million tons or 12.5 percent of the corridor total. Commodities transiting Washington and using the corridor totaled 13.4 million tons or 63.4 percent of the corridor total. Nearly 81.0 percent of the transit loads used Interstate 5 exclusively from the Washington-British Columbia border to the Washington-Oregon border.

Investment requirements to support Canadian commodities on the I-5 corridor total $554.3 million, 75.2 percent of the total state requirements. Commodities transiting Washington through the corridor require $390.6 million or 70.5 percent of the corridor requirements, commodities originating in the corridor $98.7 million or 17.8 percent and commodities destined for corridor $65.0 million or 11.7 percent (see Table 3).

U.S. 97. The U.S. 97 corridor stretches 324 miles from the Canadian border on the north to the Oregon border on the south, effectively serving as a major corridor through Central Washington. The corridor population of 546,600 represents 10.4 percent of Washington's population and employs 233,671 people or 10.6 percent of the state total.[13] Agriculture production and fruit and vegetable processing are major contributors to the economic base of the corridor.

Transportation of Canadian commodities reflects largely the agriculture base of the corridor. Three commodity groups: crops; food and kindred products; and lumber and wood products, accounted for 93.9 percent of the commodities originating in the U.S. 97 corridor. Commodities with destinations to the corridor include: chemical and allied products (predominately agricultural chemicals); lumber and wood products; livestock; and crops, accounting for 71.9 percent.

In 1994, 310,062 trucks transported more than 5.3 million tons, 21.0

Table 3. Canadian shipments in the Interstate 5 corridor by movement type, number of trips, commodity tons, commodity value, average tons and highway needs.

	Movement Type			Corridor Total
	Origin	Destination	Transit*	
Number of Trips (thousands)**	463.4	261.8	699.4	1,424.7
Commodity Tons (millions)	5.1	2.6	13.4	21.1
Commodity Value (millions)	$9,610.7	$2,701.4	$17,836.6	$30,148.7
Average Commodity Tons	13.1	14.8	19.2	14.9
Highway Needs (millions)	$98.7	$65.0	$390.6	$554.3

*Transit movements reflect total trips, tons and value that actually used the corridor, while road needs reflect actual tonmiles used in corridor.
** Includes empties

percent of the total amount of Canadian commodities in Washington, in the corridor. Commodities that transited Washington and used the U.S. 97 corridor made up 67.0 percent of the movements, those with corridor origins 18.9 percent and with corridor destinations 14.1 percent

Investment requirements to support Canadian trade flows in the U.S. 97 corridor totaled more than $113.3 million, 15.4 percent of the state total. Commodities transiting Washington through the corridor required $66.0 million or 58.3 percent, commodities with corridor origins $27.7 million or 24.4 percent and commodities destined for the corridor $19.5 million or 16.2 percent (see Table 4).

U.S. 395. The U.S. 395 corridor traverses 281 miles through the eastern portions of Washington, from the Canadian border on the north to the Oregon border on the south. A population of 614,500 and an employment base of 234,302 live and work in the corridor.[13]

Three commodity groups, crops; stone, clay, and glass products; and wood and lumber products make up 68.9 percent of the 15 commodity groups originating in the corridor. Fourteen commodities were destined for the corridor with wood and lumber products; livestock; and chemicals and allied products making up 86.2 percent.

Canadian commodities in the U.S. 395 corridor represented 14.6 percent or 3.7 million tons of the Canadian commodities on Washington highways and were valued at more than $5.5 billion. Commodity movements transiting Washington in the corridor were 3.0 million tons or 81.2 percent of the corridor movements, commodities originating from the corridor were 284,291

Table 4. Canadian shipments in the U.S. 97 corridor by movement type, number of trips, commodity tons, commodity value, average tons and highway needs.

	Movement Type			Corridor Total
	Origin	Destination	Transit*	
Number of Trips (thousands)**	65.1	56.0	188.8	310.0
Commodity Tons (thousands)	1,009.4	755.6	3,581.1	5,346.2
Commodity Value (millions)	$561.7	$454.9	$6,148.4	$7,165.1
Average Commodity Tons	16.9	19.5	18.7	17.2
Highway Needs (millions)	$27.7	$19.5	$66.0	$113.2

* Transit movements reflect total trips, tons and value that actually used the corridor, while road needs reflect actual tonmiles used in corridor.
** Includes empties

tons or 7.7 percent and 412,491 or 11.1 percent had destinations to the corridor.

Corridor investment requirements for U.S. 395 total $69.12 million. Shipments that transit Washington through the corridor require 81.2 percent, those originating from the corridor 7.7 percent and those destined to the corridor 11.1 percent (see Table 5).

Table 5. Canadian shipments in the U.S. 395 corridor by movement type, number of trips, commodity tons, commodity value, average tons and highway needs.

	Movement Type			Corridor Total
	Origin	Destination	Transit*	
Number of Trips (thousands)**	18.4	26.4	159.1	203.9
Commodity Tons (thousands)	284.3	412.5	3,016.7	3,713.5
Commodity Value (millions)	$350.8	$164.1	$4,992.5	$5,507.4
Average Commodity Tons	19.4	19.7	19.0	18.2
Highway Needs (millions)	$10.5	$9.7	$48.9	$69.1

*Transit movements reflect total trips, tons and value that actually used the corridor, while road needs reflect actual tonmiles used in corridor.
** Includes empties

Interstate 90. The Interstate 90 (I-90) corridor provides critical interstate, intra-corridor and inter-corridor connections. Interstate 90, approximately 305

miles in Washington, starts in Seattle, the I-5 corridor, and continues east through Ellensburg and Moses Lake, the U.S. 97 corridor and through Spokane, the U.S. 395 corridor, to the Washington-Idaho border. Typical movements on I-90 are long haul, interstate/provincial shipments of British Columbia and Washington with the eastern U.S. and Canada. Other movements use I-90 intermittently from one corridor to another for inter-corridor and intra-corridor movements. In 1994, nearly 195,000 trucks used I-90 to transport more than 3.6 million tons of Canadian commodities valued at $6.1 billion dollars. Highway needs to support Canadian movements using I-90 from Seattle to the Idaho border, total more than $76.0 million.

Summary. Investment requirements of the State of Washington to sustain the transportation of Canadian commodities on Washington highways exceeds $736 million. The I-5 corridor requires 75.2 percent of the highway needs, while the U.S. 97 and U.S. 395 corridors require 15.4 and 9.4 percent respectively, to support Canadian shipments on Washington highways. Corridor requirements vary significantly, whereas corridor requirements by movement type are more similar among the corridors. Canadian transit movements require 68.6 percent of Washington's highway needs supporting Canadian commodity shipments on Washington's highways. When considering movements that have Washington origins or destinations, origin requirements need 18.6 percent of Washington's highway needs while destination movements 12.8 percent (see Table 6).

Table 6. Washington State corridor investment needs by movement type, 1994.

	Origin		Destination		Transit		Total
Corridor	millions	%	millions	%	millions	%	millions
I-5	$98.7	17.8	$65.0	11.7	$390.6	70.5	$554.3
U.S. 97	$27.7	24.4	$19.5	17.3	$66.1	58.3	$113.3
U.S. 395	$10.5	15.2	$9.7	14.0	$48.9	70.8	$69.1
Total	$136.8	18.6	$94.3	12.8	$505.6	68.6	$736.7

Implications

The U.S. has a negative trade balance with Canada of $14.5 billion, and according to the EWITS and assuming that movements originating from Washington are produced in Washington, and likewise, imports from Canada

are produced in Canada, Washington has a positive trade balance of $7.2 billion with Canada. Washington trade with Canada may continue to grow since NAFTA hopes to expand trade throughout North America by reducing and removing barriers to trade. Most barriers are easy to identify, tariffs, quotas, and licensing requirements. However, other barriers exist which impact the movement of trade. An adequate transportation system allows trade between countries to move timely, cost effectively and safely. Highway development and maintenance are important components for the transport and expansion of free trade by serving as the backbone of a multimodal transportation system.

This paper determined highway investment requirements, by corridor, needed to sustain NAFTA trade flows on Washington highways. Four corridors were identified as major routes used for NAFTA trade, particularly U.S.-Canada trade. Annual corridor investment requirements to sustain current trade flows between the U.S. and Canada on Washington highways total $736.7 million. Canadian movements transiting Washington require nearly 70 percent of total highway needs on Washington highways. The I-5 corridor requires 75 percent of the state highway investment requirements, which is substantially higher than the U.S. 97 and U.S. 395 corridors.

Transit movements have little direct economic benefit to Washington, but still require the most substantial investment needs from the State of Washington. Transit movements make up nearly 40 percent of the annual trips of Canadian commodity shipments on Washington highways and haul 60 percent of the Canadian commodity tonnage. Thus, transit loads are heavier and travel more miles than commodity movements with Washington origins or destinations. Commodity shipments with Washington origins are the least heaviest by movement type, making up 45.8 percent of the trips and 25.1 percent of the tonnage.

A preliminary procedure utilizing a state-wide origin and destination study in conjunction with highway damage coefficients to determine international corridor investment requirements was presented in this paper. Transportation planners developing corridor investment strategies to sustain free trade, and the movement of that trade on highways, face challenging issues of how to fund the investment requirements given a disparity of investment requirements among the movement types. Planners will want to pay particular attention to the type of movements, transit; origin; or destination, and the particular commodities which require highway investment. Maintaining an adequate highway transportation system for the movement of goods and services is extremely important for Washington industries exporting to Canada.

References

1. Federal Highway Administration, Department of Transportation. *Descriptive Report of Cross-Border Traffic and Transportation in the Western U.S.-Canada Region.* Washington D.C.: U.S. Department of Commerce National Technical Information Service, September 1993.
2. William R. Gillis, Eric L. Jessup and Kenneth L. Casavant. "Movement of Freight on Washington's Highways: A Statewide Origin and Destination Study." Eastern Washington Intermodal Transportation Study Report #9. Washington State University/Washington State Department of Transportation, November 1995.
3. William R. Gillis and Kenneth L. Casavant. "Washington State Freight Truck Origin and Destination Study: Methods, Procedures, and Data Dictionary," Eastern Washington Intermodal Transportation Study Report #3. Washington State University/Washington State Department of Transportation, December 1994.
4. The International Bank for Reconstruction and Development/World Bank. *World Tables.* Washington D.C., 1995.
5. Cathy J. Schaeffer and Thomas I. Wahl. "The North American Free Trade Agreement: Effects on Washington State Agriculture," IMPACT Center, Washington State University, Information Series Report #67, June 1993.
6. Office of the U.S. Trade Representative. *North American Free Trade Agreement.* Washington D.C., October 1993.
7. Kenneth L. Casavant and Jerry C. Lenzi. "Procedure for Predicting and Estimating the Impact of Rail Line Abandonments on Washington Roads." Washington State Department of Transportation, Olympia, Washington, November 1994.
8. Denver Tolliver, Kevin Andres and Brian Lindamood. "Washington Rail Impact Analysis: Executive Summary." Washington State Department of Transportation, Olympia, Washington, December 1994.
9. Organisation for Economic Cooperation and Development. Statistics Directorate. "Monthly Statistics of Foreign Trade." August 1995.
10. United States Department of Commerce, NAFTA Office. Assorted "NAFTA Facts." 1995 and 1996
11. Washington State Department of Transportation. "Statewide Multimodal Transportation Plan." Olympia, Washington, public review draft, 1994.
12. BST Associates. *1995 Marine Cargo Forecast.* Prepared for Washington Public Ports Association and Washington State Department of Transportation. Seattle, Washington, January 1995.
13. Office of Financial Management. *Washington State Data Book.* Olympia, Washington, 1993.

NAFTA Trade Corridors - The "Unlevel Playing Field" for International Carriers

By R. Ian Kingham, P.Eng.*

I.0 INTRODUCTION

The North American Free Trade Agreement (NAFTA) became effective in 1994 committing Canada, United States and Mexico to a program that would remove trade barriers. Free trade was the focus, with substantial consideration given to the removal of tariffs. Beyond tariffs, it was recognized that if goods were to move more freely, a host of transportation issues need to be addressed. While these issues remain unsettled, international carriers operate in an unlevel playing field with their international competitors.

Earlier, the Free Trade Agreement (FTA) between Canada and the United States became effective (1989). This too addressed tariff removal but did nothing to resolve issues on the movement of goods, such as standardization between the two countries on trucking weights and dimensions. It was assumed that both countries would maintain their current regulations concerning rail, truck, air and marine goods movement and that the transportation industry would accommodate itself to a more international economy. The entry of Mexico into the NAFTA Agreement immediately broadened the transportation issues that need to be addressed, principally those involving safety and the environment. With the geographic context of the trade agreement being substantially increased, issues involving operating authority for truckers and agreements between railroad companies became more important.

The purpose of this paper is to assist planners in understanding the existing barriers to the efficient movement of trucks and automobiles across borders. Information is drawn from the Congressional Study[1], Statistics Canada, The International Road Transport Union (Switzerland)[2] and others.

---

* *President, GMK Transportation Planning and Engineering Ltd., Victoria, B.C. and Port Angeles, Washington.*

2.0 TRADE BETWEEN THE NAFTA COUNTRIES

2.1 The Importance of International Trade to Canada, United States and Mexico

In terms of population, gross domestic product and trade, the United States is the giant of the three Countries as provided in Table 1. Exports, however, expressed as a percentage of gross domestic product, are the least for the three countries showing that the United States is less dependant on international trade than its two partners. Canada, on the other hand, has the largest dependence on international trade. It is apparent from these statistics that the United States has the least to lose and Canada the most to lose in negotiations with its trading partners.

2.2 Existing and Future Trade Flows

The value, in Canadian dollars, of trade flows between the three partners and the rest of the world are shown in Figure No.1[3]. Trade between the United States and Canada is three times that of trade between the United States and Mexico and almost a hundredfold times the trade between Canada and Mexico. These statistics show the relatively small base for trade between Canada and Mexico at the time of NAFTA signing.

Growth in trade between NAFTA Countries between 1988 to 1990 is shown in Figure No.2. Whereas trade between Canada and the United States has remained stable, trade has increased significantly between Mexico and its partners. Exports from Mexico were increasing at a more rapid rate than imports into Mexico.

The growth in trade between United States and Mexico is complex. The percentages are large because of enterprises in Mexico close to the border called Maquiladoras. These manufacturing concerns receive components, 95 percent from the United States, assemble them at border manufacturing sites and export them into the United States. Maquiladora enterprises represent 50 percent of non-oil exports and 20 percent of Mexico's manufacturing.

In more recent years, trade has been increasing at an increasing rate between the United States and Canada. Figure No.3 shows that exports from Canada to the United States have grown at a rate of 15 percent annually from 1990 to 1994[4] compared to 1.4 percent from 1988 to 1990. Whether this results from the Free Trade Agreement, or from a weakening Canadian dollar, is not known.

Table 1

Economic Overview (1993, U.S. $)

Characteristic	Canada	U.S.A.	Mexico
Population (in millions)	28.7	257.9	90.0
GDP (billions $)	532.9	6379.4	302.9
Trade (billions $)	273.7	1045.3	101.3
Exports as a % of GDP	26	7	14

GMK

Source: Ref. 2

Figure 1

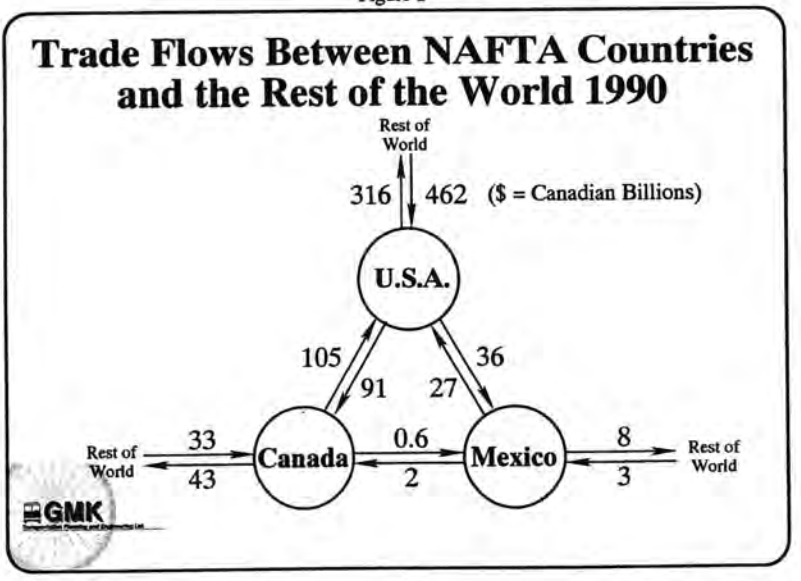
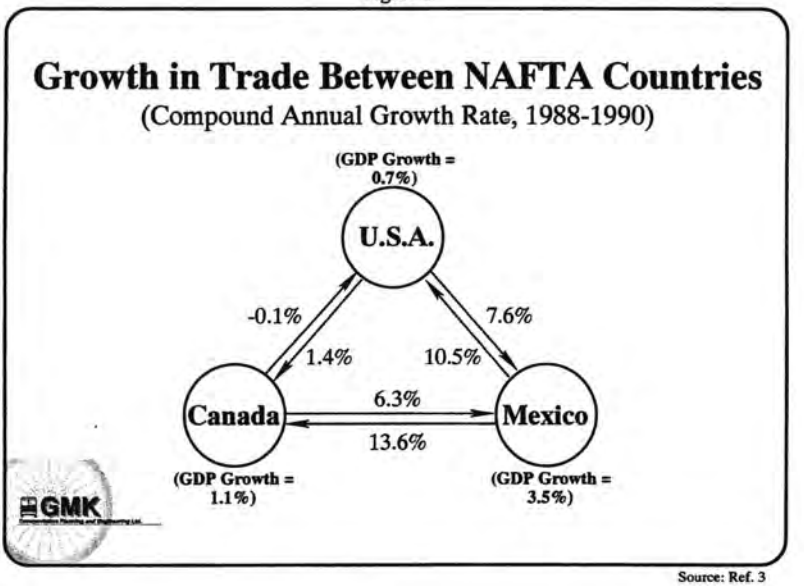
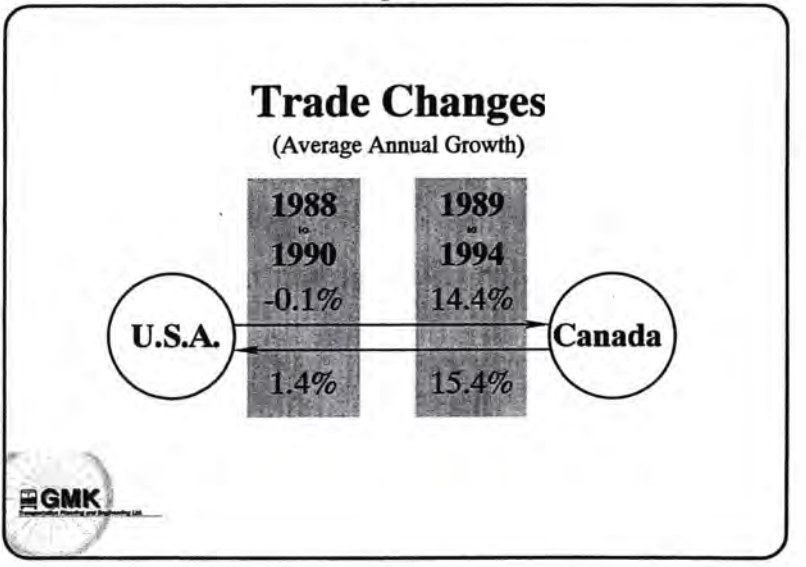


Figure 2



Source: Ref. 3

Figure 3



3.0 HOW GOODS MOVE

Of the five modes to move goods, (rail, trucking, marine, pipeline and air), trucking has the lions share, as shown in Table No. 2. Percentages for each country show that as much as 81 percent of the dollar value of all goods moved, moved by highway. These percentages do not include the movement of oil by pipeline. From the data sources available, a firm figure for rail movement between the United States and Mexico could not be ascertained. It is probably low, as indicated, because of the substantial movement, by truck, of components to the Maquiladora Enterprises.

Further information on truck mode split is provided in Table No. 3. Between Canada and Mexico, 66 percent of Mexican exports to Canada moved by truck whereas 45 percent of exports from Canada to Mexico moved by truck. Canada's exports are machinery and auto parts/automobiles and resource commodities, best transported long distances by rail because of weight. Mexico's exports are also manufactured goods and, in addition, agricultural products that are best transported by truck.

Table 2

How Goods Move ($)

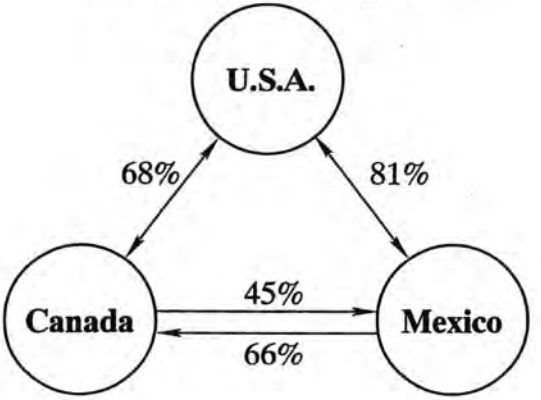
Country	Mode: Highway %	Rail %	Water %	Air %
US/Canada	68	20	2	10
US/Mexico	81	5 *	10	4
Canada/Mexico	60	26	11	3

* The US Congressional Study suggests that this figure could be as high as 12%.

Source: Ref. 1 and 2

Table 3

Truck Mode Split



4.0 TRANSPORTATION BARRIERS TO INTERNATIONAL CARRIERS

4.1 The Cabotage Rule

Although NAFTA is levelling the playing field for commodities, it did nothing for levelling the playing field for truckers in Canada, United States and Mexico. Canada and the United States, but not Mexico, do grant operating authority to foreign truckers. Nevertheless, a foreign trucker cannot pick-up and deliver goods within the host country. This is the well known cabotage rule. A Canadian trucker carrying goods from Canada to Mexico is prohibited, on the return, from picking up goods in the United States and delivering in the United States. For Canadian truckers, this problem is magnified because of Mexico's policies restricting foreign truckers.

Mexico, at the time of NAFTA signing, did not permit any foreign truckers to operate anywhere in Mexico. All goods moving southbound had to be delivered to a Mexican carrier at the border. Goods moving northbound had to be transferred from a Mexican carrier to a United States or Canadian carrier on the United States side of the border. For a Canadian carrier, these goods picked up in the United States could not be delivered in the United States because of the cabotage rule. Recognising this disadvantage to Canadian truckers, Mexico and Canada signed a Memorandum of Agreement in March 1994 that permits Canadian truckers to pick-up and deliver in the six Mexican border states. The Memorandum of Agreement accelerated what would have happened under NAFTA, that permits all foreign truckers into Mexican border states by December, 1995. Mexico also has limitations in foreign ownership in trucking firms. Under NAFTA, Mexico was to permit 49 percent foreign ownership by December 1995. This would not change until January, 2001, when foreign ownership could be increased from 49 to 51 percent. Consequently, trucking between Mexico and its trading partners must be by agreement and cooperation. Multi-national trucking firms, and the potential efficiencies that could result, are not possible until January 2001.

The planned December 1995 changes did not take place. The United States, in return for Mexico permitting United States truckers into its border states, was to permit Mexican truckers into the United States border states. On December 19, 1995, Secretary Pinya unilaterally overruled the NAFTA provision, presumably over concern for damage to roads and bridge by Mexican trucks carrying loads in excess of the United States limits (but legal in Mexico). Mexico has filed an official protest under the provisions of NAFTA.

Other changes called for by NAFTA are travel throughout Mexico by foreign carriers by January, 2000. By January, 2004, all foreign restrictions on ownership will end.

The cabotage rule also has significant impacts on the cruise ship business. In the Alaska cruise business non-United States carriers can operate out of Vancouver and consequently not violate the Rule. This is one factor that limits the ability of the Port of Seattle to become a cruise centre.

Unless the cabotage rule also can be relaxed, Canadian and Mexican truckers will be at a disadvantage to United States truckers. United States truckers have the biggest market to serve and, consequently, the greatest probability of picking up back-loads.

4.2 Truck Weights and Dimensions

Because of the United States action of prohibiting Mexican trucks in United States border states, the issue of very low load restrictions in the United States has shot to the forefront of the unequal playing field. Table 4 provides a comparison of weight and size restrictions in the three countries[5,6]. The United States has the lowest axle and gross vehicle weight restrictions of any country in the world. Although previous studies in the United States[7] have recommended relaxation of these limits, Congress has not acted since the last increase in 1982. These studies showed that benefits to the economy (the trucking industry) far outweigh the increased pavement and bridge costs.

TABLE 4

Comparison of NAFTA Country Vehicle Weights and Measures (kg)

Vehicle Weight and Dimension	United States	Mexico	Canada
Weight: (kg) Steering	-	6,500	7,300
Single	9,080	-	9,100
Tandem	15,430	19,500	17,000
Tridem	-	22,500	24,000 max.
GV (max.)*	36,320	62,000	62,400
Length: (m) Semi-trailer	14.6	16	20
Overall	17 + tractor	23	31

* In Canada and Mexico, eight axle double trailer combinations are permitted the maximum gross loads.

United States' height and size restrictions have a dramatic impact on Canadian and Mexican carriers. They must operate with less pay load or use different equipment. Why the United States differs from the rest of the world comes from both political and technical sources. The political one is preoccupation with a concern that productivity gains in the trucking industry will change the competitive environment between truck and rail such that goods travelling by rail now would be diverted to truck. This in turn could lead to an increase in the number of trucks and consequently a reduction of pavement and bridge life. This concern for pavement wear goes back to the AASHO Road test in the 1950's and the technical reason why load limits in the United States are the lowest in the world.

The AASHO Road test was an accelerated full-scale, outdoor, statistically designed experiment to relate pavement structure thickness with weight and load repetitions. Most of the pavements failed in the two Springs when weather warmed and frost left the ground. Consequently, the thickness/load repetition relationship consisted of two data clusters associated with each Spring. The researchers had difficulty maintaining traffic the first Spring because of rapid failures of grossly inadequate pavement thicknesses. Consequently, the load repetition relationship is suspect and was declared such in the Canadian analysis of the results[8]. For the empirical relationships developed, it was a judgement call on the selection of a curve shape through what was essentially two points. Furthermore, extrapolation was necessary to obtain the thickness/weight/load repetition relationships for highway cost allocation and pvaement design. Consequently, the United States' weight limits, determined on the basis of highway cost allocation, are suspect.

The United States had a huge investment in the experiment to obtain the relationship and consequently, was "under the gun" to do it regardless of the data limitations. This has lead to the weight restrictions of today. Canada and Mexico could be more objective with the AASHO Road Test results and make very different decisions on allowable weights and sizes that have been very beneficial to their economies.

4.3 Border Delays/Clearance Procedures

4.3.1 Trucking

Trucks, unlike rail cars, must be cleared individually, that in turn makes the delay greater per unit of freight carried. Canadian and United States border clearance procedures are constantly improving but there still exists a need to reduce the number of full inspections required to catch offenders.

On both the Canadian and Mexican borders, demonstration projects are being conducted with new technology to identify trucks, their loadings, and drivers. Briefly, a computer chip containing this information in the drivers tractor is read far in advance of the border. This information is transmitted automatically to computers that are accessed by customs and immigration officials at the border. With such a paperless system, most trucks could be waived through the border with minimal delays.

Combining custom's and immigration's responsibilities in one person for routine checks has been a characteristic of the Canadian system for some time. On the other hand, the United States system requires immigration pre-clearance outside of the United States and customs clearance at the point of entry. Changes are being considered in the United States and Canada where one official could act for immigration and customs for both countries. A pilot program is planned for the new international terminal in Vancouver's airport to reduce delays for passengers passing through Canada for the United States.

The environment at the Mexican border is very different. A Canadian truck must stop on the United States' side of the border and wait while a freight forwarding person crosses the border with the paper work and makes the necessary arrangements with Mexican customs to clear the load. With clearance, the truck is driven over the border and again stopped to transfer the trailer to a Mexican tractor. Often this procedure can take a day or more. It is a critical factor in "just-in-time" deliveries. Mexico is also experimenting with paperless clearance procedures.

On both borders, proper identification of the driver as an approved business visitor is important. Research is just beginning on storing all the necessary driver history in the computer chip.

4.3.2 Rail

Rail has enjoyed a competitive advantage over truck. One reason being that border clearance costs per unit of freight can be substantially less. Unit trains can be cleared as one. Customs and immigration deal with fewer people and the potential of violations is considered to be less.

The same electronic processing is being demonstrated for rail. This is expected to provide further efficiency.

4.4 Alliances and Foreign Ownership

A shipment from Canada to Mexico south of the border states, using a Canadian carrier, requires alliances between Mexican, Canadian and likely United States' transporters, whether it be truck or rail. This creates inefficiencies although in the case of rail, alliances have been in place for many years[9]. Canada's railroads have had their wholly owned United States subsidiaries. Railroads have been oriented to north/south movements in recent years. Excellent business relationships exist between United States railroads and the Mexican National Railway. Furthermore, NAFTA permits Canadian and United States railroads to; (1) market their services in Mexico, (2) operate unit trains with their own locomotives, (3) construct and own rail terminals, and (4) finance infrastructure.

The picture for trucking is not as good. Mexico will not permit controlling foreign ownership until the Year 2001. Alliances between trucking firms are a new condition of work for some carriers. Until efficient alliances can be established with Mexican carriers or until Mexico allows foreign carriers to operate in Mexico, rail will have a competitive advantage over trucks.

5.0 "HARMONIZATION" REQUIRED BY NAFTA

The following five harmonization committees were established by NAFTA to focus on problem areas that need standardization between the three countries. These are as follows:

1. Vehicle weights, dimensions, brakes, tires and emissions
2. Driver standards, medical requirements
3. Dangerous goods movement
4. Road signing and traffic control devices
5. Rail safety

Tough issues to resolve exist with the vehicle weights and dimensions as discussed previously. For the agreed on sizes and weights, the Committee will have to develop methods that are common in the three countries for issuing permits for overweight and oversize. Vehicle brakes, tires and emissions are other vehicle elements requiring standardization. Last but not least, this Committee must deal with how to control violations. A report is due from this Committee in January, 1997.

Standards for transporting dangerous goods are common between the United States and Canada. Negotiations will be needed to reach an agreement with Mexico.

6.0 TRI-NATIONAL COMMITTEES

In addition to that required by NAFTA, the three countries have established five tri-national committees to oversee a greatly expanded transportation system. These are as follows:

1. Operations
2. Border facilitation
3. Automated data exchange
4. Science and technology (paperless clearance)
5. Maritime and port operations

The operational group is concerning itself with the operational aspects of transporting people and goods across the two borders, such as application for operating authorities, vehicle registration, fees and taxation, insurance, etc. The border facilitation group is working on improving border clearance procedures, coordination of agencies at the borders, standardization of operations, immigration and temporary entry, customs and the planning of border infrastructure facilities. The automated data exchange group is working on such things as the exchange of information on drivers' records and linking the licensing information system in the three countries. The Science and Technology Group is working on paperless clearance procedures, the use of electronic identification for vehicles and global positioning satellite systems.

7.0 CONCLUSIONS

There is a long way to go to level the playing field for Canadian and Mexican truckers. Rail carriers have had the initial advantage over trucking with few international barriers and a history of working together across borders. Consequently, the current environment favours rail. However, should load limit restrictions be relaxed in the United States, and border clearance improved, trucking will become more competitive with rail. This could result in a greater mode split for trucking.

8.0 REFERENCES

1. Federal Highway Adminstration, USDOT, *Assessment of Border Crossings and Transportation Corridors for North American Trade: A Report to Congress*. Publication No.FHWA-PL-94-009, Washington, DC, 1994

2. LeBrun, S., Nix, F., Jones, J., *The North American Free Trade Agreement (NAFTA) and Road Transport*, International Road Transport Union, Geneva, Switzerland, September, 1994.

3. Braxton Associates, *NAFTA: Prospects for the Canadian Trucking Industry*, Transportation Services Division of External Affairs and International Trade Canada, Ottawa, 1993.

4. Szamosszegi, Andrew, *The US/Canada FTA - A Good Deal A Good Model* Economic Strategic Institute, Washington, D.C., 1996.

5. Federal Highway Administration, United States Department of Transport, *Comprehensive Truck Size and Weight Study: Summary Report for Phase I - Synthesis of Truck and Size (TS + LO) Studies and Issues*, March 1995.

6. Federal Highway Administration, United States Department of Transport, *Comprehensive Truck Size and Weight (TS + W) Study, Phase I Synthesis, Pavements and Truck Size and Weight Regulations*, Working Paper 3, February 1995.

7. "Truck Weight Limits; Issues and Options", *Transportation Research Board Special Report 225*, National Research Council, Washington, D.C., 1990.

8. *Report of the Observer Committee of the Canadian Good Roads Association on the AASHO Road Test*, Canadian Good Roads Association, June 1962.

9. Sauvé, Pierre, *NAFTA and the Railroads: A Canadian Prospective*, Presentation to the Winter Meeting of the Association of American Railroad Superintendents, Kansas City, MO, January 25, 1993.

IN DEFENCE OF A TRANSPORTATION SYSTEMS APPROACH TO TRANSPORTATION POLICY DEVELOPMENT

by Derek Lett*

The opinions expressed are solely those of the author and do not necessarily reflect the official position of the Ministry of Transportation, Ontario or of the Government of Ontario.

This paper advocates the use of a transportation "perspective" as a means of countering the challenges to policy development in the transportation sector. A transportation "perspective" is a broad, long range, multi-modal, strategic overview of the transportation system, in which the issues affecting all modes are addressed in an integrated manner. It is essentially an application of the "systems approach" of transportation planning to the field of transportation policy development.

The Policy Development Problem

Policy development has long been referred to as an inverted pyramid, the top of which contains broad government goals which narrows as it approaches the policy implementation apex. Exhibit 1 illustrates this inverted pyramid as it is applied to the integrated transportation policy and planning approach used in Ontario.

The policy process is filled with hills and valleys which tend to relate to the urgencies of the issues on the social and political agendas. The challenge of governments is to maintain or strive for equilibrium in the application of the policy instruments available to them - taxation, regulations, subsidies - in order to deal effectively with competing demands for limited resources.

In order to achieve that balance in a political, social, economic and technological environment that is rapidly changing, policy makers may be tempted to engage in reactive and piecemeal analysis of individual, modal-specific issues. This challenge is particularly relevant in today's fiscal environment in which debt ridden governments world-wide, must turn to downsizing and privatizing public

---

*Policy Advisor, Ministry of Transportation, Ontario.

INTEGRATED TRANSPORTATION POLICY AND PLANNING

PROVINCE OF ONTARIO

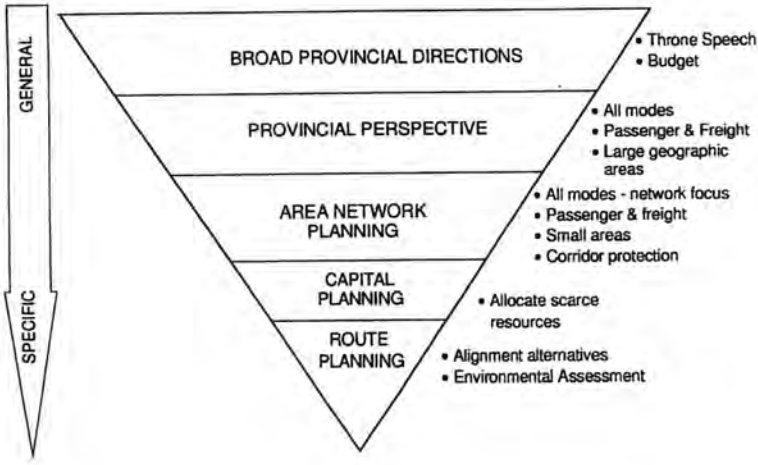


EXHIBIT 1

services. Fiscal imperatives have become much more of a driving force in making policy and investment decisions. The potential dangers associated with this approach are related to treating issues in isolation.

When policy issues are treated in isolation, two types of problems can emerge[1].

1. The same issues have a tendency of recurring. They are difficult to resolve to the satisfaction of all interested parties. Hence, at the policy implementation stage, the problems that arise are most often related to those areas where the policy formulation process has generated the greatest controversy. Therefore, in an era of severely limited financial resources and competing transportation demands, quick fixes to complicated policy issues, by their very nature, will only be temporary.

2. Issues treated in isolation may conflict with one another, generating new issues causing unintended consequences and distortions and hence, a further need for corrective policy development. The resulting patchwork of policies designed to address these now interlinking or interrelated issues reinforces the public perception that government policies, and programs, are uncoordinated and inefficiently administered.

The Systems Approach

The systems approach in transportation planning is used to develop alternative transportation networks for a selected land use plan. These networks ideally, should reflect different policies and take the form of complete systems serving an entire area.

The systems approach and the network choices that result from its use is based upon the careful consideration of such factors as the rate of trip generation, trip length, land-use characteristics, cost, transport design and operational criteria. These factors help determine the critical and alternative networks. Although the approach is applied to highway planning and may vary from study to study, according to Michael J. Bruton[2], the process used is usually as follows:

i. Identify the strategic links in the highway network. These are the links with the best levels of service; provides the best connections between major communities; and minimize costs - travel time, air pollution, energy consumption, and both travel and network construction costs.

ii. Establish the design principles to be applied at the area network planning stage, such as continuity of movement throughout the system.

iii. Establish design standards and criteria for highways such as design speed, capacity, lane width and number of roads to be constructed.

iv. Outline the constraints that are likely to influence the design of alternative highway networks.

v. Produce layout plans of alternative highway networks.

The systems approach has been around since the early 1970's. It emerged at about the same time as large-scale computers and fostered the belief that large scale transportation systems could be formalized, understood and manipulated by models[3]. It was, and to a large extent, continues to be, a diagnostic approach.

In its incipient stage, the approach was applied to the evaluation of transportation systems in mainly urban environments and focused on short-distance passenger traffic. It attempts to identify factors which hamper *optimal* performance of the transportation system by applying such measures as level of service, time, financial and energy costs to problem areas.

It is in its narrowest application, when the approach is used to define transportation problems solely in the context of systems performance, that the systems approach has come under attack since the 1980's. As the perception of the transport problem changed (from level of service to air pollution & energy concerns, to mobility for persons with disabilities and disadvantaged groups), transportation system planning also had to change. The performance-based systems approach is criticized for ineffectively incorporating these socio-economic factors in its modelling and prescriptions.

Those who challenge the systems approach[4] present at least three compelling arguments. They argue that transportation planning and policy-making cannot be divorced from broader socio-political and economic systems. The broader social and economic issues (economic growth, social mobility, transfer of wealth, etc.) are the result of problems external to the transportation system. Yet the solutions sought are restricted to options within the transportation system.

For example, the demand for transportation is a derived demand, that is, it is derived from the demand for products and services which depend on transportation to provide access to markets. According to the critics, the systems approach would define the problem of economic growth in terms of capacity issues and suggest system expansion as a means of fostering more efficient movement, and by extension, greater ease of market access. The problem with the solution option is that its focus is on the supply-side functions. It fails to

examine how the existing, unexpanded system can be made more efficient by transportation demand management such as road pricing, number of lanes available, etc..

It is also argued that the systems approach does not address the issues of equity very well. Accordingly, transportation system planning often fails to recognize and integrate the segmented mobility needs reflected by the inequities between parents and children, men and women, workers and pensioners, the healthy and persons with disabilities. This deficiency is obviously more applicable to passenger transportation issues than to freight. This is because technological improvements, structural change and fewer objectives are believed to be involved in the latter.

The final argument against the systems approach relates to its predictive capacity. Forecasting future behaviour could be based on erroneous or biased information designed to achieve a result that supports or undermines a certain political position. It can then be argued that since behaviour in the real world is not necessarily randomly distributed around the means of objective measures, the decisions which ultimately result from the forecasting exercise could indeed be tainted by subjectivity and bias.

Whereas these criticisms of the systems approach to transportation planning and policy development are quite valid, they do not in themselves require that the framework be entirely discarded. The only area of weakness these points expose in the systems framework is the inadequacy of the approach to properly deal with important socio-economic and political factors. The fact is, the socio-economic data is already included in such measures as origin-destination, travel patterns and demographic statistics. These are collected and included in the computer traffic modelling exercise. However, because the primary focus of that exercise is on the efficient movement of people and goods, the importance of the socio-economic data is often lost. In the field of policy development however, the opportunity exist to give these factors a higher profile.

The transportation "perspective", as described below, represents the cross over of the systems approach to transportation policy development. It is the wider application of the systems approach, as opposed to the narrower application on which the criticisms above are based.

Transportation "Perspectives"

A transportation "perspective" is the integrated planning approach being used in Ontario to identify a strategic transportation network (STN). It combines "needs-

based" social and economic policy objectives with the hard data on which systems modelling is based, thereby addressing the major weakness of the systems approach and producing a comprehensive transportation system overview.

The Ministry of Transportation has recently embarked on an initiative to create a long range transportation plan for the entire province. This plan is being created on a region by region basis and requires a movement away from project-based, incremental planning to large scale master planning. The systems approach, because of its comprehensiveness, is ideal for large scale master planning and for addressing big transportation issues. The larger the issues, the greater the need for a comprehensive approach.

Prior to this initiative, long range transportation planning and policy development in Ontario concentrated mainly on highway construction and maintenance. For example in the mid-1970's, under the auspices of the then Ministry of Treasury, Economics and Intergovernmental Affairs, a series of long-range (20 year time horizons) transportation plans were developed as part of a plan to enhance regional economic development throughout the Province. These plans examined the regional economies and the role of the transportation infrastructure, much like what is currently planned in the perspectives. However, the focus of those exercises was on highway performance and capacity improvements. The role of rail, air and marine transportation only received honourable mention.

By the late 1980's and early 1990's, the focus had changed to integrated planning and policy development. From the "Getting There" project of 1990, whose objective was to develop a new vision for the future of transportation, to the year 2015; to the internal business process review of 1991-92 at the Ministry of Transportation; to current focus on long range integrated planning and policy development, including such projects as transit integration in the Greater Toronto Area; there has been a progression to a more comprehensive approach to understanding, planning and developing transportation policies in the Province of Ontario.

The current regional perspectives and the STN they identify will be used as input into policy decisions about infrastructure investment priorities, policy or regulatory changes and as background for more detailed area, corridor or route planning studies. Four regional perspectives (The Greater Toronto Area, Northern, Southwestern and Eastern Ontario) are planned[5] and will be brought together, like pieces of a jigsaw puzzle, to form a provincial perspective. Exhibit 1 illustrates where and how the provincial perspective fits into the policy and planning framework in Ontario.

The rigorous technical analysis and modelling that is associated with the systems approach are not necessary for undertaking the broad overview that is required of a perspective. However, the approach is useful for determining which elements of the transportation network is "strategic" now and in the future. It should be noted however, that the strategic network will reflect the preferred choices of the travellers and shippers, rather than influence those choices.

One challenge associated with using a transportation perspective is that local and/or regional interests and concerns cannot always be given the attention they deserve. Perspectives are, by definition, broad studies over large geographic areas. Hence, getting municipalities and even provincial staff based in the regions to substitute their parochial interests in favour of a broader provincial interest, becomes a task for skilled negotiators. In Ontario, a small-scale, very focused consultation process for the on-going studies is planned for early Spring, 1996.

Nevertheless, this broad overview approach has one very important advantage. While it may be clear what the physical transportation infrastructure is in any region, it is often less clear how that infrastructure actually functions and what the interrelationships are between modes (supply) and industries and communities (demand). A transportation perspective allows for the examination of these relationships and provides the means through which transportation system management and investment priorities can be judiciously executed. An examination of the Northern Ontario Transportation Perspective illustrates the approach.

Northern Ontario - An Illustration

Northern Ontario is a vast geographic area which, as a region, contains many of the diversities and challenges for transportation planning that are associated with entire countries. This is evident in a brief overview of the area to illustrate the environment in which long range passenger and freight transportation planning and policy development must take place.

At over 800,000 sq. kms., the region comprises over 90% of the total land mass of the Province of Ontario. However, with only 867,700 people (1994 estimate) it is home to only 8.5% of Province's population. In addition, much of the region's population (over 50%) is concentrated in the five largest urban centres of Sault Ste. Marie, Sudbury, North Bay, Timmins and Thunder Bay. A further unique element in the population distribution of this region is the people who live in some remote communities where there is no permanent road access. These isolated communities are almost totally reliant on the air mode for transportation of both passengers and freight.

The Transportation Network of Northern Ontario



EXHIBIT 2

A number of the region's communities are dependent on a single industry or even a single firm for their employment base and their existence. As a result, employment levels tend to fluctuate with markets for the products. At the same time, the region contains the bulk of the province's natural resources. Most of Ontario's annual 20-25 million tons of timber yield comes from Northern Ontario. Thirty-seven of the province's thirty-eight metal mines are located in the area. Its output is worth over $4 billion a year and it contributes over $41.5 billion in value added (at constant 1986 dollars) output to the provincial economy.

In terms of the transportation infrastructure, the region has a fairly comprehensive supply of roads, rail, airports and marine ports to serve both the remote communities and urban concentrations (Exhibit 2). On a map, this transportation network would appear sparse by comparison to the intricate networks in southern Ontario. For the most part, the region is served by only two major provincial highways (Highways 11 and 17) running in a parallel fashion east-west and then north-south. Similarly, the two core lines of the major railways (Canadian National and Canadian Pacific[6]) also run parallel and are complemented by the main lines of the Wisconsin Central and Ontario Northland. The region also possesses 66 or 59% of the province's 111 publicly-owned airports and 12 active marine ports which handle significant cargo volumes.

In summary then, Northern Ontario is an area that is quite large, geographically; is sparsely populated for the most part but contains extremes in population settlement, from urban centres to very remote settlements; is rich in renewable and non-renewable natural resources on which its economy and the survival of most of its communities are dependent. Added to this are some of the major transportation policy and planning issues affecting the region. These include, the elimination of the provincially-owned and operated norOntair air service, the possible abandonment of major rail lines, the impending loss of federal subsidies for some passenger rail services and the uncertain future of intercity bus services to many small communities in the region.

The challenge is to deal with these major transportation issues in a comprehensive manner. The transportation perspective will facilitate this endeavour. A transportation master plan for this region will be created, projecting 30 years into the future, to 2026, using provincially significant criteria for determining the strategic importance of the major transportation links or sites (i.e. highways, railways, airports, marine ports, intermodal terminals, etc.).

A Project Team, comprised of staff from two Ontario government ministries,

with knowledge and expertise in fields of transportation planning, policy development, community economic development and transportation logistics, was established to develop such a master plan. By applying the systems approach, the Project Team set about to firstly identify the strategic components of the existing and future transportation network. Those components are deemed most important for their contribution to economic growth and development and social mobility.

More specifically, the key components of the transportation system were divided into three categories - strategic or provincial, regional and local. Implicitly, this provides a ranking, from the view of the provincial government, from the most important to the least important. In order to identify the components which fit into each category, an evaluation framework based on four provincially important goals was adopted.

(i) regional and provincial economic development;
(ii) addressing user needs;
(iii) management of limited resources; and
(iv) environmental sustainability and sound land use management.

These four goals were further refined to include specific objectives, criteria and measures. For example, four specific objectives were attached to goal 1 - contribution to economic development -:

a. *Support economic competitiveness of Ontario's industries by addressing demand for the movement of people and goods, including needs for convenience, responsiveness, time efficiency and reliability;*

b. *Provide support to Ontario's tourist/recreation industry;*

c. *Promote the integration of transportation systems that can improve the efficiency of passenger and goods movement; and*

d. *Promote and enhance access to Ontario's international crossings.*

To each of the above objectives, specific criteria and quantifiable measures were then developed. For example, the criterion related to the second objective - provide support to Ontario's tourism/recreation industry - was the identification of the transportation links or enhancements to the key tourism/recreation centres. The related measures included traffic data (Summer Average Daily Traffic minus Annual Average Daily Traffic volumes) and cataloguing changes to the trip purpose map.

Of the four provincially important goals adopted, it was the first two - economic development and user needs - that were crucial in deciding which links or sites are included in the strategic network.

Further, under those two goals, the link or site's **function** rather than its level of activity was deemed most critical if it was to be considered strategic. In this case, function was defined by use for long distance, international or inter-provincial travel and/or connection to a trade gateway. In other words, the link or site must play a role in support of trade (the movement of goods and people) beyond the regional boundaries in order to be considered strategic. This is not to say that such "systems" factors as facility volumes, population served, employment concentrations and key industries supported, to name but a few, were not important. In fact, that kind of information forms the backbone of this type of exercise. In this case however, since such information varies greatly between regions in the province, its use was confined to that of supporting evidence used to describe and/or confirm the identified strategic routes and facilities.

Criteria and minimum thresholds were established for each mode and each category. Exhibit 3 illustrates the criteria and measures for the "strategic" and "regional" network. All the remaining elements that are not illustrated are considered "local". For the most part, the regional and local network play a supporting role by acting as feeders to the strategic core.

The actual cut off measures used reflect what the Project Team considered appropriate for the region based on several factors. These included annual average daily traffic (AADT), summer average daily traffic (SADT), commercial vehicle traffic, value and amount of goods shipped through all the major corridors or sites and on the major transportation links. In other words, there is no technical rationale for the actual numbers chosen. It is expected that the cut off values used in other regions of the province will be different, based on the unique characteristics of the region.

The strategic transportation network that was finally identified for Northern Ontario included a highway network of over 3,680 kms. of provincial roadway; all of the main rail lines of three of the four rail companies operating in the region; two airports and two marine ports. This network provides highway access to over 80% of the region's population; handles approximately 83% of the region's rail freight traffic and approximately 82% of its marine trade tonnages. Further, in terms of activity levels (passengers handles, aircraft movements, etc.) the two airports identified as strategic are indeed the most important in the region.

What results from this exercise is a comprehensive overview and understanding of both the extent and nature of the existing transportation infrastructure, and the types and volumes of services using that infrastructure. More importantly, the evaluation approach, while focusing on the interrelationships between the transportation infrastructure and transportation services, including level of activity, reveals which aspects perform a strategic function and provides an overview of industry and the social needs of an area.

Conclusion

It is within this systems framework that such broad and important transportation policy issues as economic development, system efficiency, safety and user needs can be analyzed and addressed. The framework provides a degree of comprehensiveness that would otherwise be lacking if these issues were to be analyzed and addressed independently. It ensures that some kind of synergy exists between policy positions so that decisions in one area complement, and do not work against decisions in another area.

This framework allows intelligent policy development and planning to take place because it places major policy decisions and directions in a context that identifies and understands the modal, social and economic forces and relationships at play in a particular region.

STRATEGIC & REGIONAL TRANSPORTATION NETWORK EVALUATION

	STRATEGIC		REGIONAL	
Mode	**FUNCTION**	**MEASURE**	**FUNCTION**	**MEASURE**
Highways	International; Inter-provincial; Long distance travel; Connects to trade gateways.	* Commercial vehicles/day > 200 OR * AADT >2000	Links pop'n ctrs of 5,000 or more; Connects two strategic links; Important feeder to a strategic link.	* Commercial Veh/Day > 100 OR * AADT > 1000
Railways	International; Inter-provincial; Main line.	> 10 Million tons/km	Secondary Spur	1>10 million tons/km
Airports	International; Inter-provincial.	Passengers/yr > 180,000	Secondary	Passengers/yr 50,000 > 180,000
Marine Ports	International; Inter-provincial.	Cargo/tons/yr > 5 million	Secondary	Cargo/tons/yr 0.5 > 5 million

AADT = Annual Average Daily Traffic

EXHIBIT 3

ENDNOTES

1. These points were raised by P.M. Bunting in an unpublished paper entitled "Developing a Policy Framework for Transportation" (P.M. Bunting & Associates: Kingston, On., April, 1994).

2. Michael J. Bruton. **Introduction to Transportation Planning**. 3rd Edition. (Hutchinson and Company Limited: Great Britain, 1985), p 167.

3. S. Reichman and I. Salomon, "Travel Demand or Need for Access as Future Determinants of Transport Policy-Making" in **Transportation Planning in a Changing World**, ed. P. Nijkamp and S. Reichman (Gower Publishing Co. Ltd.: Vermont, 1987).

4. Ibid.

5. The Greater Toronto Area, Northern and Southwestern studies are all currently in progress but at different stages. The Eastern Ontario perspective has yet to be started. Other similar planning studies have also been completed prior to the launch of these perspectives. For example the "Transfocus 2021" study covered the Niagara-Lake Erie area and was completed in April, 1995. A passenger study entitled "A Passenger Strategy for South Central Ontario" was also completed by mid 1995.

6. Now renamed the St. Lawrence and Hudson Railway.

PUBLIC TRANSPORTATION IN THE FACE OF DETERIORATING ECONOMIC CONDITIONS AND COLONIAL-DEPENDENT SPATIAL POLICIES IN CAMEROON

Ambe J. Njoh
University of South Florida

INTRODUCTION

Some of the most daunting problems facing development planners and policy makers in Cameroon reside in the transport sector. In the country's two largest cities, Douala and Yaounde, traffic congestion is generally as intense as in the cities of advanced nations. At the same time, the access of rural dwellers as well as residents of urban peripheries to important urban areas of activities such as employment, banking, and recreation, is anything but guaranteed. While these problems are not unique to Cameroon, they are aggravated in her case by her unique colonial experience and an economy that has been experiencing negative trends since the mid-1980s.

This paper identifies some specific elements of the country's economic problems and colonial experience that have implications for its urban public transport systems. Additionally, it proposes some measures deemed capable of facilitating the intra-urban movement of people, goods and services in the face of economic crises that show no signs of disappearing within the foreseeable future. An important objective of the paper is to contribute to ongoing efforts addressed to improving urban transportation in Cameroon in particular and Sub-Saharan African countries in general.

NEGATIVE TRENDS IN THE CAMEROONIAN ECONOMY

For twenty years following independence in 1960, Cameroon enjoyed a steady agriculture-led economic growth rate of five percent. With the export of petroleum, the growth rate jumped to eight percent per annum from 1980 to 1986. Since then, things have taken a turn for the worse. From 1986 to 1989, the price of the country's major exports, oil, cotton, coffee, and cocoa declined by 45 percent. From 1985 to 1987, her export earnings were halved, while tax receipts fell dramatically. These problems, combined with a depreciation of the United States dollar *vis-à-vis* the CFA franc, Cameroon's currency plunged the country into deep economic turmoil. Some of the country's economic woes include the loss of two-thirds of its export earnings from 1986 to 1988; a fall in real economic growth by 1.7 percent in 1986/87, 8.6 percent in 1987/88, and 6.0 percent in 1988/89 (US Department of Commerce, 1990). During fiscal year (FY) 1989-90, the country's GDP fell by 6.3 percent while the unemployment rate stood at 40 percent in urban areas and 25 percent elsewhere. Furthermore, investment has fallen by an estimated 50 percent since 1986. Recent World Bank estimates show that the country had a negative per capita growth rate of minus 6.6 percent, the lowest in Africa during the decade, 1985-1994 (InterPress Service, 1996).

Since 1988, Cameroon has been working with the International Bank for Reconstruction and Development (The World Bank) and the International Monetary Fund (IMF) particularly, under the umbrella of the Structural Adjustment Program (SAP), to restore the country's economic growth. As part of the SAP, wages have been drastically curtailed and employees laid off in especially the public sector. Early

retirement of government employees has also been instituted as part of the program as well. Additionally, under the SAP, the government has embarked on efforts to encourage entrepreneurship by simplifying or eliminating a number of business regulations (US Dept. of Commerce, 1991). For instance, import licensing has been abolished except in the case of goods that are subject to 'quantitative controls'. Also as part of the programme, the government has moved to privatize a number of public and para-public agencies. An important initial step in this direction was taken some six years ago (1990) when as many as fifteen state agencies were identified for privatization. Since then, other steps in this connection have been taken including the recent (1995) invitation to interested foreign parties to bid for the country's ailing government-owned and operated telecommunication system. Finally, as part of the programme, the government has attempted to simplify the process of securing business permits in sectors such as transportation, tourism and finance. These developments have far-reaching implications for the country's urban public transportation system. A discussion of the most salient of the implications follows the next section, which reviews colonial transport and urban development initiatives in the country.

COLONIALISM, TRANSPORTATION AND URBAN DEVELOPMENT

An appropriate backdrop for discussing urban public transport problems in Cameroon is the country's colonial history particularly, as it relates to transport and urban development. Cameroon lays claim to a unique colonial history. The country was annexed by Germany in 1884 and remained a colony of that European colonial power until 1919, when as part of the League of Nations' (now United Nations) rulings following the outcome of World War I, the country was divided into two unequal parts of one-fifth and four-fifths, with the larger portion going to France and the smaller one, to Britain. The country thus remained a colony of these two European powers from 1919 to 1960 when it became independent. Each of the three colonial powers contributed in varying degrees to the development of the country's transportation infrastructure (Table 1) and urban settlements.

Colonial Transport and Urban Development

The foundation of Cameroon's modern transport system was laid by the Germans. The latter, like the French and British after them, accorded high priority to transportation because of its ability to facilitate the evacuation of natural resources as well as incorporate the peripheral economy into the international capitalist system (cf. Mabogunje, 1981). The colonial transportation development initiatives were biased in favour of rail. Thus, despite the crude technology available at the time, the Germans succeeded in developing the portion of the *Nordbahn* (present day, TransCameroon rail line) between Douala and Nkongsamba (1906 - 1911) and between Nkongsamba and Windenmeng, a few kilometers away from Yaounde, the capital city (by the outbreak of World War I).[1] The French extended the line into Mbalmayo and later into Yaounde (1927). Even the British, who accomplished very little in the area of transportation, favoured the railway although their effort was limited to a few narrow-gauge rail lines in coastal colonial plantations in present day Fako Division.

Apart from their bias in favour of rail, colonial developers were more interested in inter-regional as opposed to intra-regional linkages. Inter-regional linkages served important functions such as facilitating the movement of badly needed labour from the hinterlands to colonial plantations in coastal areas, territorial defence, and above all, linking the territory through the seaports to its colonial metropolis. This offers at least partial explanation for the fact that colonial transport lines including the TransCameroon railway established by the Germans; the narrow strip of

Table 1: Transport Development in Cameroon by Colonial Era

POLITICAL ERA	MAJOR TRANSPORT DEVELOPMENT	MODE	REMARKS
German (1884-1919)	Built Bonaberi - River Rail Line.	Rail	The rail line was intended as a means of linking two modes of transport, water & rail.
	Completed rail line b/w Douala & Nkongsamba (160 km). Rail		Owned & operated by a private company Kamerun Eisenbahn und Betriebsgesellschaft The line is part of present day Transcam.
	Started Douala-Yaounde rail project (1911).	Rail	By the outbreak of WW I the project had reached Windenmeng near Y'de.
	Completed 280 km of Kribi-Yaounde Road.	Road	The Germans laid 484 km of roads before leaving Cameroon in 1919.
	Work on Victoria (now, Limbe)-Kumba-Mamfe -Bamenda Road.	Road	Labour input was mostly forced, segment b/w Mamfe & Bamenda was wide enough for only one horse-drawn cart.
	Dev. of Bota (Limbe) seaport.	Water	
French (1919-1960)	Extension of Trancam Douala-Mbalmayo-Y'de	Rail	The French essentially continued from where the Germans ended.
	Improvement of Douala Seaport.	Water	The Douala (Bonaberi) port is not a natural sea-port.
	Additions to roads started by the Germans.	Road	French concentrated on the agriculturally fertile coastal areas.
British (1919-1960)	Improvements to Victoria-Kumba-Mamfe-Bamenda Rd.	Road	Improvements were very minimal.
	Built narrow gauge rails in CDC Plantations.	Rail	These narrow gauge rail tracks were located exclusively in the agricultural fields.

untarred road developed during the colonial era to link Bamenda in the hinterland and Limbe (formerly Victoria) in the coast (through Kumba and Mamfe); the 280-km stretch of road between Kribi and Yaounde developed by the Germans; and the narrow-gauged lines in the colonial plantations of Tiko and Limbe have seaports as their termini.

The impact of colonialism on the urban system of African countries is already well-established (see e.g. Wekwete, 1995; Kironde, Chokor, 1993; 1992; Simon, 1992; Mabogunje, 1981; Kanyeihamba, 1980) and need not be rehearsed here. We thus limit our discussion to specific attributes of the impact having implications for urban public transportation. Although all major urban centres in Cameroon possess some distinct German architectural relics, most of the urban design, administrative and legal features were created during the French/British colonial era.

The French differ in some fundamental ways from the British. As Mazrui (1983) notes, while the British hail from the Germanic or Teutonic side of Europe, which believes in racial superiority, the French come from the Latin or Romanic side, which subscribes to a philosophy of cultural superiority. These aspects of ethnic culture affected the nature and orientation of the urban development strategy adopted in Cameroon by each of these colonial powers: for instance, the British were more

inclined to promote physical segregation along racial lines; while the French preferred segregation on the basis of culture. Thus, galicized indigenes were actually accepted in French circles and could in fact live side by side with French colonial officials or immigrants. In contrast, no matter how anglicized a native became, living side by side the English was never permitted. Hence, in the Anglophone portion of the territory, the British quickly moved to institute town planning policies designed to separate native settlements from European ones. The two were separated by very large parcels of undeveloped land or 'green belts'. The impact of such policies are clearly visible in the country's Anglophone townships such as Bamenda, Mamfe, Kumba, Buea and Limbe, where the former European settlements (presently Government Residential Areas or GRAs) and administrative centres remain physically, socially and culturally separated from the multifunctional settlements, where the social use of space is a function of indigenous culture. Despite their cultural differences, the British and French were equally relentless in their bid to 'modernize' the Cameroonian urban system by introducing symbols of European values and social controls. In this connection, the European tradition of zoning to separate different functions and activities was instituted with identical rigour by both in their respective portions of the trust territory. How policies such as these accentuate public transportation problems is discussed below.

IMPLICATIONS FOR PUBLIC TRANSPORT DEMAND AND SUPPLY

The public transport system involves modes of collective as well as personal transportation except private cars used as private means of transportation (cf. World Bank, 1975). We argue that colonial urban development policies, an inherited transportation infrastructure originally designed to serve colonial economic and administrative objectives, and negative trends in the economy render efforts to ameliorate the country's urban public transport system extremely difficult. The discussion is organized under the following broad and overlapping headings: 1) colonial legacy; 2) spatial organization; 3) transport service demand; 4) transport service supply; 5) traffic management; and 6) revenue generation for urban public transport projects.

Colonial Legacy

An important legacy of colonialism in Cameroon is the colonial transport infrastructure, which constitutes the nucleus of the country's contemporary transport system. Rather than being and remaining a facilitator, the colonial transport infrastructure is an impediment to the country's urban public transport system. The infrastructure was designed to serve colonial economic objectives, particularly the inter-regional transportation of goods. Thus, hardly any attention was paid to the linkage of points within urban areas unless such areas were inhabited by Europeans or Europeans and galicized indigenes (in the French-controlled portion of the territory).

The indigenous leaders of Cameroon have hardly deviated from the transportation development 'blue print' left behind by the colonial authorities. Thus, in Limbe the formerly exclusive European residential settlements (now GRA) of Bota were, and continue to be, served by tarred and lighted streets while the multi-functional and historically indigenous settlements of New Town and Mokeba Farms were, and remain, without tarred streets. The same contrast exists in Douala between Bonanjo, where Europeans and galicized indigenes or '*evolues*' lived during the colonial era, and indigenous settlements such as Bepanda. Not only have the country's indigenous leaders seldom deviated from the colonial transport plan, they have added very little to the colonial infrastructure. Yet, rather than diminish, the need for transport services within, and between, the multi-functional and former European

settlements continue to grow with rapid rates of urbanization. Negative trends in the country's economy and steps to remedy the situation through measures to reduce government spending, such as laying off and prematurely retiring public sector employees, means more dependence on public transportation. For one thing, those civil servants previously able to afford private modes of transportation can no longer do so and must therefore depend for their transport needs on public modes. The many currently unemployed former public sector employees are likely to be in search of employment, thus increasing the need for urban transport services.

Spatial Organization

Two features characterize the spatial organization of Cameroonian urban areas. The first is the rigid compartmentalization of land use activities brought about largely by zoning policies introduced during the colonial era and forcefully maintained by the country's indigenous leaders. The second is the dispersion of land use activities over vast areas resulting from colonial racial residential segregation schemes and unplanned ribbon developments along inter-regional roads and railways. Both features have negative implications for urban public transportation. Particularly, the rigid compartmentalization and spatial dispersion of land use activities dramatically increase the cost of developing and maintaining urban transportation systems. Dimitriou (1990: 23) implies this when he states that 'the operation and development of a city's transport system are ...influenced by the organizational structure and geographical distribution of a settlement's major activities ...' This is because the cost of elements comprising a transportation system such as infrastructure is directly related to their length and surface area (cf. Njoh, 1995). Thus, *ceteris paribus,* a two-kilometer stretch of road will cost twice as much as a one-kilometer stretch. Therefore, the dispersion of urban land use activities over vast areas simply makes the task of developing and maintaining urban transportation infrastructure in Cameroon very considerable especially given the present state of the country's economy. Also, such dispersion renders the provision of public transportation, which depends on high volumes on fixed routes, more difficult.

The spatial organizational structure of Cameroon's urban areas also results in the lengthening of distances between activities. This means longer trip duration for passengers, goods and services. In addition, it creates an unnecessary dependence on motorized modes of transportation. This latter problem is not limited to the country's largest cities, Douala and Yaoundé, where their level of development and concomitant economic activities may justify such dependence. Rather, it is commonplace in the smaller urban areas as well. In this regard Njoh (1995: 345) notes that

> In Kumba, most of the city's formal sector employees must commute at least 8 kilometres every day to work at Government Station, an area set aside for public and para-public offices, and located on the north-western outskirts of the city.

Approximately the same distance separates Government Station, Bamenda, from the main city or what was alluded to during the colonial era as 'native settlements'. The problem in Bamenda's case has to do not only with distance but also topography. A spiral road links Government Station located at the peak, and the main city located at the foot, of a very steep, high and rocky hill. The location of government offices and living quarters for public officials on hill tops was typical of the British colonial administration. The administrative towns of Bamenda, Wum and Nkambe of the North-West Province are reminiscent of this situation (Tandap, 1976). Thus, in most cases, the lack of resolve or energy to make the necessary climb up the hill compel residents to depend on motorized means of transportation, thereby spending scarce financial resources that would otherwise be used to satisfy other basic needs.

The impact of the country's spatial organizational structure on urban public transportation can therefore be summarized as follows:

- increased cost to both the state and passengers;
- undue cost-related burden on the unemployed and poorer groups of the population;
- increased trip duration for passengers, goods and services; and
- decreased productivity and efficiency.

Transport Service Demand

One fundamental way in which Cameroon's economic crisis and attempts to reverse it have affected urban public transportation in the country is by increasing demand. This increase can be appreciated from at least three perspectives. First, those who prior to the economic crisis depended on private modes of transportation such as cars and motorcycles can no longer afford the cost of fueling and maintaining these vehicles and must depend on the urban public transport system. Second and paradoxically, the increase in unemployment levels and the practice of prematurely retiring public sector employees as part of ongoing IMF/World Bank-recommended steps to reduce government spending has increased, rather than diminished, the demand for public transport services. Former public sector employees, especially retirees, most of whom currently live in rural areas where they were born but had lived in urban centres throughout their entire working life, frequently return to these centres to collect retirement checks and for urban-based services such as banking, shopping, health care, and recreation. Additionally, the unemployed, whose ranks have tremendously increased, are constantly on the move in search of, or doing mostly temporary, work in the informal sector. In both cases, the result is an increase in demand for public transport services. It is important to point out that the demand problems posed by the present situation, which is characterized by high levels of unemployment, irregular employment, and a rapid increase of informal sector activities, are different from those posed prior to the mid-eighties, when a substantial portion of the working age population was employed in the public sector. For example, whereas in the latter case, peak traffic hours were constant (7:30 to 9:00 am, when people went to work and school; 12:00 to 2:30 pm, lunch time; and 5:00 to 6:30 pm, at the close of the working day), in the former case, peak hours are difficult to predict. Finally, a rapid increase in the population, partially a product of the idleness brought about by high levels of unemployment, has also increased the demand for public transport services. For instance, children born when the country's economic woes begun in the mid-1980s are about eleven now, and use public transport services to attend school.

This latter problem is accentuated by policies that seek to compartmentalize urban land use activities. Thus, schools are often located a considerable distance away from residential areas. In fact, it is not uncommon to find a secondary school in Cameroon located in the middle of the woods, far removed from the urban area, where its students reside. For example, a significant proportion of the students at Government Technical High School Ombe commute from Limbe (about 16 km away) and Tiko (about 9 km away); while some of those attending the Government Bilingual High School in Molyko must commute from Buea (approximately 9 km away). As stated earlier, policies promoting the physical dispersion of land use activities were introduced in the country by colonial administrators and continue to be vigorously enforced by indigenous authorities. During the colonial era, the question of transportation was not serious as students were housed on campus. In the 1990s, not only are the government, religious organizations and proprietors of private schools unable to operate dormitories, parents are also increasingly unable to afford the additional cost of boarding. Hence, students must commute long distances requiring motorized means of transportation.

Transport Service Supply

Cameroon's economic crisis has substantially weakened the ability of the government and private entities to provide and/or maintain necessary urban transport services and facilities. Government's dwindling resource base means it can no longer construct, maintain and rehabilitate major urban transportation infrastructure such as streets, bridges and motor parks. Nor is it capable of supplying and servicing public transit vehicles. In fact, the country's only para-public mass transit agency, *Société de Transports Urbains du Cameroun* (SOTUC), which provides bus transit services to the few areas with tarred streets in Yaoundé and Douala, faces a very uncertain future as government subsidies that kept it afloat prior to the economic crisis have markedly diminished.

Government's ability to supply transport facilities and/or services is particularly debilitated by two crucial factors central to the country's current economic crisis and IMF/World Bank recommended measures to reverse the situation. First, most road building/maintenance equipment and/or materials must be imported. This triggers a foreign exchange problem, itself compounded by the devaluation of the CFA Franc, Cameroon's currency. Second, government's revenue earning ability has considerably waned and continues to grow weaker in part, as a result of fallen prices for the country's main export products. This, at least partially, explains why the para-government public transit agency, SOTUC, is no longer able to secure the revenue necessary for purchasing new buses and/or maintaining those in the existing fleet; and why national and local government authorities are increasingly unable to construct and/or maintain city streets. Thus, the urban transport problems of Cameroon like those of other developing countries, are likely to be associated with the problem of declining economic growth and stagnation than with anything else. Under such circumstances, it has been argued that 'the inadequacy of resources to provide for the maintenance and management of transport services and infrastructure' is likely to supersede all other transportation problems including those generated by increased vehicle ownership in developing countries within the foreseeable future (Dimitriou, 1990: 17).

Apart from the government and para-public entities, private individuals play a crucial role in the supply of urban public transportation services in Cameroon. Privately owned and operated taxicabs are second only to walking as the mode of intra-urban travel in the country. The government regulates how much operators can charge passengers per trip. Such rates, by law, are posted on the vehicles. However, the government rates are seldom abided by. The actual fares are usually subject to haggling and differ depending on the type of service and the time at which it is provided. Services provided between 10 p.m. and 5 a.m. are generally more expensive. There are two main kinds of taxi services available. One is an express service locally known as *'course'*, which transports the passenger directly to his/her destination without stopping to pick up other passengers along the way. Another is a regular service locally known as *'ramassage'*, which provides transportation service to the passenger from point A to B with stops in-between as necessary to pick up and/or drop off other passengers. The problem of diminishing revenue earning ability on the part of the government has surfaced at a time when individual citizens have witnessed their incomes drastically reduced either as a result of falling export commodity prices or salary cuts or lost employment in the public sector. This means that now, more than ever, private operators of public transportation vehicles find themselves increasingly unable to afford the purchase and maintenance of such facilities.

Traffic Management

Cameroon's traffic management problems have been compounded by the deteriorating state of her economy. While the country's urban centres have grown and proliferated since independence, the main thoroughfares have remained the same especially in terms of width. Hence, the nagging levels of congestion prevalent in especially the country's two main cities, Douala and Yaoundé. In advanced countries, as much as 15 to 25 percent of the space in urban areas (over 30% in newer low density U.S. cities) serves as streets, but only 10% (and very rarely, 15%) of the urban space in LDCs serves the same purpose (World Bank, 1975). In Cameroon, with high levels of unemployment and a significant increase in informal sector activities, this means the available limited right-of-way must be shared by an assortment of activities such as, parking, hawking, loading, unloading, and various traffic including, pedestrians, manually operated trucks, taxicabs, private automobiles, buses, heavy and light-weight trucks. Additionally, the lack of funds on the part of the national government means the task of installing new, or restoring broken, traffic signals, road and street signs including street names, becomes insurmountable. Thus, because of inadequate signage, the prevalence of nameless streets and poor driving habits, not only is there an increase in traffic accidents, traffic is also significantly slowed thereby increasing travel time. In fact, during peak hours, speed in the central areas of the main cities often falls to as low as 10 to 15 km per hour.

Revenue Generation for Urban Transport Projects

Growing economic problems have, among other things, contributed to significant leakage in the mechanism for generating revenue for urban public transportation projects in Cameroon. For instance, the toll gates that have recently been erected to collect funds for such projects are not sufficiently staffed as a result of government cut-back schemes. This results in rampant toll fare evasion. Furthermore, the few available toll gate employees resort to diverting toll gate funds for their own personal use as a means of augmenting their meagre wages that continue to shrink as a result of the salary reduction schemes adopted by the government in response to IMF/World Bank demands. In order to divert such funds without being caught, toll gate attendants customarily collect less than half the official fee of CFA 500 (i.e. about CFA F 250 or CFA F 200) as long as the paying party agrees not to be issued a receipt. Leakages exist in other areas including public transport vehicle registration fees *(vignette)*, transport vehicle taxes *(patente)*, and taxicab/bus driver licensing and license renewal fees. Paradoxically, leakages in these areas are caused mostly by agents of the state or those in positions of responsibility. For example, members of the police force, the military and high ranking government officials operate public transport vehicles without defraying the necessary fees. Reduced earnings in especially the public sector may be responsible for this as well as the escalation of fraudulent activities relating to the importation of transport vehicles and related goods. Although these problems pre-date the country's economic crisis, we contend that they have grown and proliferated with the emergence of the crisis in the mid-1980s.

POLICY IMPLICATIONS

The need for caution in recommending any action for transportation planning cannot be overstated. This is because transportation planning, especially in developing countries, is anything but an exact science. Cognizant of this, we therefore limit our recommendations to those with some Cameroon-specific evidence of working rather than the generalized and familiar measures that abound in the relevant literature. Thus, we suggest that the necessary steps be taken to:

- promote transport-friendly urban design schemes;
- encourage the private ownership/operation of public transport vehicles;
- encourage other modes of transportation;
- elicit the direct participation of citizens in street development and maintenance;
- improve revenue generation strategies;
- improve traffic management.

These proposed remedies and the problems discussed above are summarized in Fig. II.

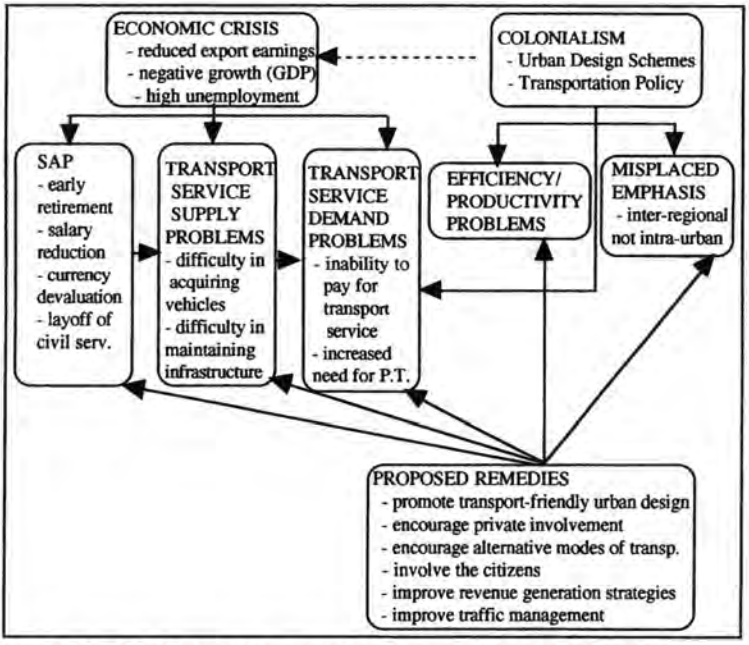


Fig. I: Model of Colonial/Economic Impediments to, and Proposed Strategies for Improving, Urban Public Transportation Effectiveness in Cameroon

Promote Transport Friendly Urban Design Schemes

As argued above, by being highly compartmentalized and comprising sparsely distributed units, urban centres in Cameroon render the task of public transportation service delivery difficult at best and impossible at worst. We therefore propose that authorities encourage highly dense, more compact small urban areas as a means of facilitating this task (cf. World Bank, 1975). To do so, is to return to more or less Cameroonian traditional human settlement patterns. These patterns were predominant in pre-colonial Cameroon and remain common features of the country's rural built environment. Traditional Cameroonian human settlements are often compactly arranged in a semi-circular formation around a courtyard (Tandap, 1976). In siting

buildings, care is taken to ensure proximity to a basic human needs such as a constant source of water for domestic use as well as the use of domesticated animals. This automatically leads to the location, within close proximity of dwelling units. Therefore Cameroonians are more likely to identify with a compact settlement pattern than with a sparse one such as was promoted by colonial urban development policies and currently sustained by contemporary zoning legislation. A compact settlement pattern is less costly to furnish with transportation facilities.

Furthermore, we recommend a mixed land use system, which will allow habitation, light home-based industries, wholesale, retail, and offices in the same district. This, as opposed to a segregated land use pattern based on strict zoning legislation, will significantly reduce the distance to work and other activities thereby reducing dependence on motorized modes of transportation. In fact, the location of offices in residential neighborhoods, albeit for different reasons, is nothing new in Cameroon. With the multiplicity of government agencies that occurred between the mid-1970s and mid-1980s, the lack of space in government-owned buildings compelled authorities to house public offices in residential neighbourhoods.[2] There is no evidence that these offices are any less productive than their counterparts located in exclusive office parks.

Private Ownership of Public Transport Vehicles

That the Cameroonian government currently faces serious resource constraint problems dictates a need for alternative means of quantitatively and qualitatively improving supply to meet the growing demand for public transport services. While private individuals play a significant role in the supply of transport services, the supply of public bus transit services in the two largest cities, Douala and Yaoundé is monopolized by the para-public agency, SOTUC. Additionally, private suppliers of public transport services in urban centres throughout the country, must contend with a lot of stringent rules and regulations. The effect of this is a significant reduction in the number of transport vehicles that would otherwise be supplied. As a means of achieving a better match between the demand and supply of public transport vehicles, we suggest that the rules and regulations be relaxed. Under conditions of severe resource constraints such as are characteristic of Cameroon today, the private or informal supply of public transport vehicles present some unique advantages. In comparison to the formal sector, the supply of public transport vehicles by the informal sector possesses the following characteristics (cf. Banjo, 1994):

- lower overhead cost;
- no dependence on the government for financial aid;
- high accessibility to users;
- personalised services;
- high intensity use of vehicle;
- low to medium running cost; and
- low capital outlay.

Apart from these advantages, the relaxation of pre-conditions for operating public transportation vehicles, hence the encouragement of private participation in the public transportation market, have many positive implications for the Cameroonian economy as a whole. For instance, there will be more competition where there is more participation on the part of private individuals, and, hence, an increased and more differentiated supply of public transport vehicles tailored to respond to specific demand characteristics (Banjo, 1994). Additionally, the involvement of private individuals reduces the need for government subsidies often attached to the provision of public transport services. Thus, not only does the involvement of private individuals free government funds for use in other areas, it increases efficiency and

productivity by guaranteeing individuals personalized services at minimal cost. That there is considerable involvement on the part of private individuals despite existing institutional constraints signifies that an elimination of the constraints will go a long way in encouraging the entrepreneurship that is central to IMF/World Bank proposals.

Experience elsewhere contains a large number of examples of public transport systems that have benefited enormously from the involvement of private individuals. In Hong Kong for instance, as part of the reforms that took place after the 1967 political upheavals, the government legalized mini-buses as a form of public transport. This policy was hardly nine years old in 1976, when the number of mini-buses or 'public light buses' (PLB) had grown to 4,350, accounting for the movement of as much as one-third of the state's public transport passengers. "So profitable were these PLBs by 1980 that their owners could recover in a single year two-third of a vehicle's purchase price" (Urban Edge, 1982: 2). In Argentina's capital city, Buenos Aires, at least 75% of the public bus trips are made by privately owned micro-buses or "*collectivos*", which usually contain 23 seats. In Cairo, Egypt, a union of taxi drivers and minibus operators operates fixed-route jitney services. As of 1979, some 800 vehicles were being operated under the scheme. In Calcutta, India, most of the bus services are provided by private buses. The more telling fact about this case is that while the government-owned, Calcutta State Transport Corporation (CSTC), which services the city's best roads, receives a subsidy of US $1 million per month, the privately operated buses require no subsidy whatsoever. One possible explanation for the impressive record of privately operated buses is their greater willingness to make the necessary repairs/maintenance, collect the fares, and work harder (Ibid: 2).

Encourage Other Modes of Transportation

One problem with the Cameroonian urban transport system is the availability of an extremely limited variety of transport vehicles. This problem appears common to countries in Africa in comparison to those in other comparable parts of the world. The problem of limitation in variety of passenger vehicle type, an example of deprivation, has been noted in Sierra Leone, where, as in the case of Cameroon, 'the only alternative to travel by bus, truck, or a combined passenger/goods pickup, is walking' (World Bank, 1984: 9).

Prior to the abolition of the federal system and the subsequent 're-unification' of the Anglophone and Francophone portions of the country in 1972, West Cameroon (the Anglophone part of the country or present day North West and South West provinces) had a richer mix of public transport vehicles, ranging from non-motorized options such as heavy duty hand carts to motorized vehicles such as heavy duty truck chassis mounted with locally manufactured bus bodies (locally known as *ngongoros*). Following 're-unification' in 1972, the country's economic boom, and the absence of institutional support for innovation and creativity in the transport sector, these informal modes of transportation were replaced by 'modern' and sophisticated vehicles imported from mostly France and Japan. This development has had many negative implications, ranging from the elimination of labour-intensive jobs in the vehicle construction industry to the inability to adequately provide demand-oriented services in areas without paved roads. As a means of redressing this situation, it is necessary for authorities to fully support the informal transport sector by investing in research effort aimed at developing and/adapting alternative modes of transportation or paratransit systems. Paratransit or informal sector transport mode currently found in comparable parts of the world range from the manpowered rickshaw (found in India and Indonesia), to small vehicles derived from motorcycle technology. Also included are larger vehicles such as full size four-wheel drive

chassis fitted with bus bodies to carry up to 60 passengers (e.g. the *fula-fulas* of Kinshasa, Zaire) (White, 1990).

EncourageCitizen Participation

We have already argued that the Cameroonian government cannot by itself accomplish the task of public transport service delivery. Hence we have suggested, among other things, the involvement of private individuals. To be sure, we are not advocating that 'lumpy' and indivisible elements of the system such as the provision and/or maintenance of streets be placed in the hands of entrepreneurs or private citizens. Ideally, such elements should remain the responsibility of the government. However, we see the possibility of government enlisting community support in this connection given the uncertain economic conditions currently engulfing Cameroon. The government can create an environment conducive for the private delivery of urban public transportation in several ways. It can, for example, seek to rehabilitate the existing infrastructure, particularly the streets.

We suggest that local communities be given the responsibility to provide the necessary labour input into street maintenance while the government retains the role of supplying the material input. The labour input on the part of local communities can be in the form of skilled/unskilled labour. Where more sophisticated labour inputs are necessary, the government can tap the expertise of the Army Corps of Engineers, who, must be paid whether they performed such otherwise productive duties or not. Institutional reform efforts in the transport sector in LDCs should strive towards involving citizens or road users in many ways including the management and maintenance of roads. To do so is to empower citizens and give them a stake in their communities. Once this is done, it becomes possible for the government to enlist communal support (without coercion) for such important projects as street or road maintenance and upkeep. According to Africa Transport (1994), the 'Newsletter of the Sub-Saharan Transport Policy Program', involving road users in the management of roads is not only one way of winning public support for among other things, more funding but also one of four basic building blocks or pre-requisites (the others include, stabilizing road, identifying responsibility, and strengthening road managemen) for the more fundamental reforms necessary for improving the effectiveness of road management in Africa.

Efforts to elicit citizen participation in community development efforts in Cameroon have always been very successful. Through various grassroots community groups based on churches, cultural and ethnic organizations, cooperatives and so on, citizens have successfully completed community development projects, such as water supply schemes, irrigation, school building, and rural road building throughout the country. For instance, during its 14-year history, one such grass-roots community-based organization, the Bafut Village Community Project (BVCP) has constructed 72 kms of community roads, 16 primary and secondary schools, 25 safe water systems and launched a promotional rice cultivation programme using irrigated water (Musa, 1996).

Improve Revenue Generation Strategies

As heretical as it may seem, the Cameroonian government is capable of generating significant revenues by doing no more than simply revamping the system by which it collects fees and taxes associated with transportation. One specific suggestion in this connection is that serious steps be taken to curb import tax fraud. Such steps may include computerizing the information storage and retrieval system of the Customs Department and introducing severe penalties to deal with such fraud. We have already noted that the toll gate policy is marred with a lot of leaks. Thus, given the manner in which they are currently operated, toll gates can hardly be considered a

revenue source for the government. We therefore suggest that the gates be eliminated. Since toll fees are supposed to compensate for damage to streets or roads caused by vehicles, we propose that vehicles be charged based on their weights and distance traveled and that the fees so charged be collected in two ways (cf. Prests, 1969): as part of a gasoline tax, to account for distance traveled; and as part of the vehicle license fees, to account for the vehicle weight . In this connection, the fuel levy should be collected at the wholesale, as opposed to the retail, stage. This will significantly reduce the chances of tax fraud. We further suggest that such levies on fuel and license fees be separated from government sales and excise taxes, and be deposited directly into a special 'Road Maintenance Fund'. Examples of such an operation can be found in the Central African Republic, Ghana, and Zambia (Africa Transport, 1994). The revenues so collected need to remain with municipal governments which, we believe, need to have more say in urban transportation management in particular and urban policy issues in general.

Improve Traffic Management

To attempt to deal with the physical problems identified above by widening existing or building additional, streets is not only financially prohibitive, given the country's economic problems, but will not constitute a judicious use of extremely scarce resources. We therefore suggest that serious efforts be made to maximize the utility of existing streets. This can be accomplished through improved traffic management policies. The national government in partnership with local governments can significantly improve public transportation traffic flow by formulating and implementing policies addressed to significantly increasing the speed of urban public transportation vehicles. Such policies include,

- discouraging private automobiles from especially, busy urban areas by for instance, charging high parking fees;
- making all busy streets 'no parking zones';
- prohibiting heavy trucks from using certain streets;
- prohibiting loading and/or unloading on certain streets during peak traffic hours;
- ensuring that hawking is restricted to sidewalks;
- converting some downtown streets, especially the narrow ones into 'one way' streets; and
- introducing 'bus only' lanes.

Pedestrians constitute the majority of road users in most downtown areas. Thus, the speed of public transportation vehicles sometimes depends on the behavior of pedestrians. We therefore suggest that especially urban citizens be educated about road use. In Calcutta for instance, officials have advocated a similar strategy and schools in the Metropolitan area have already incorporated traffic education in their syllabus. Traffic flow also depends on how well traffic signals function as well as the presence/absence or visibility/obscurity of road/street signs. Efforts should therefore be made to install street and road signs in appropriate and highly visible points along urban streets in the country.

Again, this is one area in which community input can be sought. Specifically, we propose the improvisation of street signs by using materials such as left-over roofing corrugated aluminum sheets, empty containers of oil, tar, fertilizers and so on. Local sheet metal workers can be charged with the responsibility of flattening and cutting these materials to specified sizes and shapes, while local painters paint and label the signs. The rest of the community can then be required to make in kind or in cash contributions. The latter can go towards the purchase of paint and other essential materials for the project. As mentioned earlier, such participation is not unusual in Cameroon. In fact, in the BVCP case cited above, villagers are required to similarly contribute in kind or cash to twenty-five percent of the cost of infrastructural projects

such as bridges, health post and schools, and twelve percent for roads (Musa, 1996). The rest comes from external sources, particularly the national government.

CONCLUSION

This paper has discussed factors inhibiting the performance of the Cameroonian urban transport system as a function of colonial spatial development policies and contemporary economic problems. The system's performance can be significantly improved by de-emphasizing government's current role as regulator while reinforcing its role as facilitator. It is recommended amongst other things that efforts be made to adopt policies that minimize spatial dispersion, encourage private ownership and operation of public transport vehicles, directly involve citizens, improve revenue generation in the transportation sector and traffic management.

NOTES

1. There are plans to extend the Center Line to Moundou in Chad and the Western Line to Bangui, the capital of Central African Republic.
2. Examples of offices operating out of residential neighborhoods include, but are not limited to, the departments of Housing, Surveys, and Lands of the Ministry of Town Planning and Housing in the cities of Kumba and Limbe, and the Divisional Officer's Office, Bamenda.

REFERENCES

Africa Transport (1994) 'Management and Financing of Roads: An Agenda for Reform' Newsletter of the Sub-Saharan Africa Transport Policy Program (No. 9, October), United Nations Economic Commission for Africa (UNECA) and World Bank, Washington, DC.

Banjo, G. Adegboyega (1994) 'Deregulation of Urban Public Transport Services: Some Realities for Policy Makers in African Cities'. Third World Planning Review, 16 (4), 411-428.

Chokor, Boyowa A. (1993) 'External European Influences and Indigenous Social Values in Urban Development and Planning in the Third World: The Case of Ibadan, Nigeria'. Planning Perspectives, 8, 283-306.

Dimitriou, Harry T. (1990) 'Transport and Third World City Development', in Dimitriou, Harry T. (ed.), Transportation Planning in Third World Cities. London/New York: Routledge.

----------(1990) 'Transportation Problems in Third World Cities' in Dimitriou, Harry T. (ed.) op. cit. InterPress Service (1996) 'Developing Countries Not Developing'. International Electronic Network (Internet), Jan. 2.

Kanyeihamba, G.W. (1980) 'The Impact of the Received Law on Planning and Development in Anglophonic Africa'. International Journal of Urban and Regional Research, 4, 239-266.

Kironde, Lusugga, J. M. (1992) 'Received Concepts and Theories in African Urbanisation and Management Strategies: The Struggle Continues'. Urban Studies, 29 (8), 1277-1292.

Mabogunje, Akin L. (1981) The Development Process: A Spatial Perspective. New York: Holmes & Meier.

Musa, Tansa (1996) 'Cameroon Development: Oxen and Participation'. PeaceNet World News Service, IGC Networks, San Francisco, CA (Via International Electronic Network (Internet)), Jan. 2.

Njoh, Ambe J. (1995) 'Building and Urban Land Use Controls in Developing Countries: A Critical Appraisal of the Kumba (Cameroon) Zoning Ordinance'. Third World Planning Review, 17 (3), 337-356.

Prests, A.R. (1969) Transport Economics in Developing Countries: Pricing and Financing Aspects. London, Weidenfeld and Nicolson.
Simon, David R. (1992) Cities, Capital and Development: African Cities in the World Economy. London: Belhaven Press.
Tandap, Lucas T. (1976) 'The Evolution of Rural Settlements in Cameroon'. Ekistics, 249, 106-108.
Urban Edge (1990) 'Cameroon: Urban Development'. The Urban Edge, January/February, 1990.
-------(1985) 'Focusing on Incomes and Productivity: A Fresh Look at Urban Employment'. The Urban Edge, 9 (8), September/October.
-------- (1982) 'Improving Transportation for the Poor' The Urban Edge, 6 (1), January .
Wekwete, K.H. (1995) 'Planning Law in Sub-Saharan Africa--A Focus on the Experiences in Southern and Eastern Africa' Habitat International, 19 (1), 13-28.
White, Peter R. (1990) 'Inadequacies of urban Public Transport Systems' in Dimitriou, H. (ed.) op. cit.
World Bank (1986) The Supply and Quality of Rural Transport Services in Developing Countries: A Comparative Review. Staff Working Paper. Washington, DC: The World Bank.
World Bank (1975) Urban Transport: Sector Policy Paper. Washington, DC, World Bank Headquarters.
U.S. Department of Commerce (1990) 'Marketing in Cameroon' Official Report, Overseas Business Reports (OBR) No. 90-10, U.S. Department of Commerce, International Trade Administration, Washington, DC.

Performance Indicators of the Transport Sector

by Jean-Pierre Roy

1.0 Introduction

Transport plays a vital role in the development and functioning of the Canadian economy as a whole, and the furtherance of the competitiveness of Canadian shippers. The performance of the transport sector is thus critical in this regard. This paper develops a framework to evaluate the transport sector's performance by simultaneously taking into account productivity, price changes, and the financial results of carriers. Emphasis is placed on the examination of the interrelationships between these performance indicators. Measurement and methodological issues are discussed. Empirical results are shown for the period 1981 to 1993.

2.0 Concept of Economic Performance Measurement

In the long term, at equilibrium, the revenues of a firm are equal to its costs, where all factors of production are properly remunerated, including a return on the capital input. This relationship is expressed in the following dual function:

$$\text{Revenues } (P*Q) = \text{Cost } (X*F)$$

where
- P is a vector of the unit price of output
- Q is a vector of the quantity of output
- X is a vector of the unit price of factor input
- F is a vector of the quantity of factor input

One of the crucial performance indicators used is Total Factor Productivity which simply expresses a relationship between the index growth of Q and that of F. TFP is a more robust and comprehensive indicator of performance than the more conventional, but partial, concept of labour productivity.

When input and output quantity indices (F and Q respectively) are known, input and output prices (X and P) can be derived by dividing the revenue (R) and cost (C) indices by the appropriate quantity index. Conversely, when price indices are known, we can derive the quantity indices in the same manner. In fact, both approaches were used, depending on the data availability. Finally, unit cost indices can be developed by dividing input price indices by productivity indices.

3.0 Output Measurement Issues

Outputs of the transportation industry consist, in general, of two broad revenue categories; freight and passenger. The existence of aggregate measures such as tonne-kilometre or passenger-kilometre creates the impression that output measurements are readily available. However, these measures do not allow for differences in traffic characteristics across commodities, size of shipments, length of haul, stage length or markets served. The aggregation problems are compounded when the firms' output shifts over time. It also creates serious problems of comparisons between firms, as this author has demonstrated in "The Effect of Commodity Mix on Yield and Unit Labour Cost in the Railway Sector".[1]

The use of aggregate output indices supposes that, in freight, a coal movement costing $20/tonne is equivalent to a movement of automobile products that costs $150/tonne. Or, in the passenger area, it gives the same weight to a short haul movement of less than 450 kilometres which costs between 25 and 30 cents per kilometre as to a Trans-Pacific flight of 7,000 kilometres yields some 9 cents/km.

Transport Canada had undertaken some research in the measurement of price changes in trucking and rail.[2] The effort has been extended to the air industry[3] where price changes have been measured by market segments, by carrier and between scheduled and charter services. The effect of variations in stage length has also been removed from changes in yield. In the context of this exercise, estimates of price changes have been developed for the bus and water transportation industries. However, the estimates are volatile and should be revisited. Output indices were derived by dividing the index of revenues by the price index.

The choice of an output measure has a considerable bearing on productivity results. In the air industry, for instance, aggregate measures of seat kilometres in scheduled services indicate a 23% growth between 1981 and 1993. Our desegregated measure of output shows an overall growth of 8.6%. This translates directly into productivity differentials since the numerator figure remain the same. Tretheway et al[4] have documented the evidence on the magnitude of errors (8.3% over the 1981–1990 period) which can be introduced by the use of aggregate data in rail.

4.0 Input Measurement Issues

Likewise, inputs usually consist of multiple and heterogeneous factors (e.g. labour, capital and fuel), which cannot be added together. The technique used to aggregate the indices of individual inputs was to combine them by weighting each input factor according to its respective share of total costs, by using the Törnquist approximation to the Divisia Index.

4.1 Labour and Fuel

Labour makes up, on average, 36% of the revenue of the transport sector, the highest share being recorded by the rail industry (48%), the lowest by the water transport industry (26%). Fuel accounts for 12% of the revenue of the transport sector. The air industry has the highest fuel cost (19% on average) down however from 27.5% in 1982.

It is straightforward to develop a labour quantity index, since the number of employees or person-hours is readily available. If detail by job category was available, individual indices were constructed and a composite index was derived using Törnquist procedures. With respect to fuel, an aggregate volume index was built, based on the thermal content of each fuel used.

4.2 Capital

The model supposes that the fully allocated costs of the firms are equal to their revenues. In this case, the service price of capital will be equal to depreciation and the operating margin. However, nothing guarantees that the operating margin of the firms will be sufficient to ensure the repayment of the debt and provide a sufficient return to the shareholders. What is often observed in the transport sector is that carriers' prices (determined by market forces) fall more rapidly than their input costs (controlled mostly by management). The solution consists of imputing the Cost of Capital (COC) incurred by the firms.

Because we are concerned with the operations of transport firms, the COC is developed for capital goods (e.g. property, plant and equipment) used in each transport sector as measured by their net book value. The method of financing (i.e. by equity, debt or deferred taxes) was determined from the firms' capital structure. For assets financed by debt, the actual amount of interest paid by firms was used, whenever such data was available. Otherwise, the cost of long term borrowing by corporations was applied, using a five-year moving average to smooth the effect of year-to-year fluctuations in interest rates. To reflect the relative riskiness of equity versus debt, the return on equity was deemed to be equal to the long term bond rate plus a premium of 2%, after tax. Finally, the rate on accumulated deferred tax balances and, in the case of railways, government grants and donations, was set at 0%, since these funds can be viewed as an interest free loan from the government.

This produced a blended cost of capital rate which, multiplied by fixed assets, yielded the COC of the firms. The service cost of capital (SPC) was the COC plus depreciation. The COC was substituted to the operating margin of the firms to determine their fully allocated cost and the share of each factor of production was computed from fully allocated costs, not total revenue.[5]

The ideal method of measuring the index of capital utilization is to use capital stock figures in constant $ derived from perpetual inventory methods. These data exist only at the aggregate level for each transport industry and do not allow for a segmented analysis by industry components, but in rail, thanks to the work of Tretheway et al[6]. The concept of capital stock is similar to that of net book value, except that the latter is expressed in original dollar cost. We asked Statistics Canada to prepare a special compilation of the stock of capital in original dollar costs and in constant dollars. The ratio of original/constant dollars by industry was applied to the net book value of assets for each industry component, to yield an estimate of capital stock in constant $. The price index of capital followed from the division of the index of SPC by the capital volume index.

4.3 Leasing

Increasingly, transport firms are relying on leasing as a source of capital. Nowhere is this trend more pronounced than in the air industry, where some 45% of capital costs are incurred from leasing. On average, in all transport, the proportion of the cost of capital arising from leases now exceeds 25%.

The indices for capital assets leased are prepared differently. Since the price of individual assets is unknown, the annual equivalent rental or lease payment can be estimated on a per unit dollar basis, using formula to calculate the periodic payment to pay off a loan, on the basis of the cost of the funds (long term bond rate plus 2%) and the service life of assets. The leasing price index was estimated by adjusting for changes in capital prices from Statistics Canada. The leasing volume index is then the result of dividing the index of rent/lease payments by the leasing price index.

4.4 Other Input

The standard approach towards the residual category (38% of operating revenues) consists of converting these expenditures into constant dollars by using the GDP implicit price index. It has been demonstrated [7] that it is a valid approach in rail, where the residual category is limited to 25% of the total cost. However in other modes, such as trucking, this residual category may account for as much as 50% of the firms' cost. Included in the residual category are costs such purchased transport services which essentially are services purchased from rail (4%) and owner-operators (23%). In the air industry, a major cost component was promotion and commissions to ticket agents (14%). Other costs include food and beverages (4% of the cost of airlines and 1% for the shipping industry), government fees (4% for marine, and 2.5% for the airlines).

What we did was to split the residual category into identifiable components and adjust the estimation process of quantity and price indices to each case. For

instance, in trucking, we selected the price index of the inputs used by the trucking industry as a proxy for the price of owner-operator services, on the basis that trucking firms will choose to use external resources to their firms if the prices of the resources are at least comparable to the cost of their internal resources. In the airline industry, the volume index of commissions paid to ticket agents was deemed to be equal to the index of passenger revenues. Other promotion costs were linked to the price index of advertising services.

5.0 Financial Performance

The financial performance of the firms deteriorates when the price of inputs exceeds output prices over the long run, and the gap is not made up by productivity gains. In financial terms, a firm is viable if its return on investment is comparable to its cost of capital (COC). The return on capital (RC) can be expressed as net operating income, divided by the net book value of capital invested in the enterprise. Net income, sometimes referred to as the OM, is the difference between operating revenues (OR) and operating costs (OC). Thus the viability of the firms can be tested by comparing the COC with the operating margin of the firms.

A firm is considered viable if it generates an OM sufficient to earn its COC. In other words, it meets the test of revenue adequacy. If the OM is insufficient, then a "margin deficiency" occurs.

The linkage between these measures is the Asset Turnover Rate (ATR), which is the ratio of operating revenues to the net book value of capital assets. ATR indicates the degree to which a firm is using its assets to generate revenues. A firm with excess capacity would be expected to have a lower ATR than a similar company where capacity is fully utilized.

This approach of expressing the cost of capital in relation to the operating revenues of the firms has three advantages as follows:

- it determines the adequacy of the firms' profitability using the simpler and more accessible measure of operating margins; and
- it facilitates intermodal comparisons, by showing the degree to which a high operating margin is required for highly capitalized firms, compared with others that need lower operating margins because they need fewer assets to produce a unit amount of revenue.

The analysis of the transport industry's financial performance incorporates the following elements:

- Net fixed assets, and the portions financed by equity, deferred taxes, and debt;
- The cost of capital, and the service price of capital;

- Asset turnover rates (the ratio of operating revenue over net fixed assets);
- Operating margins, and margin deficiency or surplus; and
- The coverage of interest expenses by the operating margin.

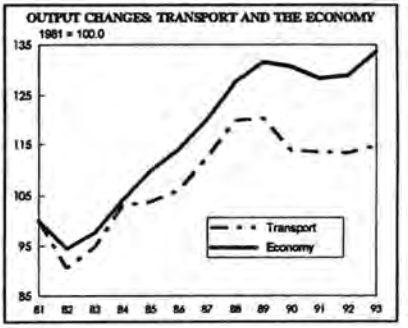
Most of these data elements have been developed in the course of developing economic performance indicators, which indicates the linkages between the two forms of analyses.

6.0 Empirical Results

The results, shown below, cover the intercity bus and airline industries defined as passenger carriers and the rail, trucking and water transport industries defined as freight carriers. It excludes government carriers, such as Marine Atlantic or Via Rail. Urban carriers such as transit systems or taxicab operations or special services such as school bus operators are also excluded. So were charter operators and small carriers in the air and trucking industries. The carriers analyzed in this paper represented about 72% of the gross revenues of business transport services.

6.1 Revenues, Output and Price Indicators

The transportation sector represents a considerable business activity in Canada. In 1993, the gross revenue of transport industries amounted to $23 billion. As shown in the opposite figure, the output pattern of the transport sector follows the cycles of the business sector of the economy[①] However the cycles seem more accentuated in transport. In addition, the growth of transport trailed by a considerable margin that of the rest of the economy, reflecting the maturity of transport services. Had transport industries' revenues grown at the same pace as the rest of the economy, in 1993, the revenues of the sector would have been 27% ($ 6.3 billion.) larger.



The mediocre performance of the transport sector is particularly evident since 1987, as can be seen in Table 1, next page. The prices of services are flat especially in freight. Output grows at a pace which is less than half of the growth rate of the Canadian economy.

① *From now on, the term "economy" will be used to refer to the business sector.*

Table 1
Price and Output Indicators of Major Transport Industries (Selected Years)

	1981	1984	1987	1990	1993	AAG in
Revenues in $ Million						
The Intercity Bus Industry	400	471	505	572	598	3.4
The Airline Industry	3,919	4,310	5,287	6,634	6,083	3.7
CN and CP	5,380	6,379	6,513	5,872	5,894	0.8
Trucking	4,720	5,593	7,429	7,821	9,122	5.6
Water Transport Industry	1,399	1,178	1,152	1,242	1,109	-1.9
Major Transport Industries	15,818	17,930	20,886	22,141	22,806	3.1
Price Indices (1986=100)						
The Intercity Bus Industry	73.4	89.5	101.6	111.3	117.9	4.0
The Airline Industry	85.3	93.0	103.5	114.3	121.9	3.0
CN and CP	84.7	93.9	101.4	100.1	93.2	0.8
Trucking	84.0	92.4	96.9	99.9	103.6	1.8
Water Transport Industry	92.7	94.6	99.2	116.5	127.0	2.7
Major Transport Industries	84.9	93.2	100.2	105.0	108.1	2.0
Economy	*83.8*	*97.9*	*103.4*	*114.0*	*115.3*	*2.7*
Output Indices (1986=100)						
The Intercity Bus Industry	112.3	108.2	102.4	105.8	104.4	-0.6
The Airline Industry	90.6	91.4	100.7	114.5	98.4	0.7
CN and CP	101.9	109.0	103.0	94.1	101.5	-0.03
Trucking	82.1	88.4	111.9	114.4	128.6	3.8
Water Transport Industry	138.6	114.4	106.6	97.8	80.1	-4.5
Major Transport Industries	95.5	99.0	104.6	105.4	105.5	1.0
Economy	*87.8*	*91.4*	*105.3*	*114.7*	*117.3*	*2.4*

Within the transport sector, the strongest output performance is shown by the trucking industry, its output advancing by 62% (3.8% a year) between 1981 and 1993, compared to 27% (2.4% annually) for the economy. A strong elasticity of 1.65[2] can be measured between trucking and the economy. The worst performer is the domestic shipping industry, its output declining by 42%. The bus and rail industries also exhibit declining output, however rail output rebounded in 1993, with a strong 4% growth. Airline output peaked in the late

[2] *This estimate was derived from a log linear model which explained 97% of the variation in trucking output.*

80's, its 1993 output being 16% lower than the 1989 level, but still 9% higher than in 1981.

For the users of transport services, the rail industry offered the best deal, its prices increasing only by 0.7% a year and in fact in nominal decline since 1987. This represents a 20% drop in real terms. The pace of increase in trucking prices doubled that of rail, still a 13% drop in real terms. (The truck-rail differential is reduced considerably when the prices of long haul and high volume services are considered). The shipping industry showed price increases similar to those in the economy. In fact, together the three freight modes registered price increases half of those in the economy. Both the bus and air industry posted annual price increases, 4.0% and 2.8% respectively, higher than in the economy. Overall, transport prices fell 0.7% annually between 1981 and 1993, an average savings of $ 2 billion since 1988 for transport users.

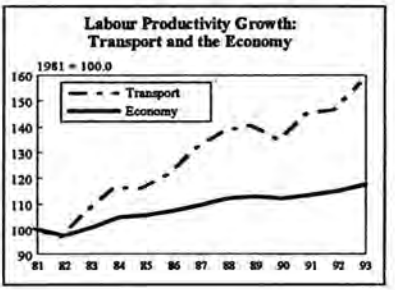
6.2 Productivity Trends

Total productivity of the transport sector in Canada grew by 15% over the 1981-1993 period (1.1% per annum). Since the mid-Eighties, productivity has been trending downwards in passenger carriers, while the gains of freight carriers have been accelerating since 1986. (See Table 2 for more detail)

Table 2
Total Factor Productivity Growth in Transport (AAG in %)

Period	1981-86	1986-93	1981-93
Passenger Carriers	0.3	-1.7	-0.8
Freight Carriers	1.9	2.2	2.1
Total transport	1.4	1.0	1.1

Comparisons between the transport sector and the economy are limited to labour. The transport sector exhibits a much stronger productivity growth (52%) than the economy (17.5%). The opposite graph provides evidence of the changes on a yearly basis. In spite of higher salary increases in transport, the sector's growth of unit labour cost was only a third of that of the economy (1.3% vs. 3%).



At the industry level, (Table 3, next page, for detail), the best productivity performance is shown by the rail industry, with an annual growth of 2.8%.[3] The pace of productivity gains by CN (3.6%) was stronger than CP's (2.9%). These results are in conformity with those of Tretheway&Walters. While we suspect that CP's output growth is understated and CN's overstated, the magnitude of the differences is such that they cannot be the result of estimation errors.

The air industry had the worst performance, its productivity falling by 1% a year since 1981. However its the productivity has been advancing by 2.3% a year between 1991 and 1994, which would indicate that part of the difficulties encountered in the mid-80's were due to the restructuring of the industry. The gains were not evenly distributed between Canadian and its partners and Air Canada and its affiliates. Finally, replacing our output measures by an index of passenger seat-kilometres would have yielded a productivity growth of one percent a year, which would indicate that the productivity fall between 1988 and 1991 was due in part to lowered load factors.

Until the late-Eighties, the productivity growth of trucking industry was the strongest of all transport industries. However, these gains seem to have stalled either in 1989 or 1990. What is unique is that this halt is common amongst all industry segments (by region, small & large, specialized & general, etc.) and is also reflected on individual factors of production.

The productivity of the water transport industry has been volatile, peaking in 1988 and dropping considerably (11%) in the last two years. This may be due to estimation problems of output changes in the industry. The downward trends in bus productivity are more evident. However, the charter segment of the industry shows a more positive trend.

For the whole transport sector, the capital productivity growth is slightly above that of the variable (labour, fuel, etc.) factors. This corresponds to a reduction in the capital output ratio. By industry, the bus and airline industries exhibit strong negative capital productivity trends, which suggest major excess capacity problems. The capital productivity gains in the shipping industry reflect the erosion of the asset base in that industry. The rail industry show the strongest gains (3.7% per annum) in variable productivity, followed by the trucking annual gains of 1.5%. The variable productivity of the airline industry is stable, in spite of solid gains in labour and fuel productivity, indicative of trends in outsourcing in the industry.

---

[3] *The productivity growth of the railways has an upward bias created by the massive buyout of 1991-92 which produced a labour productivity gain of 25% between 1991 and 1993. The cost of these buyouts ($1.4 Billion) was not accounted for in the analysis.*

Table 3
Productivity Indicators of the Transport Sector – Selected Years

	1981	1984	1987	1990	1993	AAG in %
Variable Productivity (1986 = 100.0)						
The Intercity Bus Industry	100.7	111.2	99.2	99.4	102.9	0.18
The Airline Industry	97.1	104.7	97.1	96.7	95.0	-0.17
CN and CP	86.3	101.2	106.5	112.4	133.6	3.71
The Trucking Industry	94.3	101.4	106.0	114.9	112.8	1.50
Water Transport Industry	95.4	107.0	112.3	102.1	92.7	-0.24
Major Transport Industries	93.2	103.9	103.1	105.1	106.2	1.10
Capital Productivity (1986 = 100.0)						
The Intercity Bus Industry	99.8	106.9	101.1	82.0	68.7	-3.06
The Airline Industry	106.0	86.8	88.5	67.6	54.7	-5.36
CN and CP	106.0	111.0	98.5	97.2	104.0	-0.16
The Trucking Industry	86.0	103.3	102.3	96.7	103.3	1.54
Water Transport Industry	79.0	100.8	105.1	124.7	131.3	4.32
Major Transport Industries	94.6	97.7	103.0	104.5	109.7	1.24
Total Factor Productivity (1986 = 100.0)						
The Intercity Bus Industry	100.5	110.6	99.4	97.0	97.0	-0.29
The Airline Industry	98.1	101.7	95.8	91.8	87.4	-0.96
CN and CP	89.9	103.3	104.6	108.5	125.9	2.84
The Trucking Industry	93.4	101.6	105.7	112.9	111.8	1.50
Water Transport Industry	95.6	105.5	114.8	108.7	101.7	0.52
Major Transport Industries	93.3	102.9	103.1	105.0	106.8	1.13
Input Price Index (1986 = 100.0)						
The Intercity Bus Industry	76.4	100.5	98.8	109.5	114.4	3.42
The Airline Industry	85.6	99.1	99.1	110.5	108.3	1.98
CN and CP	71.5	92.1	102.7	114.9	121.1	4.50
The Trucking Industry	80.2	94.9	102.3	112.7	115.9	3.12
Water Transport Industry	83.4	99.7	105.0	118.7	122.5	3.25
Major Transport Industries	74.8	95.7	102.0	113.8	116.4	3.35
Unit Cost Index (1986 = 100.0)						
The Intercity Bus Industry	76.0	90.8	99.4	112.9	117.9	3.73
The Airline Industry	87.3	97.5	103.4	120.3	124.0	2.97
CN and CP	79.4	89.2	98.3	105.9	96.2	1.61
The Trucking Industry	85.8	93.4	96.8	99.8	103.7	1.59
Water Transport Industry	87.3	94.5	91.4	109.2	120.4	2.72
Major Transport Industries	84.0	93.0	98.9	108.3	109.0	2.20

6.3 Unit Cost Trends

Unit costs are influenced by the behaviour of productivity and factor prices. Input prices increased at an annual rate of 3.4% a year over the 1981-1993 period in the transport sector, following closely the rate of inflation in the Canadian economy during the same period (3.5%). The air industry faced the lowest input increase (2% per annum), as a result especially of falling leasing prices, which are affected directly by lower interest rates and equipment prices. The drop in fuel prices was also a factor. The rail industry had the highest input price increase (4.5% a year). Salary cost was a major factor, as well as the cost of their owned capital. ④

Total unit cost of the transport sector advanced by about 2.2% a year. There was a significant increase in the early half of the 1980's (3.5% per annum), but the pace decelerated considerably during the 1986-1993 period (1.2%). Over the 1981-93 period, the unit cost performance of the rail and trucking industries was the best in transport and essentially the same (1.6% annual growth). However, since the mid-eighties, the rail industry has outperformed trucking by a 7.5 percentage point margin. The bus industry has had the highest unit cost increase (3.7% a year). The airline industry recorded an annual 3% increase in its unit cost. However, between 1991 and 1994, the industry has made a major turnaround by reducing its unit costs by 8.5%. Unit costs in the shipping industry have been growing steadily at 2.8% a year.

6.4 Financial Indicators

The transport sector did not start the 80's on a strong financial position. In 1981, the earnings of the sector were $435 million shy of its economic cost of capital. The recession of 1982 exacerbated the situation by raising the cost of capital to new highs and reducing transport demand by almost 10%. As a result, the operating margin shortfall (of its imputed cost of capital) reached a staggering $1.1 billion or 7.4% of the revenues of the sector. The sector's operating margin was equivalent to only 40% of interest payments on its debt. The situation did improve up in the mid-80's, but the prolonged 1989-92 recession plunged the transport sector into a profound morass culminating in 1992 with operating ratios reaching almost 100%. However, 1993 marked the beginning of a turnaround, at least for rail. In 1994, the situation continued to improve for rail and the airline industry recorded positive operating margins after four consecutive years of operating losses. (For more detail, see Table 4, next page)

---

④ *Our approach has generated intermodal differences in the capital price index which seem excessive. One factor is the price index of leased capital which falls over time and thus affects negatively the productivity of industries making an extensive use of this capital source, such as the airlines. The differences in the price index of owned capital have a minor effect on all modes, except in the rail industry where the productivity growth would be reduced by 7% with a simulation of the average service price of capital in transport. The productivity growth of the air industry would gain 2%.*

Table 4
Financial Indicators of the Transport Sector – Selected Years

	1981	1984	1987	1990	1993
Imputed Cost of Capital/ Fixed Assets (%)					
The Intercity Bus Industry	16.4	15.2	12.7	13.1	11.6
The Airline Industry	13.4	12.9	11.4	10.9	9.1
CN and CP	9.5	10.8	10.8	9.9	8.9
The Trucking Industry	15.0	13.5	12.6	12.3	11.5
Water Transport Industry	12.1	12.1	9.8	9.9	8.0
Major Transport Industries	11.5	11.9	11.2	10.5	9.2
Revenue Adequacy (Required Operating Margin in % of Operating Revenues)					
The Intercity Bus Industry	7.9	7.6	6.5	7.1	6.9
The Airline Industry	7.9	10.3	7.9	7.1	6.2
CN and CP	10.3	12.4	14.7	15.4	13.4
The Trucking Industry	4.3	3.5	2.7	2.9	2.5
Water Transport Industry	10.5	12.6	8.5	7.0	5.0
Major Transport Industries	7.8	9.1	8.5	8.0	6.6
Operating Ratio (%)					
The Intercity Bus Industry	97.1	95.3	92.7	95.8	94.5
The Airline Industry	97.1	97.2	94.6	100.8	101.5
CN and CP	91.6	86.3	86.0	94.4	93.8
The Trucking Industry	97.2	97.1	96.7	96.7	97.2
Water Transport Industry	93.4	103.6	96.0	97.1	101.9
Major Transport Industries	95.0	93.4	92.6	97.5	97.5
Operating Margin Deficiency In $M (Cost of Capital - Operating Margin)					
The Intercity Bus Industry	20.2	13.7	(4.1)	16.6	8.5
The Airline Industry	195.0	323.5	132.4	528.1	458.3
CN and CP	95.4	(83.2)	41.5	576.5	428.3
The Trucking Industry	72.7	(1.2)	12.1	52.2	(44.1)
Water Transport Industry	54.5	190.9	51.4	50.4	76.6
Major Transport Industries	437.8	443.8	233.4	1,223.8	927.6

The industry added $6 billion to its fixed assets between 1981 and 1993. Because of limited profits or financial losses, this expansion of the asset base was financed mostly by debt, which raised the percentage of assets from 46 to 53%. The increased leverage of the transport sector, and lower interest and taxation rates have contributed to the decline in capital cost rate, from an estimated 11.5%

of assets in 1981 to 9.2% in 1993. The strongest declines were in industries (bus, air and water) which experienced rising debt-equity ratios. However, the transport sector has become more vulnerable to market swings

To meet the economic cost of capital, the transport sector required a profit margin of less than 7% in 1993 down up from 7.8% in 1981. In other words, the operating ratio should be 93%. Significant differences exit within the transport sector. The rail industry requires by far the highest operating margin at 13.4%. This is understated, because CN, as a crown corporation, was assigned a return on equity equal to the cost of its debt. The lowest rate (2.5 to 3%) is in trucking, because of high turnover rates of 4.1 in 1993 up from 3.8 in 1992. We suspect in this case some data reporting problems. The airline industry has a relatively low operating margin requirement of 6.9%. This is due to the fact that the industry makes an extensive use of leasing as a source of capital. A more precise measure is the service price of capital owned and leased. In the airline industry this ratio reaches 18.5% of its fully allocated costs, compared to 23% for rail.

The transport sector has been far from meeting its required margins, or in other words, to earn an economic return to capital. The margin shortfall has averaged 4.1% of the operating revenues of the sector. This shortfall, which can be considered as a transfer form shareholders to transport users, reached $1.6 billion or 7.2 % of the sector revenues in 1992. For the air and rail carriers, the margin deficiency ($775 and $ 675 million respectively) reached crisis levels in 1992. For the air industry, "healthy" price increases were not sufficient to offset the stronger growth in unit cost. In rail the small increases in unit costs were still excessive for an industry which since 1986 has had to face declining output prices.

The trucking industry achieved operating ratios higher than the industry average. Yet, on average, It is has been close to meet the cost of the its debt and yield an adequate return for the shareholders, because of a small asset base relative to its revenues. At the regional level, Atlantic firms pass the test of revenue adequacy in all the years, but one, of the period analysed. Ontario and larger firms show high margin deficiency rates especially in the latter years. Overall, the financial position of the trucking industry has been relatively stable over the period analysed, essentially because the industry has been able to balance its costs with the price of its services.

The pace of acquisition of new assets in the bus industry was surprisingly high for an industry which is not showing a good market potential. As a result, the asset turnover rate is second only to rail in the transport sector, which means that the bus industry requires an operating margin above the transport average to meet the cost of its capital. In the 80;s, the industry did improve its profitability for a while, mostly through price increases superior to its input costs. This was short lived. As productivity dropped in the second half of the 1980's, operating ratios

climbed abruptly from 92% in 1988 to 96% in the following years up to 1992. The increase in operating ratios was such that operating margins barely covered the estimated interest costs of the industry, compared to a ratio greater than 2 in 1982. In 1993, however, a combination of falling unit cost and slight price increase produced a drop in operating ratio such that the industry interest coverage exceeds again 2 and margin deficiency is reduced to 1.4% of the industry operating revenues.

The financial performance of Canadian water carriers had a lot to be desired for the period under review. In addition to the decline in their asset base, they have increasingly become dependent on debt financing (nearly 59% in 1992). The industry achieved operating ratios less than 100% only in 1981 and in the 1987-91 period. On average, yearly industry earnings have been $108 million short of earning an adequate rate of return. This represents 8.8% of industry revenues throughout the 1981-93 period.

7.0 Conclusions

The Canadian economy has benefited in the 80's and the early 90's from a relative reduction of transport prices, especially in freight. These reductions have been made possible by lower unit costs increases in transport than in the economy. However, another half of the reduction consisted of a transfer ($1 billion a year between 1988 and 1993) from shareholders to transport users. This strained the financial viability of the carriers. As the economy picked up in 1993 and 1994, a turnaround is evident in those two years. The slowdown in 1995 and its negative impact on transport are indicative of the increased sensitivity of the transport industry to economic cycles.

The reduction of economic regulation and the devolution of services outside the public domain mean a shift of the government role to that of oversight and policy. In this context, the issue of performance of the transport sector will take on an increasing importance. We must avoid the facile alternative of using indicators simply because they are readily available. The development of indicators must be supported by a rigorous analytical framework, such as the one offered by the concept of total factor productivity. The disadvantage of this approach is the availability, timeliness, quantity and quality of the data required for these measurements. The data problems which we alluded to in this paper showed that it is not a minor issue. There are also methodological questions, such as the price of capital, which will need to be addressed, but their resolution will be constrained by data availability.

References

1 Annual proceedings of the Canadian Transportation Research Forum, 1993.

2 Evaluating Surface Freight Transportation Rates, 1981–1992, Economic Research, November 1992. This working paper has been updated for 1993.

3 The Canadian Airline Industry, Price and Output Measurements, Raynald Ouellet, Working Paper, Economic Research Branch, Transport Canada, November 1995.

4 Aggregation and Accuracy in Measuring Total Factor Productivity, Evidence from Rail Productivity Studies, W. Walter and M. Tretheway, Annual Proceeding of TRF, 1995.

5 In "Measuring Productivity Performance and Financial Returns for Canadian National Railways, 1956–1991, M. Tretheway, W. Walters, University of British Columbia, 1994", the authors proposed a more comprehensive method of calculating the price of capital which includes property tax and capital gains. These data were not available for other modes and we choose for consistency purposes to ignore them, even for rail.

6 Same source as above.

7 Rail Material Input Price and Quantity Indices, B.R. Khanam, Working Paper, Economic Research Branch, Transport Canada, 1992.

Statistical Sources

Statistics Canada' surveys of transport industries

Bank of Canada's Monthly Reviews

Major Carriers' Annual Reports

Informetrica

Centre for Transportation Studies, University of British Columbia

Transport Canada's data bases

Acknowledgements

The author wishes to thank the staff of the Economic Research Branch, Hiroshi Kurasawa, Joe Kruger, Raynald Ouellet, Christian Beauregard and Peter Kennedy and Vivian Wei.

An Overview of the Short Line and Regional Railway Industry in the United States

by Andrew A. Jennings*

As a major restructuring of the two major Canadian railway systems progresses, one of the restructuring issues facing Canada is the future of its light density rail lines. In the U. S., where the railway industry is not as concentrated as in Canada, many of the lighter density lines have been sold to newly formed short line and regional railways. Expansion of the limited program of line sales by Canadian National (CN) and Canadian Pacific (CP) to short lines and regionals is one option for dealing with the light density line issues. This paper provides an overview of the short line and regional railway industry in the U.S. to facilitate discussions on the future of light density lines in Canada.

Through their restructurings, CN and CP are attempting to become more competitive by reducing their costs to levels similar to U.S. railways. In 1994, the CN and CP had average costs[1] of 2.32 cents per revenue ton mile[2], 9.2% above the average costs of the twelve U.S. Class I[3] railways. A major driver of the cost differences between CN and CP and U.S. Class I carriers is the difference in density of operations[4]. The U.S. Class I railways increased their average density by 76 percent in the decade between 1985 and 1994, the latest 10 years for which data is available. The density increase comes from two components: a 38 percent increase in revenue ton miles, and a 21.5 percent decrease in miles operated. Key to the reduction in miles was the sale[5] of lines to short line and regional railways. Although the Class Is reduced their mileage by 21.5 percent, the U.S. lost only approximately 10 percent of its rail route mileage in the same decade[6] because of the growth of short lines and regionals.

---

* Andrew A. Jennings is a Senior Consultant to Mercer Management Consulting, Inc., and Director of Market Development for Rail Systems, Inc. of Hollis N. H.

[1] Excluding the U.S. Class I subsidiaries of CN and CP, Grand Trunk Western (GTW), and SOO Line (SOO)

[2] Unless otherwise indicated, U.S. dollars and English units of measurement are used throughout this paper.

[3] Class I carriers were those with revenues exceeding $255.9 million in 1994

[4] In terms of revenue ton miles per route mile of track, CN and CP have less density than ten of the twelve major U.S. roads, and have less than half the density of UP.

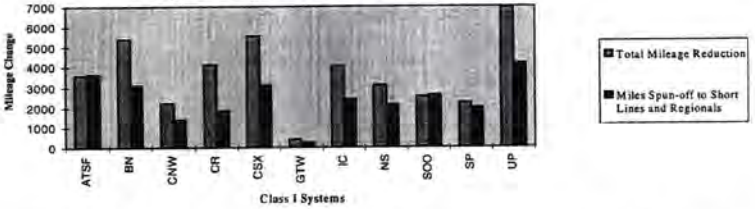
[5] The term sale used in this paper also includes the long term lease of lines to new carriers.

[6] Because federal reporting requirements for small railways in the U.S. ended after 1979, consistent reporting of the number of carriers and miles operated is not available. The best source for such information comes from a voluntary survey of U.S. railways done by the Association of American Railroads.

Canada saw a greater reduction in rail mileage than the U.S. during that decade as the mileage of main lines and branch lines was reduced by almost 15 percent. Traffic grew at half the rate of the U.S., a 19 percent increase in revenue ton miles. The sale of track by the major Canadian railways was far less aggressive than that of U.S. carriers, with only eight transactions in the 1985-1994 decade.[7]

Mileage of eleven[8] of the twelve Class I U.S. railways shrunk in the 1985-1994 decade[9]. Exhibit 1 compares the mileage spun-off with the total mileage reduction of each carrier. Some of the spun-off mileage was subsequently abandoned by the acquiring carriers. This exhibit clearly shows that spin off to short line and regional railways has been a major part of line reduction efforts of the Class I railways.

Exhibit 1
Change in Mileage 1985 - 1994 U.S. Class I Railways



Comparing the U.S. and Canadian experience, it is easy to speculate that if Canadian roads had pursued a more aggressive policy of line sales, less of the Canadian rail network would have been abandoned, a somewhat larger volume of traffic would be currently handled on the Canadian rail network, and the CN and CP would be more efficient, lower cost carriers.[10] Hypothetically, an aggressive program of spin-offs of lighter density lines to short line and regional railways in Canada would achieve these benefits by:

- Preserving lines which the major carriers would otherwise abandon, and thereby providing feeder traffic to the major carriers that would otherwise be lost, thus increasing densities on the major carrier's main lines
- Providing opportunities for investment in lighter density lines which would not occur under ownership by the major carriers
- Injecting an entrepreneurial management into the management and operations of lighter density rail lines

---

[7] In 1995, three short lines were formed from what was CP's Sherbrooke - St. John corridor.
[8] Kansas City Southern is the only Class I that has not shrunk.
[9] Comparable systems, including merger partners, were compared. CSX includes C&O, B&O, and SBD; SOO includes MILW; SP includes DRGW and SSW; UP includes MKT, MP, and WP.
[10] Differences in the regulatory environment and labor laws of the two nations may have effectively discouraged CN and CP from pursuing this strategy in Canada.

STRUCTURE OF THE U.S. RAILWAY INDUSTRY

Table 1 provides a breakdown of the U.S. Railway Industry[11]. The short line and regional segment is the third largest segment, behind the twelve Class I freight railways and the passenger segment. The short line and regional railways generate a little more than 10 per cent of the Class I revenues, using a route structure that is more than 35 percent as long as the Class Is, and employing a work force approximately 12 per cent the size of the Class Is. Private railways and in-plant switching operations play an important role in the U.S. freight industry, perhaps as large an economic contribution as the short line industry. Unfortunately, good statistics on the private railways and in-plant switching operations are not available. A few short lines have tourist passenger operations, thus blurring the distinction between short lines and operating rail museums, the smallest segment of the industry.

Table 1

Structure of U.S. Railway Industry (1994)

		Number of Operators	Mileage	Revenues $million	Employment
Freight Railways					
Class I		12	123,335	$29,931	189,240
Regional		32	19,842	$1,745	10,701
Short line	a	487	25,599	$1,422	13,070
Private / Industrial		1,622 b	c	Not applicable	Not available
Passenger Railways					
Amtrak	d	1	25,146 e	$2,077 f,g	24,991
Commuter Rail	h	14	2,817 e	$2,081 f	21,934
Museum / Preservation	i	83	791	Not available	j

Notes

a Statistics are for those carriers that voluntarily report. There are probably an additional 30 carriers that did not report, and other carriers that are inactive
b Number of locations with Industrial Operations
c Unknown, but a 1991 survey identified more than 5,400 miles of track operated at approximately 825 locations
d FY 1994
e Most miles are also served by freight operators
f Costs are used as a size measure because revenues do not approach cost coverage
g Includes costs associated with contract operations, including commuter contracts
h 1993 statistics
i Operating museums only; excludes freight operators listed above
j Not available; extensive volunteer involvement should be recognized

[11] Detailed information is readily available only on the Class I carriers, Amtrak, a few publicly traded regional railways and short lines, and the commuter rail carriers. The other figures are based on surveys, estimates, and information provided by sources which are not necessarily consistent with each other.

While short lines and regionals are a small segment of the U.S. railway industry, they are a significant industry with more route miles than all the rail route mileage in Canada. They generate about 60 percent of the revenues of the Canadian railways (excluding U.S. subsidiaries), and employ slightly under half of the number of Canadian railway employees (excluding VIA). The maps in Exhibit 2 compare the geographic scope of short line and regional railways with the twelve Class I railways. Certain areas of the U.S., including Alaska and northern New England, receive rail service only from short lines and regionals.

In contrast to the U.S. structure, the two major Canadian railways generate a higher proportion of Canadian railway industry revenues than the twelve major U.S. carriers. Carriers[12] other than CN and CP generate only 8.5 percent of Canadian freight revenues, operate 13 percent of the main and branch line trackage, and employ 11 percent of the employees.

U.S. RAILWAY INDUSTRY EVOLUTION

While CN and CP have been the two major rail carriers in Canada since the 1920's, the U.S. railway industry has continued to evolve. The large carriers have grown, creating more concentration in the industry, while simultaneously new railways have been created. The seven largest U.S. carriers in 1968 generated approximately 58 percent of the industries revenues; in 1994 the top seven generated 84 per cent of the freight revenue[13]. The concentration of carriers continued in 1995 with the mergers of BN and ATSF and of UP and CNW, leaving only 10 Class I carriers.

While the larger railways have been consolidating, new railways have been created. A consistent count of carriers is not available, but a reasonable estimate is that the number of carriers has grown from about 425 in 1968 to around 550 in 1994[14]. Far more than 125 new carriers were created during that time frame[15]. Many of the carriers operating in 1968 (and some formed since 1968) have been consolidated, reorganized under new names, or liquidated.

The creation of new railway companies in recent decades has been driven by two principal motivations: a desire to preserve service that otherwise would have been lost through abandonment of lines, or a desire to improve returns by divesting lines that are not meeting corporate objectives.[16]

---

[12] Excluding VIA and the Canadian commuter rail operations.

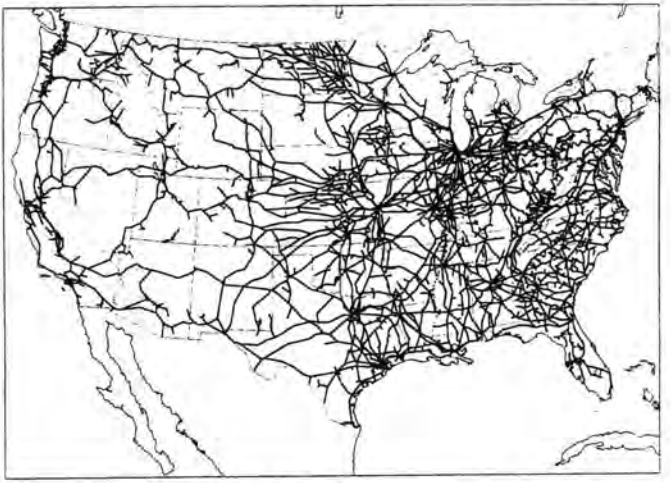
[13] In 1968 most of the major U.S. carriers operated passenger service in addition to freight service.

[14] Excludes carriers owned by a single Class I. Also excludes 35 additional railways which do not appear to be operating currently. Some of the short lines and regionals considered as active may also have limited operations.

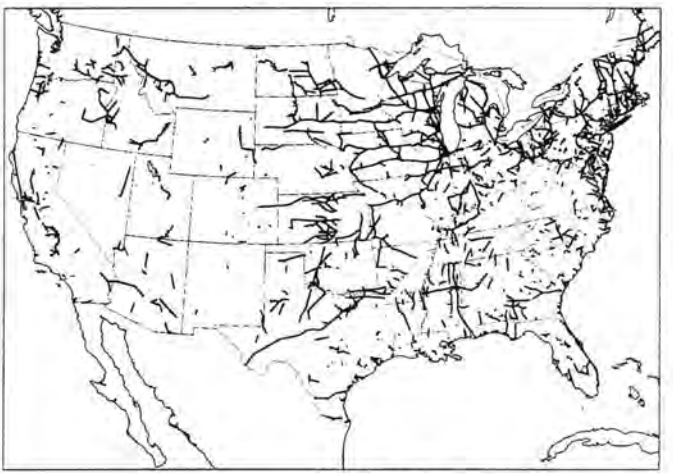
[15] From 1985 through 1994, U.S. Class I carriers were involved in 242 line sale transactions.

[16] Some of new railways were created to serve new mines, utilities, or industries, or by the conversion of private switching operations to common carrier operations.

Exhibit 2
Comparsion of Class I Routes with Short Lines and Regionals



Class I Routes



Short Line and Regional Routes

Note: Maps prepared by MultiModal Applied Systems, Inc., 347 Elizabeth Avenue - Suite 105, Somerset, NJ 08873

The first motivation, a desire to preserve service that would have been lost through abandonment of lines, was readily apparent in many lines that were created in the 1970s. Following the bankruptcy of most of the major railways in the U.S. northeast and the subsequent formation of Conrail, as well as the bankruptcy of two midwestern roads, (Rock Island and Milwaukee Road), many communities were faced with the loss of rail service. Creation of short line and regional railways that could operate at lower cost and included a local development emphasis appeared to be attractive alternatives to loss of rail service. Public-private partnerships were developed to assist in the formation of these lines.

An example of a line that was created to preserve rail service is the Livonia, Avon and Lakeville. In 1964, local interests purchased a 24-mile line serving the namesake towns in central New York state to preserve rail service to two struggling feed mills. Within two years of the acquisition of the line, the feed mills failed. The new railroad demonstrated the ability of a short line to survive difficult circumstances. In 1968, railway operating revenues were $19,250, of which $18,076 came from an excursion passenger operation it had established. Its operating ratio that year was 153 percent. Subsequently, industrial development efforts landed a significant, new industry. Today, the railroad is a prosperous 8-mile line handling approximately 1,500 carloads per year.

Not all new short lines have been successful. For example, in 1974 the Illinois Central (IC) decided to sell a rail line, Louisiana Midland, which IC had acquired as an imposed condition to a merger with another carrier. Sale of 67 miles of track, three locomotives, forty freight cars, and surplus maintenance of way material for the sum of $1,500[17] was considered a bargain by IC as it avoided the significant costs of an abandonment proceeding that would surely have been contested. With entrepreneurial management, success seemed assured. Unfortunately, by 1987 the line had died with hardly a whimper.

The second motivation, the desire to improve returns by divesting lines that are not meeting corporate objectives, appears to have motivated most of the line sales made by the Class I railway carriers in the late 1980s and early 1990s. A line may not meet corporate objectives for a variety of reasons. Table 2 outlines some of those reasons and the perceived advantages that a new short line or regional might have.

When the carrier selling a line is motivated to improve its returns, it is also interested in continuing to handle the traffic originating or terminating on the line sold. A major factor in selection of a buyer is whether the buyer can

[17] No zeros have been left out. The cash consideration was only fifteen hundred dollars.

Table 2	
Major Carrier Problem	Short Line / Regional Solution
Investment in line (e.g. rehabilitation) does not meet corporate investment hurdles	Operate at lower cost so investment can be justified
	Focus of particular investment to find lower cost solution; solution may involve higher operating costs
	Use lower investment hurdles
	Attract new capital to line (often a local operation can attract government funds not available to major carriers)
High labor costs on a particular line	Use simpler work rules
	Pay lower wage rates (often local wages are below nationally negotiated rates)
	Provide pay systems with greater incentives
	Provide improved working conditions (regular calls, no assignments away from home)
	Reengineer operations to better meet customer needs
	Provide closer supervision
	Cross-train employees
	Contract out services that can be more effectively provided by outsiders
High equipment costs	Use used equipment
	Match equipment more closely to line needs (the optimal locomotives on a large system may not be the optimal equipment for a specific line)
	Improve utilization
Lack of growth	Focus marketing to improve growth
	Entry into markets not practical for major railways (for example, tourist trains or two-car sludge trains)
	Improve match of customer needs to service
High taxation costs	Local governments are often more responsive to needs of locally headquartered companies

improve the satisfaction of customers on the line. Surveys have generally indicated that most short line and regional buyers have been able to provide service and rates that are as good or better than the sellers.[18]

SHORT LINE AND REGIONAL RAILWAYS: A PROFILE

With more than 500 short lines and regionals operating in the U.S., there is wide variation among the lines. A precedent exists for almost any model for developing a short line or regional railway.[19] Significant variations exist in length, traffic density, revenue, employment, ownership, and management style. In looking at the range of variations, however, important differences emerge between the lines that were spun off between 1985-1994 and those with a longer history, and those that were created by reorganization of older short lines. Approximately two-thirds of the short lines and regionals in the U.S. are operated independently of their customers or government. Often, multiple short

[18] See, for example, State Short Line Railroads and the Rural Economy, or A Survey of Shipper Satisfaction with Service and Rates of Shortline and Regional Railroads.
[19] For a complete profileof short lines as of the early 1980s, see Small Railroads.

lines and regionals operate under common control of a holding company[20]. The proportion of lines owned by governments and shippers continues to fall while the number of short lines and regionals expands because few of the spin-offs of operations by the larger railways have been to either shippers or governments. Most of the short lines and regionals are closely held companies.[21]

Railways owned by customers are becoming less common in the U.S. Several short lines in the Western U.S. that were owned by shippers closed when their principal source of traffic closed. Other lines were sold when the parent company downsized. Packaging Corporation of America and Stone Container are examples of shippers that sold their short lines. At the end of 1994 shipper-owned lines amounted to approximately fifteen percent of the total.

Direct public ownership of an operating rail company is not as common in the U.S. as it is in Canada. There are approximately 30 lines which are owned and operated by port authorities, other special purpose authorities, cities, and states. The largest freight operation owned by a government agency in the U.S. is the Alaska Railroad, at just under 500 miles in length and with more than $40 million in freight revenues. With two exceptions[22], the spin-offs by Class I railways in the 1985-1994 decade have not been to government-operated companies. A more common model involves public-private partnerships, in which a government agency buys a rail line and bids out the operation to potential freight operators. This model has been used to preserve lines that were under threat of abandonment.[23] The model has also been used to establish joint freight and light rail transit operations in the San Diego and Baltimore areas, and to preserve corridors for future transit development.

Exhibit 3 gives a distribution of short lines and regionals by line length. Three regionals, Wisconsin Central, Delaware and Hudson, and Springfield Terminal have more trackage than the smallest Class I, Grand Trunk Western. However, most of the short lines and regionals are limited in scope. Approximately 45 percent of the companies are 25 miles or less in length and more than 60 percent of the companies are 50 miles or less in length. The Class I spin-offs in the last

---

[20] The American Shortline Railway Guide identifies 59 groups of short lines under common control. Railtex at the end of 1994 operated the most carriers (26), but this has dropped to 24 in 1996.

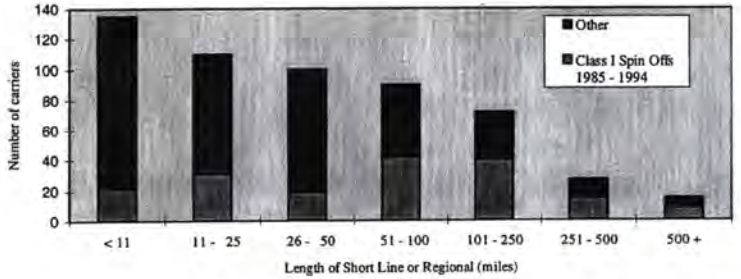
[21] Florida East Coast is traded on the New York Stock Exchange. Holding companies Delaware Otsego, Emmons Holdings, Rail America, and Railtex are traded on the NASDAQ, along with Wisconsin Central, and Providence & Worcester. Pioneer Railcorp, another holding company is traded on the Chicago Stock Exchange.

[22] Columbia Terminals and Semo Port Railroad. Both carriers combined operate 31 miles of line and handled fewer than 1,000 carloads in 1994.

[23] This model was used by the State of Vermont when the Rutland Railway collapsed in the early 1960s, by the states of Massachusetts, New York, Pennsylvania, Ohio, Michigan, and others following the creation of Conrail, and in Idaho, Montana, Michigan, South Dakota, and Washington after the bankruptcy of the Milwaukee Road.

decade have been generally larger than other carriers, averaging 130 miles in length while the other carriers average 65 miles in length.

Exhibit 3
Number of Short Lines and Regionals by Length of Line



More than 80 percent of the short line and regional railways reported under $5 million in annual revenues.[24] Using crude estimation techniques based on carloads and average haul, it is likely that half of the short lines and regionals have revenues under $1 million annually.

Short line and regional railways tend to be individually dependent on a few key commodities[25]. For approximately 25 percent of short lines and regionals, 90 percent or more of their carloads come from a single commodity (at the two digit STCC level). For approximately two-thirds of short lines and regionals, at least half of their carloads come from a single commodity. When the commodity mix of short lines and regionals is compared with the mix of the large Class I carriers, certain significant differences appear. The Class I carriers carry proportionately much more intermodal traffic and transportation equipment (largely automobiles and auto parts) than their short line and regional counterparts. Food products and chemicals are also more important for the major railways than the short lines and regionals. Conversely, metallic ores are much more important to the short lines and regionals[26] than they are to the Class I carriers. Grain, forest products, paper products, petroleum products, and steel are also more important to the short lines than the Class I carriers.

---

[24] Financial information since deregulation is not readily available. Most carriers voluntarily report their revenues to the Association of American Railroads, but that information is only disclosed as being in a wide revenue range.

[25] Specific commodity information is also not readily available since deregulation. Some commodity information has been voluntary reported and distributed by the AAR. Approximately 500 short line and regional carriers have reported total carloads and the percent of those carloads in each of the three top commodity groups at the two digit Standard Commodity Code Classification (STCC) level.

[26] Iron ore follows the same pattern in Canada. It is the largest single commodity hauled by Canadian railways, but virtually all of that commodity is handled by the iron ore railways in Quebec.

Short lines and regionals are not only individually dependent on a few key commodities and a few key customers, they are also very dependent on traffic interchanged with their connections. Approximately 500 carriers reported to the Association of American Railroads (AAR) the percentage of interchange traffic. For more than 70 percent of the carriers reporting, all their traffic was handled jointly with connecting carriers. Only six percent of the carriers handled more carloads in local service than they handled jointly with other carriers.

More than 40 percent of the short lines and regionals (including 50 percent of the carriers spun-off between 1985-1994) are dependent on a connection with a single carrier. With the mergers of Class I carriers in 1995 and the likelihood of additional mergers of Class Is, the percentage of short lines and regionals connecting with a single carrier is likely to increase.

Although few definitive statistics are available, it is clear that short line and regional carriers operate at a much lower average density level than Class I carriers. Even the Wisconsin Central, the largest of the regionals in mileage, has an average density which is just more than half of the density of the lightest density Class I, the SOO Line. SOO in turn has only 40 percent of the density of the average U.S. Class I carrier.

The Class I carriers report their density by mile of track to the ICC[27]. By assuming that there is only one track mile per route mile for the lighter densities, and by estimating of the distribution of track miles to route miles for heavier densities, an estimate of densities by route mile can be developed. From carloadings and estimated hauls, crude estimates of the average densities of short lines and regionals can be made[28]. Table 3 summarizes these estimates. Despite

Estimated Density - Class I's and Short Line and Regionals

	Class I Carriers		Short Lines and Regionals		
	By Route Mileage	Percent	Number of Carriers	Route Miles	Percent by Route Mile
Density - GTM / Mile					
Over 20 Million	45,500	36.9%	10	2,300	4.9%
5 to 20 Million	29,100	23.6%	19	7,400	15.7%
1 to 5 Million	21,500	17.4%	95	14,100	29.9%
Under 1 Million	27,200	22.1%	406	22,800	48.4%
Unknown			19	500	1.1%
Total	123,300		549	47,100	

[27] A double track line would be reported as two track miles per route mile. If 40 million gross ton miles traversed a double track route, 20 million gross ton miles per track would be reported.

[28] Gross ton miles on a particular short line was estimated by multiplying carloads by the average haul (from Profiles of U.S. Railroads if available, otherwise estimated), and then multiplying by 160 gross tons per loaded car mile (accounting for empty return). This procedure clearly overestimates the gross ton miles of carriers such as Florida East Coast which has a significant amount of intermodal traffic.

the imperfections in the data, it is clear that the density of short lines and regionals is less than that of the Class I carriers and that the short lines and regionals operate a large percentage of the lighter density trackage in the U.S. Although the lines spun off in the 1985-1994 decade are generally longer than other short lines and regionals, they do not appear to have any significant density advantage over other short lines and regionals.

The lack of density is also found in revenue generated per mile. While Class I revenues per carload were relatively flat from 1989 through 1994, their revenues per mile of road operated have increased by 23 percent to $243,000 per mile. Regional railways, because of their lighter density, average only $95,000 per mile in revenue. Short lines average only $52,000 per mile.

Because short lines and regionals operate at lower densities than Class Is, key assets required to operate a short line (including employees, rights of way, locomotives, and rolling stock) must be managed differently. Short lines and regionals operate approximately one-third fewer employees per mile than the Class Is, reflecting the lower revenue densities. In general, the carriers which were spun-off in the 1985-1994 decade appear to be more productive than their regional and short line counterparts. The spin-offs appear to be able to operate with fewer employees per mile and to generate more production per employee. The spin-off have two luxuries that probably contribute to their efficiency - an ability to make a start with a clean slate, and connections that tried to simplify their relationships with new short line and regional partners.

Locomotives are essential to operating a short line, but it is not necessary to own one to operate a short line. Of the carriers that reported information to the AAR, more than 10 percent did not report locomotive ownership. In most cases, the locomotives apparently were provided by an affiliate of the railway for operational purposes.

Short lines and regionals rarely buy new locomotives. The twelve Class Is, with five times the number of locomotives than the hundreds of short lines and regionals have, are a source of used locomotives. The lower utilization of the short line and regional fleet places less wear on such units. The technological advances that have entered the diesel electric field have not provided the lower costs or increased productivity required to justify the investment in a new locomotive. The mainstays of the short line and regional fleet are 20- to 40-year-old units, many which have been rebuilt during their lifetime. Affiliates of short lines and regionals have entered the locomotive rebuilding business and the parts business for locomotives whose manufacturers have exited the locomotive business. Even the short line owned by and serving the principal U.S. locomotive manufacturing plant uses a pair of fifteen-year-old locomotives.

Sixty-five percent of short lines and regionals do not have any freight cars, but rely on their connections and on shipper supplied cars. However, short lines do supply about 20 percent of the approximately 650,000 freight cars supplied by railways out of a total freight car fleet of 1.2 million cars. Fewer of the carriers spun-off in the 1985-1994 decade (25 percent) supply freight cars than the other short lines and regionals (40 percent).

Short lines and regionals have played a key role in a significant shift in the supply of freight cars. In the late 1960s, railways supplied almost 1.5 million freight cars out of a total U.S. fleet of 1.8 million. Predecessors of the current Class I carriers provided 95 percent of the railway-provided cars, with short lines and small Class Is providing the remainder. Changes in regulated car hire rates, particularly incentive per diem, attracted leasing companies into the supply of general service cars, particularly box cars. The leasing companies entered into sharing agreements with short lines and regionals to use their marks, and provided them with thousands of freight cars. Quickly, an oversupply of general service cars resulted that the industry has only recently worked off.

SUCCESS FACTORS

The limited financial data available makes it difficult to make definitive statements concerning the success of short line and regionals. Not all short lines and regionals are successful, and a few exit the industry each year. Prior to deregulation, the range of financial results was public information, and showed great variance among carriers. An analysis of the 1968 financial results of 61 short lines in the northeastern U.S. which were not subsidiaries of other railways showed 22 of the 61 losing money, and 15 of the 61 earning more than 15 percent on sales after taxes[29]. From the limited 1994 data I have been able to gather from reports available from publicly traded companies and reports filed with states, it appears that the very mixed pattern of success continues, with a significant number losing money and some being very profitable.

It is possible to identify a number of factors that distinguish the successful short lines and regionals from the less successful. The first distinguishing factor is a significant traffic base, large enough to support the infrastructure of the line. It is important for the management of a short line to recognize that the traffic it handles is both competitive to other modes and subject to shifts for reasons unrelated to transportation. The major reason behind the abandonment of many short lines and regionals is the disappearance of a major portion of its traffic,

[29] Nine of the 15 with earnings of 15 percent or better on sales no longer operated in 1994. Three more of the 15 merged with other carriers, with the properties of one of those three abandoned. Twelve of the 22 losing money in 1968 are not operating today, and two have been merged with other carriers. Overall, 32 of the 61 are not operating today, and eight of the 61 have lost their identity through merger or purchase by a new company.

such as may occur with the closure of the principal mine or industry on its line. Rarely does a short line that was created to preserve service on a line threatened with abandonment become one of the more profitable short lines, even though many have achieved their goal of preserving service[30]. Effective managers will constantly be developing traffic (through conversion from competing modes or the development of new industry to serve) to replace business that is lost and to develop economies of scale and density. Lines with a good traffic base, but which are losing money, often get a second and third chance at survival as successor organizations attempt to avail themselves of the business opportunity available.

Many of the short lines and regionals spun off between 1985-1994 have had outstanding business development results. The largest regional, Wisconsin Central, doubled its carloadings from 1988 (its first full year of operation) to 1994, excluding the impact of acquisitions. Indiana Rail Road more than quadrupled its business from 1987 to 1994. Others that were formed in this time period have had the opposite results; for example, the Florida West Coast lost more than 80 percent of its business after the closure of the two largest plants on its line. Although it is extremely difficult to compare the short lines and regionals with similar lines operated by Class I carriers, I believe that overall, the short lines and regionals develop more business on lighter density lines because of their focus on the communities they serve.

The second key success factor is the management itself. In many respects, short lines are more difficult to manage because of their smaller scale. Because they cannot afford the depth of management, the general manager of a property must be expert an on a wide range of issues, including infrastructure, mechanics, operations, contracts, human relations, government relations, and business development. Short lines and regionals that have tried to mirror the organization of large railways have usually found those organizations to be too costly. Cost control is a requirement of short line management, given competitive factors and the lighter densities of line.

A third key factor in the success of a short line or regional is an effective working relationship with connecting rail lines. As mentioned above, most traffic handled by short lines and regionals is handled jointly with connecting rail lines. With rail customers increasingly insisting on seamless service, it becomes critical that all interfaces (operating, data exchange, traffic, and equipment) between short lines and regionals be as smooth and effective as

[30] A notable exception to this generalization is the Georgetown Railroad. It was formed in 1958 by local interests to operate an eight-mile line under threat of abandonment. An aggregate industry was developed on the line, and Georgetown now handles more than 40,000 carloads per year. The railroad and its affiliates have become leaders in the development of technology for rail haulage of aggregates.

possible. Because connecting rail lines can be competitors as well as partners, no short line or regional can survive a hostile relationship.

For new short lines and regionals, a key success factor is the relationship formed with their principal connection at the time of sale. A seller who seeks to maximize immediate returns from the sale will likely place a financial burden on the buyer, which may make it more likely that the buyer will fail. A buyer with a long term perspective is more likely to develop a partner that will grow and prosper. The seller should share in that prosperity through the interchange business that they both handle.

OUTLOOK AND CHALLENGES FOR SHORT LINES AND REGIONALS

Table 4 summarizes some of the major challenges that short lines and regionals face. In response to these challenges, I expect to see continuing evolution in the short line and regional industry. First, the industry in the U.S. should continue to grow. Class I railways will continue to spin-off lines beyond the 1,800 miles of lines that Conrail and the 4,000 miles that BNSF have announced as spin-off candidates. Second, some of the existing operations of short lines and regionals will be abandoned. Not enough is being reinvested in track to continue to operate all the lines currently in operation (not all of the lines are generating sufficient business to justify reinvestment). Third, common ownership of short lines and regionals will to increase as companies currently in business seek

Table 4
Challenges Facing Short Lines and Regionals

Challenge	Potential Impact
Growth of intermodal and rail-highway transfer traffic	Connecting carriers may shift investment away from conventional "loose car" railroading. Most short lines have been unable to develop intermodal traffic and therefore are not participating in growth. Intermodal traffic requires partners who may compete directly with short lines and regionals for traffic.
Continuing cost reduction effort by partners may reduce cost advantage of short lines	Class Is may spin off fewer lines, giving existing short lines and regionals fewer opportunities for economies of scale through expansion. Short lines and regionals will be under pressure to match partners' cost reduction.
Fewer surplus resources may be available from Class I carriers	Short lines and regionals may have to spend increased attention and effort on training. Fewer surplus locomotives and freight cars may be available, especially of the size needed for short line operations - driving up prices and required investment.
Increasing concentration of Class I carriers	Short lines and regionals will have to further develop cooperative partnerships with connections. Niche markets will become more important.

economies of scale. Finally, as the number of major carriers in the U.S. decreases, short lines and their partners will forge closer alliances as they attempt to make their joint operations appear seamless to their customers.

If the U.S. experience is applicable, the spin-off of lines in Canada to short line operators may be a method of allowing the CN and CP to become more competitive in the North American market while preserving service to communities on lighter density lines. For the spin-off to work effectively, the lines sold must have sufficient traffic density to be attractive, capable management must be in place to run the new operations, and effective partnerships must be developed between the new short lines, connecting carriers, and the communities served.

References:

American Shortline Railway Guide - Commercial Edition by Edward Lewis; PO Box 505 Aberdeen, NC 28315-0505

Annual Report of Transportation Statistics in the United States for the year ending December 31, 1968; Part 1 - Railroad; Interstate Commerce Commission, Washington, D.C.

Comprehensive Guide to Industrial Locomotives by Jay Reed; Published by Rio Hondo Publishing, 2595 Hardies Ln., Santa Rosa, CA 95403; ISBN 0-9647221-0-0

The Official Locomotive Rosters & News; 1994-1995 Edition; by James W. Kerr, Published by DPA-LTA Enterprises Inc., PO Box 3000, Champlain, NY 12919-3000; ISBN 0-919295-26-6

The Official Railway Equipment Register; published quarterly by by K-III Directory Corp., 10 Lake Drive, Hightstown, NJ 08520; ISSN 0030-0373

The Official Railway Guide; published bi-monthly by K-III Directory Corp., 10 Lake Drive, Hightstown, NJ 08520; ISSN 1069-1715

Profiles of U.S. Railroads; published annually by Association of American Railroads, 50 F Street, N.W., Washington, DC 20001 (Note: available in electronic form only)

Rail In Canada; published annually by Statistics Canada, Transportation Division, Surface and Marine Transport Section, Ottawa, K1A 0T6; ISSN 0843-4350

The Short Line; published bi-monthly; PO Box 607, Pleasant Garden, NC 27313; ISSN 0199-4050

Small Railroads by Harvey A. Levine, et al; published by Book Crafters, Inc., Chelsea, MI; 1982; Library of Congress Catalog Card Number: 81-70898

State Short Line Railroads and the Rural Economy; by Michael W. Babcock et al, published by Kansas State University, Mahattan, KS, 66506; Report No. K-TRAN:KSU-92-2R

Steam Passenger Service Directory; published annually by Empire State Railway Museum, Inc.; available from Great Eastern Publishing, Post Office Box 246, Richmond, VT 05477; ISSN 0081-542X

A Survey of Shipper Satisfaction with Service and Rates of Shortline and Regional Railroads; by Office of Transportation Analysis, Interstate Commerce Commission and Federal Railroad Administration, U.S. Department of Transportation, Washington , D.C., August 1989

CHARACTERISTICS OF U.S. GRAIN SHORT-LINE RAILROADS: IMPLICATIONS FOR THE PRAIRIE PROVINCES

by Denver Tolliver, John Bitzan, Trent Byberg

1 INTRODUCTION

The railroad network of the Prairie Provinces includes many light-density branch lines that are used for gathering and consolidating grain shipments. As the Canadian railway system changes, the future of many of these lines is uncertain. The CN privatization act and the movement away from the WGTA have created expedited procedures for abandonment of branch lines and potential line sales to short-line operators.

Many potential short lines in the Prairie Provinces are grain-dependent branch lines (on which most of the traffic and revenue is generated from grain traffic). Thus, the economics and organizational structure of potential grain short-line railroads are topics of special interest in Canada. Since the United States already has considerable experience with short-line railroads, it may be useful to analyze the U.S. experience, particularly with respect to grain carriers.

This paper presents an overview of grain short-line railroads in the U.S. and analyzes differences between grain and non-grain carriers. In some cases, comparisons are also made to Class I values. Section 2 of the paper presents an overview of rail grain transport in the U.S. and identifies some important movement characteristics. Section 3 establishes criteria for separating short-line railroads into grain and non-grain carriers. Section 4 describes and compares the network characteristics of grain and non-grain short lines. Section 5 focuses on equipment comparisons, while section 6 covers labor factors. In section 7, the cost and financial characteristics of grain and non-grain short lines are analyzed and contrasted.

2 OVERVIEW OF U.S. RAIL GRAIN TRAFFIC

Before focusing on short-line railroads, some overall characteristics of U.S. grain rail traffic are highlighted. The source is the public railroad waybill sample. This sample provides estimates of the quantities of grains terminated by large railroads

in the United States, and allows the calculation of several important movement characteristics[1].

According to the waybill sample, large railroads terminated over 10.9 billion metric tons of grains and oilseeds in the U.S. in 1994. Of this total, approximately 5.3 billion metric tons (or 49%) originated from the North-Central region, which includes the states of ND, SD, MN, KS, IA, NE, MO, and WI. Much of the U.S. wheat and barley crop production occurs in this region.

The average haul for wheat originated in the North-Central region in 1994 was 642 miles; while the average revenue was $4.59 per car-mile. These values are very close to the average haul for all wheat (638 miles) and the average revenue for all wheat shipments ($4.14 per car-mile). Later in the paper, these statistics will be compared to specific values for grain short-line railroads.

3 DEFINING GRAIN SHORT-LINE RAILROADS

The remainder of this paper focuses on differences between grain and non-grain short-line railroads. The information is derived from a survey conducted by the American Short-Line Railroad Association (ASLRA) and the Upper Great Plains Transportation Institute. The data reflect calendar year 1994 operations and conditions, and thus are consistent with the waybill and traffic information presented above. The survey includes network, traffic, revenue, commodity, equipment, financial, and operational data. Thus, it allows for a comprehensive comparison of grain and non-grain carriers across a range of variables.

The ASLRA survey collects information for both local (line-haul) and regional railroads. Regional railroads are defined as those that operate at least 350 miles of road and generate revenues of at least $40 million, but have revenues less than the Class I threshold of $256 million[2]. Local line-haul railroads are defined as those that fall below the regional railroad threshold, excluding switching and terminal railroads.

The local line-haul railroad seems to fit the classic definition of a short-line carrier—one that operates a limited number of route miles and interchanges most of its traffic with a Class I carrier. Clearly, potential exists for larger regional carriers in the Prairie Provinces. However, the local line-haul classification appears to be more descriptive of potential grain short-line gathering networks. For these reasons, the remainder of this analysis focuses on local line-haul operators.

Table 1 shows the relative importance of farm products in the traffic bases of local line-haul railroads. The indicator variable is: farm products carloads as a percentage of total carloads handled. Of the 88 respondents classified as local line-haul carriers, farm products comprised no more than 22% of the 1994 traffic base for 62 (or 71%) of these carriers. For this group, grain revenues, costs, and other transportation characteristics are unlikely to be significant factors in the carrier's overall operations. In comparison, farm products comprised more than 44% of the 1994 traffic base for 15 of the local line-haul carriers.

To more clearly delimit grain carriers, these 15 railroads were further analyzed in terms of location and volume of terminated fertilizer shipments. In some cases, county grain production data were examined to determine whether the farm products carloads consisted primarily of grain or other crops. All 15 of these railroads were found to operate in grain production areas, and were thus classified as grain local line-haul carriers. In the statistical analysis which follows, this group is compared to the remaining 73 local line-haul respondents, who are referred to as non-grain short lines.

4 NETWORK COMPARISON

In this section of the paper, four network characteristics are used to compare grain and non-grain short lines: (1) track condition or quality; (2) traffic density; (3) average length of haul; and (4) and percentage of carloads interchanged. Track quality is measured by the percentage of route miles equipped with rail weighing less than 90 pounds per yard. Much of this track apparently consists of older jointed-rail over which trains operate at slower speeds. Traffic density is defined as carloads handled per mile, and is an important factor in achieving economies of utilization. The average length of haul is defined as the average distance that a shipment travels on the short-line's network. The percentage of local traffic (one minus the percentage of carloads interchanged) describes how much operational control the short-line railroad has over its traffic, and is a proxy for yard switching costs.

Table 2 summarizes the results of the network comparison. In terms of track quality, grain short lines have a higher percentage of route miles with 90-pound or heavier rail (77% of route miles as compared to 63% for non-grain carriers). The mean length of haul is also slightly higher for grain short lines. Moreover, there are slight differences in density and the percentage of carloads interchanged. However, none of the differences among means shown in Table 2 are statistically significant.

The testing function used (the t-statistic) helps explain whether the differences between the mean values for grain and non-grain short lines are meaningful, given the variances that exist within each group. For example, a seemingly large difference between grain and non-grain short lines in terms of average haul (50.1 miles versus 32.92 miles) is not statistically significant because the standard deviation for the grain group is 48.6 miles. The statistical significance of a t-test is actually judged by the corresponding P-value. Traditionally, three levels of P-values are evaluated: 0.10, 0.05, and 0.01. These values are referred to as "moderately significant", "significant", and "highly significant", respectively.

In Table 3, the mean values for short-line railroads are compared to the values of two U.S. Class I grain carriers: the BN and the Soo Line Railroad. The Soo Line reflects CP Rail's U.S. operations in the northern plains. As the data show, the BN has the highest quality track, with 91% of its route miles constructed with 90-pound rail or heavier. However, the mean values for the Soo Line and grain short lines are exactly the same: 77% of route miles. In terms of traffic density, the BN system has 11.76 million revenue ton-miles per mile, as compared to four million for the Soo Line. In contrast, the mean value for grain short lines is approximately 456,000, as compared to 413,000 for non-grain short lines. The average haul for the BN is 793 miles, as compared to 393 miles for the Soo Line and 50 miles for grain short lines. As mentioned previously, the U.S. average haul for wheat is 638 miles. Finally, the BN interchanges only 45% of its traffic, as compared to 64% for the Soo Line and 71% for grain short lines.

5 EQUIPMENT COMPARISON

Many U.S. short lines depend on Class I carriers for car supply. Furthermore, the power characteristics and needs of short lines are much different from those of Class I carriers. As is detailed later, short lines operate at slower speeds than Class I carriers. Moreover, many short lines don't continuously utilize locomotive power, frequently shutting units off at night. Thus, short-line carriers are able to purchase and more effectively utilize older power. For example, short lines typically pay as little as $150,000 to $200,000 for used low-horsepower units which they operate for some time before rebuilding[3]. In contrast, Class I carriers may pay between $500,00 and $800,000 for a remanufactured unit[4].

The purpose of this section of the paper is to compare the equipment characteristics of grain and non-grain short lines, and to compare them to Class I railroads. The comparison includes: cars owned per 1,000 carloads of traffic, the age of carrier car fleets, locomotives owned per 1,000 carloads, and the age of the locomotive fleet; all of which are summarized in Table 4.

As Table 4 shows, at least 45% of the freight cars owned by local line-haul carriers are more than 20 years old, and more than 90% of their cars are at least 10 years old. Similarly, over 80% of the locomotive units owned by local line-haul carriers are more than 20 years old.

Although, there are no significant differences in equipment characteristics between grain and non-grain carriers, there are differences in equipment characteristics between the BN, the Soo Line and local line-haul carriers as shown in Table 5. Whereas the BN owns 14.8 freight cars per each 1,000 carloads handled, grain short lines own less than five cars per 1,000 carloads of traffic. Similarly, whereas, the BN owns 2.45 locomotives per each 1,000 carloads handled, grain short lines own only 1.11 locomotives per each 1,000 cars of traffic. Furthermore, there major differences in horsepower rating. The average horsepower of BN and Soo Line locomotives are 2,678 and 2,450, respectively[5]. In comparison, 87% of short-line locomotives are rated at 2,500 horsepower or less.

6 LABOR COMPARISON

Three measures of labor efficiency are compared in this section of the paper: number of employees per mile, number of carloads per employee, and revenue ton-miles per employee. Employees per mile provides a broad labor comparison based on system size, while carloads per employee provides a measure of the output produced per employee, and may be considered a measure of employee productivity.

Table 6 summarizes the comparison between grain and non-grain short lines. As the table shows, non-grain railroads have slightly fewer employees per mile and slightly more carloads per employee than grain short lines. However, neither of these differences is significant at conventional levels.

Table 7 provides a comparison between the two grain Class I carriers (BN and Soo Line) and grain and non-grain short lines. As the table shows, there are large differences in both employees per mile and carloads per employee between the Class I carriers and the short lines. In both cases, the short-line railroads exhibit much lower employment per mile and much higher carloads per employee. This difference is not unexpected, as many of the savings that short lines are able to realize are due to labor savings. Short lines are able to employ smaller crew sizes, are not bound by as many work rules, and require employees to perform a variety of tasks that would normally be restricted to separate job classifications on a Class I system. However, while these differences reflect positively on short-line railroads, some of the difference in carloads per employee is also due to differences in length of haul. This

is born-out by the comparison between BN, Soo Line, and short line revenue ton-miles per employee, as shown in Table 7.

7 COST AND FINANCIAL COMPARISON

This section provides a comparison of seven cost and financial characteristics of grain and non-grain short lines: (1) revenue per car-mile; (2) operating ratio; (3) operating expense per car-mile; (4) operating and capital expense per car-mile; (5) maintenance of way cost per mile; (6) equipment cost per mile; and (7) projected capital expenditures per mile. Revenue per car-mile is a standardized measure of rates charged. In general, revenue per car-mile will decrease with distance and average cars per shipment. Operating ratio is defined as total operating expenses divided by operating revenues, and is an important financial indicator for small railroads. Operating expense per car-mile is a standardized measure of train-operating, car ownership, and track maintenance costs (excluding capital expenditures). A related measure (operating and capital expenses per car-mile) standardizes both operational and capital expenses and reflects the relative economies of traffic density of each group. Two sub-categories of operating expenses— maintenance of way (MOW) and maintenance of equipment costs— are examined in order to isolate their effects on short-line costs. However, these values are computed on a route-mile basis. Finally, a comparison of projected capital costs per mile for grain and non-grain short-lines is provided.

Table 8 summarizes the comparison of costs and financial characteristics for grain and non-grain short-lines. Table 9 extends the analysis by comparing BN and Soo Line statistics to the grain and non-grain short-line values. As Table 8 shows, all of the revenue and expense figures are lower for grain short lines. Moreover, with the exception of operating ratio and MOW cost per mile, all of the differences in cost and financial characteristics between grain and non-grain short lines are significant at conventional levels: Measures that are significantly lower for grain dependent short-lines include revenue per car-mile, operating expense per car-mile, operating and capital expense per car-mile, equipment cost per mile, and projected capital expenditures per mile.

The difference in revenue per car-mile between grain and non-grain short lines (as shown in Table 8) is partially explained by length of haul. However, much of the difference may also be due to differences in commodity demand and related pricing factors[6]. Overall, the mean revenue per car-mile for local line-haul carriers is considerably greater than the U.S. average for wheat (of $4.14 per car-mile). Moreover, the average for all U.S. wheat is considerably higher than the mean for

all commodities (which is $2.81 and $2.55 per car-mile for the BN and the Soo Line, respectively).

Differences in operating and capital expenses between grain and non-grain short lines may also be reflective of differences in commodity characteristics. However, these cost differences may also be related to operational and maintenance factors. As Table 8 shows, grain short lines experience about 1/3 of the equipment cost per mile of non-grain carriers. This may be partially due to lower car-hire costs (as a result of Class I carrier management practices), better equipment utilization, and lower maintenance costs. Moreover, the projected capital expenditures for the next five years apparently reflect a lower expected level of road and equipment expenditures by grain short lines per route mile. In essence, grain short lines may be less capital-intensive than non-grain carriers.

Although train speed was not part of the ASLRA survey, the proportion of track miles in various Federal Railroad Administration (FRA) safety classes was collected. General inferences regarding train speeds can be drawn from these values. FRA track class 1 allows a maximum operating speed of 10 mph, while class 2 allows train speeds up to 25 mph. Excepted track is also restricted to 10 mph operations. According to the survey, approximately 24% of grain short-line route miles are classified as FRA class 1, while another 23% are classified as excepted track; meaning that the average running speed in both cases is less than 10 mph. Another 42% of grain short-line route miles fall into track class 2, meaning that the average train speed is probably somewhere between 20 and 25 mph. A similar distribution exists for non-grain short lines, with 18%, 46%, and 19% of their route miles falling into class 1, class 2, and excepted track categories, respectively. Given these distributions, the system average train speed for grain short lines is apparently less than 17 mph, as compared to 25 mph for the BN system[7].

As Table 9 shows, operating expense per car-mile, and operating and capital expenses per car-mile are much lower for Class I railroads than for short lines. Overall, the higher short-line operating expenses per car-mile may be due to slower operating speeds, a greater frequency of interchange, and lower traffic densities. However, because of slower operating speeds, short-line railroads may be able to spend less in track maintenance per mile than Class I carriers, as is born-out by the comparison in Table 9. Slower speeds may reduce the probability of derailments and thus allow short lines to save on maintenance of way costs. In general, the lower Class I carrier costs shown in Table 9 are primarily reflective of economies attributable to longer hauls, higher traffic densities, larger system size, and broader commodity mix.

8 CONCLUSIONS

This paper has analyzed differences between grain and non-grain local line-haul railroads, and compared select short-line performance measures with values for two northern grain railroads; the BN and the Soo Line. In summary, the average grain short line in the U.S. operates 94 route miles, handled 9,454 carloads of traffic, generates 395 carloads of traffic per employee, has a traffic density of 95 cars per mile, has an average haul of 50 miles, earns revenues of $12.60 per loaded car-mile, and has an operating ratio of .79. Although the average grain railroad operating expense (of $8.92 per car-mile) is considerably less than the average revenue, the combined operating and capital unit cost (of $14.35 per car-mile) actually exceeds the average car-mile revenue for these carriers. Moreover, projected capital expenditures are much lower for grain than for non-grain short lines. Apparently, grain short lines are able to spend less on track and locomotive maintenance than either non-grain short lines or Class I carriers. Moreover, in 1994, were able to parlay these savings into a lower operating ratio than either the BN or Soo Line could muster. However, in the long run, differences in track and locomotive performance could obliterate some of these operational efficiencies. Also, as Table 10 shows, U.S. grain short lines have more specialized traffic bases, which tend to cause greater downside risks.

Table 10 compares Herfindahl commodity concentration indexes for the BN, the Soo Line, and grain and non-grain short lines. The concentration index ranges from 0 to 1, based on the number of commodities in the carrier's traffic base and the relative share of car loadings of each product. An index of zero represents a widely diversified traffic base, while an index of 1.0 represents a single commodity carrier. As expected, grain short lines have the highest index of .488 (representing the least commodity diversity). However, non-grain short lines also have a high concentration index of .435. In comparison, the BN and CP have diversification indexes of .177 and .119, respectively[8].

In conclusion, short-line railroads are one possible option for operation of prairie grain branch lines in the future. However, many of the economies of U.S. carriers result from labor cost savings, lower track maintenance expenditures per mile, slower train speeds, and the use of older less-expensive locomotive technology. In order for short lines to succeed in Canada, many of the same labor, equipment and operational conditions may be needed before potential operators can achieve similar efficiencies.

Table 1. Grain Traffic Base of Local Line-Haul Carriers

Grain Carloads as a Percent of Total Carloads	Class Frequency	Cumulative Frequency	Cumulative Percentage
0%	1	1	1.1%
11%	51	52	59.1%
22%	10	62	70.5%
33%	6	68	77.3%
44%	6	74	84.1%
56%	4	78	88.6%
67%	4	82	93.2%
78%	0	82	93.2%
89%	1	83	94.3%
> 89	5	88	100.0%

Table 2. Differences in System Characteristics of Grain and Non-Grain Local Line Haul Railroads

Variable	Mean Value: Grain Short Lines	Mean Value: Non-Grain Short Lines	T-Statistic	P-Value
Track Quality (Track > 90 lb.)	77%	63%	1.71	0.155
Density (Carloads per Mile)	95.4	151.8	-2.06	0.043
Average Length of Haul	50.1	32.92	1.25	0.230
Percentage of Carloads Interchanged	71%	75%	-6.32	0.750

Table 3. Comparison of Operating System Characteristics of Grain and Non-Grain Local Line Haul Railroads and Two Class I Grain Railroads

Variable	Mean Value for BN	Mean Value for CP Rail	Mean Value: Grain Short Lines	Mean Value: Non-Grain Short Lines
Track Structure (Track ≥ 90 lbs.)	91%	77%	77%	63%
Density - Carloads per Mile	197	185	95	152
Density (Revenue Ton-Miles per Mile)	11,763,532	3,999,992	455,547	413,140
Average Length of Haul	793	393	50	33
Percentage of Carloads Interchanged	45%	64%*	71%	75%

* 1993

Table 4. Comparison of Equipment Inventories for Grain and Non-Grain Short Lines

Variable	Mean Value: Grain Short Lines	Mean Value: Non-Grain Short Lines	T-Statistic	P-Value
Cars Owned per 1,000 Carloads of Traffic	4.48	5.50	-0.33	0.745
% of Cars Owned > 20 Years Old	48%	45%	0.17	0.867
% of Cars Owned 10 to 20 Years Old	44%	48%	-0.21	0.836
Locomotives Owned per 1,000 Carloads Moved	1.11	2.03	-1.29	0.198
% of Locomotives Owned > 20 Years Old	81%	82%	-0.05	0.963
% of Locomotives Owned 10 to 20 Years Old	16%	17%	-0.04	0.968

Table 5. Comparison of Equipment Characteristics of Grain and Non-Grain Local Line Haul Railroads and Two Class I Grain Railroads

Variable	BN	Soo Line	Avg. for Grain Short Lines	Avg. for Non-Grain Short Lines
Locomotives Owned per 1,000 Carloads Traffic	2.45	NA	1.11	2.03
Cars Owned per 1,000 Carloads Traffic	14.80	9.55*	4.48	5.50

*1993

Table 6. Differences in Labor Characteristics of Grain and Non-Grain Local Line Haul Railroads

Variable	Mean Value: Grain Short Lines	Mean Value: Non-Grain Short Lines	T-Statistic	P-Value
Employees per Mile	0.411	0.338	0.55	0.589
Carloads per Employee	395	472	-0.98	0.340
Thousands of Revenue Ton-Miles per Employee	1,700	4,670	-1.54	0.130

Table 7. Comparison of Labor Characteristics of Grain and Non-Grain Local Line Haul Railroads and Two Class I Grain Railroads

Variable	BN	Soo Line	Avg. for Grain Short Lines	Avg. for Non-Grain Short Lines
Employees per Mile	1.38	0.80	0.41	0.34
Carloads per Employee	143	206*	395	472
Thousands of Revenue Ton-Miles per Employee	8,485	4,992	1,700	4,670

*1993

Table 8. Differences in Cost and Financial Characteristics of Grain and Non-Grain Local Line Haul Railroads

Variable	Mean Value: Grain Short Lines	Mean Value: Non-Grain Short Lines	T-Statistic	P-Value
Revenue per Car-Mile	$12.60	$27.87	-2.86	0.006
Operating Ratio	0.79	0.86	-0.71	0.448
Operating Expenses/ Car-Mile	$8.92	$36.81	-3.51	0.008
Operating & Capital Expenses/ Car-Mile	$14.35	$39.22	-2.74	0.008
MOW Cost per Mile	$4,348	$5,570	2.03	0.253
Equipment Cost per Mile	$4,136	$12,551	-2.94	0.004
Projected Capital Expenditures per Mile	$22,920	$50,512	2.22	0.030

Table 9. Comparison of Cost and Financial Characteristics of Grain and Non-Grain Local Line Haul Railroads and Two Class I Grain Railroads

Variable	BN	Soo Line	Average for for Grain Short Lines	Average for on-Grain Short Lines
Revenue per Car-Mile	$2.81	$2.55	$12.60	$27.87
Operating Expense/ Car-Mile	$2.34	$2.84	$8.92	$36.81
Operating & Capital Expense/ Car-Mile	$2.74	$3.33	$14.35	$39.22
MOW Cost per Mile	$36,636	$23,256	$4,348	$5,570
Operating Ratio	83%	111%	79%	86%

Table 10. Comparison of Traffic Diversity of Grain and Non-Grain Local Line Haul Railroads and Two Class I Grain Railroads

Variable	BN	Soo Line	Average for Grain Short Lines	Average for Non-Grain Short Lines
Herfindahl Index (between 0 and 1; 1=ships 1 product, 0=wide diversity)	0.177	0.119*	0.488	0.435

*1993

NOTES

1. The waybill sample is a random stratified sample of terminating shipments. Railroads that terminate 4,500 carloads or at least three percent of the rail traffic within a given state are subject to the reporting requirements.

2. Association of American Railroads (AAR). **Railroad Facts, 1995 Edition**. Washington, D.C., 1995.

3. Source: personal interview of short-line railroad officers.

4. For example, the CP estimates that it costs approximately $540,000 to remanufacture a 1,500-2,000 horsepower locomotive.

5. Values for the BN and Soo Line are computed from Schedule 710 of their R-1 reports.

6. Generally, grain shipments move in large car size-blocks and experience low loss and damage costs relative to many other commodities. Moreover, a variety of grains are produced in a separate geographic regions of the U.S., suggesting potential geographic and product competition. All of these factors suggest a relatively low rail rate for grain.

7. The short-line value in this comparison is approximated by assuming the following average running speeds for the following track classes: 1 (8 mph); 2 (22 mph); 3 (36 mph); excepted (8 mph). The BN value was computed from Schedule 755 as: train miles divided by train hours, running.

8. The higher BN index probably denotes a relatively large share of coal traffic.

TRUCK SAFETY
PERCEPTION AND REALITY

Greg Little
Ministry of Transportation, Ontario*

"The Attack of the Killer Monster Trucks!" I'm surprised that this is not a classic "B" movie from the 1950's. However, today, I believe this title encapsulates one of the perspectives the motoring public has to the presence of large commercial vehicles on Ontario highways. In my career with the Ontario Ministry of Transportation, I have worked in a number of aspects of the trucking policy in Ontario. During this time I can safely say that the motoring public has become increasingly adverse to sharing the road with large trucks, i.e. large straight trucks and tractor-trailers. One letter to the minister of the day summed up the prevailing attitude as well as any I recall. The writer was corresponding to object to the decision by the courts to overturn the Federal Lord's Day Act, which had the effect of permitting unrestricted trucking on Sundays. The writer was pleading to have one day of the week in which he could have the pleasure of driving without having to worry about those large, nasty, dangerous trucks being on the road.

This paper will examine the public perception of the safety of large trucks, how this perception matches reality, and what steps might be taken to address those perceptions in the context of a truck safety program.

Public Perception

As noted above, members of the motoring public have a poor perception of the safety of large trucks. I have identified three major reasons for this unfavourable view.

First, there is the experience of the motoring public. Tractor-trailers and other large trucks can be intimidating to motorists by their behaviour while operating on-highway. Truck drivers will, on occasion, tailgate vehicles which are impeding their way on two lane highways or driving slowly in the passing lane of multi-lane highways. It can be intimidating when all you can see in your rear view mirror is the grill of a truck. Trucks may also drive two abreast on multi-lane highways and

---

* The views expressed are those of the author and not necessarily those of the ministry

block cars from passing. And then there is the problem of splash-and-spray from trucks in bad weather. This occurs when driving is already a problem for car drivers. It is then made worse by the interaction with trucks.

If that is the perception of motorist experience with trucks, what is the reality. There is no doubt that bad experiences can give the motoring public a bad perception of the trucking industry. Unfortunately, a person's perception is often formed by bad experiences rather than by neutral ones. A car driver can have months or years or driving with no bad experiences involving trucks, but one or two bad situations will colour the person's view of all trucks and truck drivers. This is also true of the experiences which are recounted to friends and acquaintances. Experiences where nothing out of the ordinary happened are not the fodder for anecdotes. In my experience, truck drivers generally operate their vehicles in a professional manner. Unfortunately, all it takes to undo good or neutral experience is one encounter with an inconsiderate or inexperienced truck driver. With tens of thousands of trucks and truck drivers there will always be a few of these. Once formed, these perceptions are not easily changed.

The second factor in public perception is the media. Accidents involving trucks tend to be more serious than those which involve only cars and other small vehicles. In 1992 (the most recent statistics available for Ontario), large trucks (tractor-trailer and truck-trailer combinations) were involved in 2.1% of collisions where only property damage occurred. However, they were involved in 5.7% of fatalities[1]. In addition, truck accidents will often have a significant impact on traffic. A single accident involving a truck in a metropolitan area or on a major inter-city highway such as the 401 in Southern Ontario, can cause a significant disruption to traffic for extended periods; the destruction of the James Snow Parkway overpass west of Toronto by an intoxicated truck driver comes to mind. These events are inevitably going to be reported by the media.

The flurry of wheels coming off trucks in Ontario at the beginning of 1995 is a case in point. A number of these events occurred in a relatively short period of time, two of which involved fatalities. The media, of course, reported these events and were on the lookout for similar occurrences - even ones of a minor nature. This was addressed in Motor Truck (June 1995), a leading trucking industry publication:

> Meanwhile, cases of loose truck tires continue to receive extensive media scrutiny in Ontario. On April 25, a tractor-trailer lost two of its truck tires 20 kilometres apart... Two other trucks were damaged when they drove over one of the runaway tires.
>
> A transport truck lost its rear-wheel assembly on May 1 when it made a turn shortly after passing through a visual inspection. That

> assembly, however, remained lodged under the trailer.
>
> Another driver was charged on May 4 when a double set of wheels fell off a tractor-trailer just after it left Hwy. 401 at Charington.

A review of headlines from earlier years further demonstrates the understandable interest displayed by the media:

> "Blast threat shuts (trans-Canada) highway: dead driver left in cab after truck hauling gas flips" - Toronto Star. July 13, 1994
>
> "12 killed in California truck crash" - Montreal Gazette, March 10, 1994
>
> "Tractor-trailer crash dumps chemicals in lake (Bay Lake in Ontario)" - Toronto Star, December 20, 1993
>
> Truck driver in fatal crash must stand trial: top court" - Montreal Gazette, August 13, 1993
>
> "Back-to-school tragedy: a tanker truck-school bus crash near Carseland kills three" - Western Report, September 16, 1991

In the media, we have a very powerful force for forming the public perception of truck safety. Is perception reality in this case? The reality is that bad news is reported by the media. Truck accidents tend to be serious, as noted above. Therefore we are going to see a disproportionate number of serious incidents involving trucks reported. These will tend to make an impression on the reader (or viewer), and lead the reader to conclude there is a serious truck safety problem. Whether or not the truck driver or truck was at fault is irrelevant to the public (even if it was reported). The involvement of the truck is enough to condemn the trucking industry. Indeed, it can be argued that trucks are as safe or safer than other vehicles. I 'll examine this in detail later.

That being said, large trucks, by their very nature, may create severe problems when involved in accidents - regardless of fault. This gives the trucking industry, and governments which regulate it, a special duty to make safety a priority.

The third factor in the public perception of trucking safety is the ongoing poor performance of the trucking industry in vehicle safety inspections. Since 1990, there has been an annual (Canada/U.S.) Roadcheck of truck mechanical fitness. These efforts have resulted in consistently high out-of-service rates of one-third and more in Ontario[2]. These results not only make dramatic reading, but they raise real

questions about the actual safety of trucks on the roads. I examine the data on vehicle condition in relation to accident data below.

Perception Versus Reality
We have been exploring the public's perception of truck safety and how this does not necessarily match reality. Let us now examine the reality of truck safety as revealed by the hard facts, i.e. the data on truck collisions (or accidents as they have traditionally been called).

1. ACCIDENT DATA

The data on vehicle accidents is the primary source of information of vehicle safety. Ultimately, vehicle accidents (or more correctly collisions) is the bottom line in vehicle safety programs. So, what do the accident records tell us?

1.1 Vehicles in Accidents

Let's look first at the absolute numbers . The Ontario Road Safety Annual Report for 1992 (most recent available) provides the numbers of accidents. Table 1 gives the number of accidents for types of vehicles and class of accident.

TABLE 1
VEHICLES IN ACCIDENTS - 1992

Type of Vehicle	Fatal	%	Personal Injury	%	Property Damage	%	Total	%
Car	977	60.4	80,935	73.6	223,629	73.1	305,561	73.2
Truck	372	23	18,852	17.2	60,221	19.7	79,445	19.00
Tractor + semi	80	5	1,254	1.1	4,426	1.5	5,760	1.40
Truck + trailer	13	0.8	535	0.5	1,908	0.6	2,456	0.60
Other	174	10.8	8,298	7.6	15,745	5.1	24,217	5.80
Total	1,616	100	109,874	100	305,929	100	417,419	100.00

From these data we can draw two conclusions. First, heavy trucks account for a small proportion of total accidents. Second, both categories of heavy trucks are involved in proportionally more fatal accidents than injury or property damage accidents. In effect (as noted above), the accidents in which they are involved are more likely to be of a serious nature.

The first point that requires clarification when looking at these data is "what trucks are we interested in?" Small trucks (up to 24000 lbs in weight) can be operated by a class G licence which is the basic licence required to operate a car. Large straight trucks require a class D licence. We are interested in heavy trucks which equate to tractor/semi-trailer, tractor and 2 semi-trailer(trains),trucks and trailers, and large straight trucks; in effect vehicles which require a class A and/or D drivers licence to operate. The data for these trucks are shown in Table 2.

TABLE 2
TRUCKS IN ACCIDENTS

Dr. Lic.	Fatal	%	P.I.[1]	%	P.D.[2]	%	Total	%
G	230	63.7	11,237	82.90	35,022	80.8	46,489	81.2
D	38	10.5	784	5.8	3,064	7.10	3,886	6.8
A	93	25.9	1,542	11.4	5,244	12.1	6,879	12.0
Total	361	100	13,563	100	43,330	100	57,254	100

Table 2 shows a similar pattern as Table 1. The class G trucks have a lower proportion of fatal accidents than property damage and personal injury accidents similar to cars. While the Class D and A trucks show the reverse, indicating accidents involving the larger vehicles have a more serious impact.

1.2 Accident Rates by Vehicle Registration

It is easy to see that there is a basic problem with using these data to analyze truck safety. The data does not account for the relative frequency (in terms of extent of use) of different types of vehicles. If there were similar numbers of vehicle used in similar amounts and in similar ways, these absolute numbers would be a good absolute measure of safety. However, this is not the case. To examine this, first let's look at the number of vehicles in Ontario. Table 3 shows the types of vehicles registered in Ontario.

---

[1] Personal Injury

[2] Property Damage

TABLE 3
VEHICLE REGISTRATION POPULATION - 1992

Vehicle Class	Pass	Comm	Farm Vehicle	Bus	School Bus	Other
Registered Vehicles	4925298	1046800	33798	19882	9448	581111
Percent	74.00	15.80	0.50	0.30	0.10	8.80

As indicated in Table 3, commercial vehicles are a smaller proportion of the total than passenger vehicles. However, they comprise a significant proportion at 15%. However, as noted above, the classification commercial includes more vehicles than we are interested in. If we again look at trucks by class of licence required, we get a different picture for both number of registered vehicles and accident rate as shown in Table 4.

TABLE 4
ACCIDENT RATES OF TRUCKS

Vehicle Class	Accident	Registered Veh.	AccidentRate (/100)
Truck Class G	46,489	9,533,453	4.9
Truck Class D	3,886	54,654	7.1
Truck Class A	6,879	98,066	7.0
Total Truck	57,254	1,106,173	5.2
Passenger	305,561	4,925,298	6.2

In Table 4 we see that the larger trucks, i.e. Class A and D have higher (poorer) accident rates than the smaller trucks. Trucks have, in total, a lower accident rate than cars. Let's now look at the same information for fatal accidents. This is provided in Table 5.

TABLE 5
FATAL ACCIDENT RATES

Vehicle Class	Fatal Accidents	Registered Vehicles	Accident Rate (per 10,000)
Truck Class G	230	953,453	2.41
Truck Class D	38	54,654	6.95
Truck Class A	93	98,066	9.43
Total Truck	361	1,106,182	3.26
Passenger	997	4,925,173	2.02

Table 5 gives quite a different picture for fatal accidents. Class D and A trucks have a significantly higher accident rate than class G trucks or passenger cars.

There are, however, problems with using the number of registered vehicles to determine accident rates. Registered vehicles do not account for vehicles from other jurisdictions operating in Ontario. In the case of tractor-trailers, this can account for a significant number. This can amount to 20% or more of tractor-trailers on Ontario highways[3].

1.3 Accident Exposure Rates

Another significant factor in determining the relative safety of vehicles is the number of accidents by the distance travelled i.e. accident exposure rate. The further a vehicle travels, the greater the exposure to accidents. Large commercial vehicles, especially tractor-trailers and trains, are generally driven much farther than passenger cars in a year. Tractor-trailers often average 100,000 kilometers or more per year to maximize their earning potential.

Attempts have been made to measure the exposure rates of trucks and other vehicles. The 1992 ORSAR reports accident exposure rates per million kilometers travelled. The report does not state if this relates to all vehicles or only passenger cars. However, given the predominance of passenger vehicles in the fleet make-up, the rate for cars would be close to this.

The accident exposure rate has been relatively consistent over the years as demonstrated in Table 6.

TABLE 6
ACCIDENT RATE PER
MILLION KILOMETERS TRAVELLED

YEAR	1987	1988	1989	1990	1991	1992	1993	1994
RATE	2.8	3.2	3.2	3.0	2.9	3.1	3.0	2.9

The accident exposure rate of heavy trucks has been calculated by the Ministry of Transportation. However, to obtain a meaningful measure for heavy trucks, it is necessary to calculate the rate for trucks of different characteristics to reflect the wide variety of truck configurations and types. In addition, the measurement is only possible at this time for the provincial highway network. Information is not currently available for truck travel on other parts of the road network in Ontario, e.g. urban roads. The information on truck accident exposure rates developed by the ministry is provided in Table 7.

TABLE 7
COLLISION RATES PER ESTIMATED KILOMETERS
FOR TRUCKS ON ONTARIO PROVINCIAL HIGHWAYS, 1989
(collisions per 1,000,000 km)

Truck Config.	Van	Flat-bed	Tank	Other	Mean
Straight	1.5	2.9	2.7	2.1	1.93
Straight & Trailer	4.0	3.3	14.8	3.6	4.10
Tractor only					12.31
Tractor & 1 trailer	0.8	1.2	0.4	0.6	0.79
A or C train	0.8	0.8	0.5	0.5	0.70
B train	1.5	1.0	0.9	1.1	1.07
Other	2.5	9.4	0.0	207.7	107.84
Total Veh.	0.9	1.5	0.8	1.7	1.17*

* Without 'other' overall mean = 1.09

As information shows, the accident exposure rate varies dramatically according to the configuration and body type. The category 'other' cannot be considered as it contains a mix of configuration and body types. This category aside, the 'tractor-only' and 'straight truck & trailer' categories have notably higher rates than the overall mean.

It is very tempting to compare this rate for heavy trucks to the accident exposure rate for all vehicles reported in ORSAR, particularly if you want to put truck safety in a favorable light. However, this must be done with extreme caution. This is a real case of 'apples and oranges'. The truck accident exposure rate is for travel only on the provincial highway system. Whereas, the overall exposure accident rate is for all types of road travel including urban. The type of highway may have an impact on both the extent and type of accidents. In addition, the exposure accident rates do not take into account the severity of accidents. As we have seen, the different types of accidents (fatal, injury, property damage) can have different rates for large trucks and other types of vehicles. The accident exposure rate deals only with total accidents. This being said, the difference in the 1989 exposure accident rate for large trucks (1.09) and all vehicles (3.2) is large enough to draw a tentative conclusion. It is not justifiable in my view to conclude, as was done in a recent report, that "All types of vehicles combined are three times more likely to be involved in a collision than heavy trucks and tractor-trailers."[4] It is probably reasonable to conclude that large trucks are relatively safe in terms of collision (accident) involvement. However, a direct comparison to other vehicles is on shaky ground at this point.

2. VEHICLE CONDITION

The factor of vehicle condition is also fertile ground for investigation of truck safety. Vehicle condition can relate to truck safety directly or indirectly. It relates directly when there are vehicle defects present in vehicles involved in collisions and the defect was a factor in the collision. There is an indirect relationship when vehicle inspections identified defects unrelated to a collision.

2.1 Vehicle Condition in Collisions

Information on the condition of vehicles involved in collisions is available from police reports. The information does not mean there is a direct cause and effect relationship between the vehicle condition and the collision. It is an indication only that the vehicle defect was present "which may have affected the operation of the vehicle and contributed to the cause of the accident."[5] In addition, the information does not reflect the true vehicle condition as it is recorded by police officers at the scene and not by a qualified mechanic. However, it does provide an indicator of the relative condition of different vehicle types.

The information on vehicle defects for fatal collisions in 1992 is provided in Table 8.

TABLE 8
1992 VEHICLE DEFECTS IN FATAL COLLISIONS

Vehicle Class	% Defects*
Class G Truck	3.9
Class D Truck	2.6
Class A Truck	0.0
All vehicles	1.2

* excluding unknown category

From this, there does not appear to be a problem with vehicle condition in the case of large trucks. In aggregate, for all fatal accidents in 1992, 92.8% of vehicles had no apparent defects, a statistic which indicates that vehicle defects are not a major problem in accidents. In addition, for heavy trucks the defect rate is less than other classes of vehicles. However, we again have to be cautious in drawing conclusions given the data limitation.

2.2 Vehicle Inspection

The Ministry of Transportation Vehicle Inspection Program involves on-road inspections of commercial vehicles. The program includes on-going inspections which are part of the day-to-day compliance activity, and the annual Roadcheck which is a North American wide safety blitz of three days duration. The ongoing inspection program is selective (targeting specific vehicles) while Roadcheck is random. The Roadcheck results are, therefore, the best indicator of overall vehicle condition of trucks on Ontario roads.

The Roadcheck results are provided in Table 9.

TABLE 9
ROADCHECK VEHICLE OUT-OF-SERVICE RATES

YEAR	INSPECTIONS	O-O-S RATE
1990	1,941	33%
1991	2,672	34%
1992	3,087	37%
1993	2,905	34%
1994	2,650	42%
1995	2,436	43%

The out-of-service rates would seem to indicate a serious on-going problem of unsafe trucks on Ontario highways. As noted in the introduction, this is a high profile issue with the Ontario public. The fact that a third or more of trucks which are inspected are unsafe enough to be taken off the road, leaves the distinct impression with people that trucks are dangerous.

Other than the annual Roadcheck blitz, the ministry vehicle inspection program is selective rather than random. In effect, the vehicles inspected are the ones which appear most likely to have problems. You would expect that these inspections would produce even higher out-of-service rates than Roadcheck. This is exactly the case. Recent truck safety blitzes have resulted in out-of-service rates of 50% and more.

These results for vehicle condition appear to be in conflict with the accident statistics. On the one hand, we have a consistently high occurrence of trucks in an unsafe condition. On the other hand, we have trucks which have a good record in terms of accident involvement. Even in the accident statistics which relate directly to vehicle condition, the defect rates (Table 8) show a low occurrence of defective vehicles involved in fatal accidents. Can we make any sense of this apparent paradox? Two possible explanations come to mind:

1. the out-of-service standards are not directly related to a condition which will contribute to an accident

2. the data on condition of vehicles in accidents is not accurate, i.e. significant defects are not being identified by the investigating police officer

Let's explore these two scenarios.

1. Out-of-Service Standards Not Related to Accidents

The out-of-service standards used in the vehicle inspection program are based on standards for the Commercial Vehicle Safety Alliance (CVSA). A level 1 CVSA inspection is undertaken in the Roadcheck Program to determine the vehicle condition. The inspections include all the major safety related components, e.g. lights, steering, suspension/frames, tires and brakes.

These are all components which directly relate to the safe operation of a vehicle. It is easy to see how a problem in any of the components could contribute to or cause an accident. The most frequent defect in the Roadcheck Program consistently is brakes out of adjustment. In 1995, problems with the air brake systems accounted for an out-of-service rate of 34.1%. It is certainly clear to anyone who has operated any kind of vehicle how improperly operating brakes could potentially contribute to an accident.

One of the major reasons why the rate for defects in vehicle inspections is significantly higher than those in accidents, is that the defective vehicle at the point of inspection only has the potential for involvement in an accident. The factors contributing to an accident are many and varied. A defective vehicle may not be involved in an accident for various reasons including:

- the correction of the defect
- the driver operating the vehicle to compensate for the defect, e.g. at a slower speed or following other vehicles at greater distances
- sheer luck

2. Data on Condition of Vehicles in Accidents is not Accurate

It is not possible to draw conclusions with any certainty from the available information regarding the accuracy of accident data respecting vehicle condition. However, without a doubt, the size of the difference between the inspection information and the accident information points strongly to the possibility that some - if not many - defects are not being accounted for in the case of truck accidents.

The accident report is completed by police officers who generally do not have specialized training required to identify vehicle defects on trucks. In addition, the police officer has other priorities in investigating an accident than doing an in-depth investigation on the mechanical condition of the vehicles involved. Finally, the conditions under which the police officer is investigating the accident are less than optimal, compared to the conditions by inspectors undertaking safety inspections. All this would strongly support the probability that vehicle defects are under reported on accidents.

established to ensure there was no bias in the types of accidents, to obtain a comprehensive picture of vehicle condition and the role it plays in accidents.

3. Improved Data on Trucking Activity

Using improved technology to track trucking activity should be given priority. The current data available on truck traffic is not adequate to develop good exposure information. Current technology should be employed, either on a permanent or periodic basis, to ensure this information is of good quality. Vehicle classification and weigh-in-motion equipment currently available would provide information on vehicle travel to enable comparative exposure data for trucks and other vehicle classes.

Programs such as the AVION/I-75 project are beginning to utilize new technologies, such as transponders and electronic licence plates. The potential for these sources to provide accurate, reliable, on-going data as a relatively inexpensive by-product, should be explored.

Industry Perception

Enhanced and improved truck safety data should be used to support the trucking industry to improve its record in the area of vehicle condition. Good data, properly analyzed, can greatly assist the government and industry to focus safety programs, e.g. safety blitzes in areas and segments of the industry that need the most attention. It can also be used to recognize the sectors that have good records or that show improvement. If the trucking industry has confidence in the data being used, it will give more credibility to any programs that are undertaken and improve the likelihood of success.

Public Perception

The enhanced and improved data on truck safety should be used to proactively educate the public on the safety record of trucks. This should emphasize the safety record relative to other types of vehicles. The credibility of the data used is important in getting the message across to the public that trucks are not necessarily the problem they are perceived to be. In addition, credible data will give credence to government programs, by being able to identify the problem and to demonstrate that the programs intend to be effective and are not just window dressing.

BIBLIOGRAPHY

1. Ministry of Transportation, Ontario. *1988 Commercial Vehicle Survey.* August 1991

2. Ministry of Transportation, Ontario. *Annual Roadcheck Reports.* Downsview, Ontario 1990 to 1995

3. Ministry of Transportation, Ontario. *ORSAR 1992 Ontario Road Safety Annual Report.* Downsview, Ontario 1993

4. Ministry of Transportation, Ontario. *Evaluation of Commercial Vehicle Safety: The Definition of Safety.* Downsview, Ontario 1995

5. Ministry of Transportation, Ontario. *Exposure and Collision Rates in Ontario.* Downsview, Ontario 1993

END NOTES

1. ORSAR 1992, page 42

2. out-of-service rates: 1990 - 33%, 1991 - 34%, 1992 - 37%, 1993 - 34%, 1994 - 42%, 1995 43%

3. 20 % of vehicles surveyed in the 1988 Commercial Vehicle Survey were registered in other provinces. More recent local surveys show a higher percentage of non-Ontario vehicles.

4. The Safety of Tractor-Trailers in Ontario, page 7

5. Ministry of Transportation, Ontario Motor Vehicle Accident Report Manual, Section 317, page 1

CANADA'S TRUCKING INDUSTRY MOVES TOWARD A NATIONALLY CERTIFIED PROGRAM OF DRIVER TRAINING

by Lloyd L. Ash, P.Eng., P.Log.*

This paper is intended to enhance the reader's understanding of current initiatives of the Canadian Trucking Human Resources Council (CTHRC) directed toward a nationally certified program of driver training for Canada's Motor Carrier Industry. The comments presented are based on two industry surveys undertaken by the author in 1994 and 1995.

WHAT IS THE CANADIAN TRUCKING HUMAN RESOURCES COUNCIL?

In 1990, the Steering Committee for the Human Resources Study of the Canadian Trucking Industry, with support from Employment and Immigration Canada, commissioned a study of perceived human resource problems within the trucking industry. As a result of consensus, in 1993, industry and labour groups joined forces to create the Canadian Trucking Human Resources Council as an independent, not-for-profit organization. Initial seed money was provided to assist with planning the organization by Industry Canada.

As illustrated, CTHRC was founded by the Canadian Trucking Association, the Private Motor Truck Council of Canada, the Canadian Auto Workers (formerly the Canadian Brotherhood of Railway and Transport Workers), Teamsters Canada, and COM-CAR Owner Operators' Association. CTHRC is tasked with the promotion and support of human resource development and training of drivers/instructors to meet the needs of the Canadian Trucking Industry.

Those interested that the Trucking Industry has enough skilled people to effectively meet operational needs include the following:

---

* General Manager, Trimac Consulting Services, Calgary, Alberta

FOUNDING ORGANIZATIONS OF THE CANADIAN TRUCKING HUMAN RESOURCES COUNCIL (CTHRC)



2

INTERESTED STAKEHOLDERS

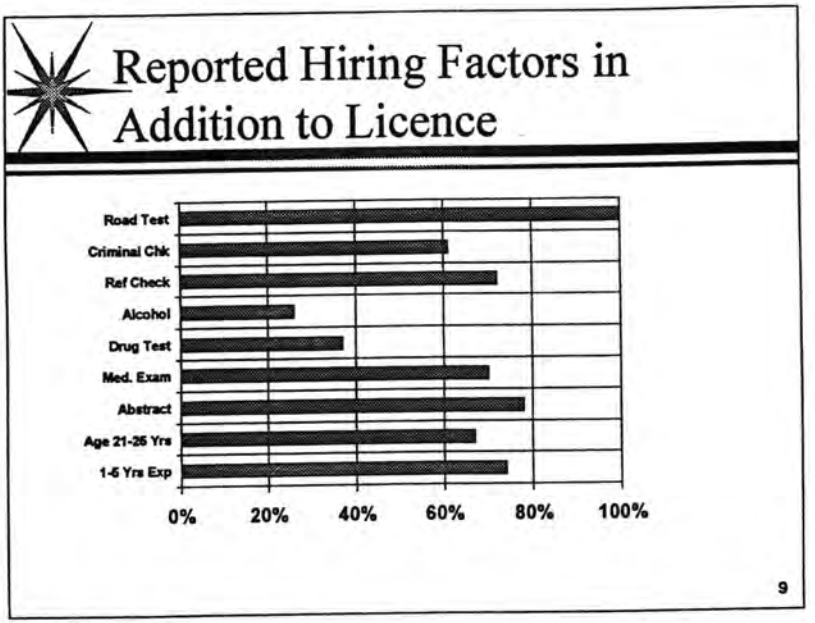
- CARRIERS
- SHIPPERS
- DRIVERS
- UNIONS
- OWNER OPERATORS
- LEASING COMPANIES
- PROVINCIAL & FEDERAL GOVERNMENTS
- TRAINING INSTITUTIONS
- INSURANCE COMPANIES
- EQUIPMENT MANUFACTURERS

3

Prior to the current harmonizing initiative by CTHRC, on behalf of the stakeholder groups identified, isolated activities were undertaken to set up and support excellence in training for drivers. These prior initiatives had been undertaken by individual trucking firms (some of whom started driving schools), selected driver training institutes, provincial governments (Quebec has a system of government operated technical driver training institutes, Manitoba has financially assisted with driver training costs, and Ontario has assisted in developing a program for business education needs of Owner-Operators) and the Government of Canada (Employment and Immigration Canada sponsored the first "Earning Your Wheels Curriculum of Driver Training").

RECENT "TYPICAL" PROFILE FOR THE POSITION OF DRIVER

When asked to describe their "hiring standards" for drivers, the criteria tabulated below are the most frequently discussed by fleet managers. In addition to the need for the required license to operate, most firms desired to hire drivers having between 1 and 5 yrs of directly relevant experience, an "acceptable" driving abstract, satisfactory references, and a medical examination. Many firms consider the types of violations noted on a candidate's abstract, not just the number of demerit points totalled by the particular jurisdiction. As compulsory drug testing and alcohol testing becomes mandatory in the U.S., firms were beginning to require this for their drivers as well as a straight medical exam. Most companies preferred candidates that were in the 21 to 25 years of age category and many cited the need for a minimum age of 21 for drivers operating to/from the U.S. An additional factor that was often important was the need to have no criminal background or to have an appropriate "pardon" that would enable a driver to haul into the U.S. Virtually all fleets required a prospective driver to undergo a pre-hiring road test with the company's driver instructor or safety supervisor. For some private trucking fleets, a contracted driver training school instructor was used to assess the candidate.



A very critical factor cited in most of the interviews with trucking firms was related to the candidate's attitude. The firm was entrusting significant capital resources, liability responsibilities, and customer assets to the prudent care and control of their drivers. They wished to ensure that wise judgement would govern the driver's actions.



As noted above, one means of assuring appropriate candidates for new drivers were found was to select drivers from a more restricted candidate "pool". Historically, many drivers were found from farm backgrounds, or the "promotion" of lower wage scale workers known to the organization (such as warehouse workers, dock labour, and driver helpers), or from a family where an established driver (older parent, spouse, or sibling) provided a role ethic / sponsorship for the candidate.

Because of several factors and changes, most of the trucking firms contacted noted that these traditional hiring sources were not yielding adequate candidates to meet the firm's ongoing needs (1994/1995). At the same time, many candidates from "driving schools" that applied were being found unacceptable (depending on the particular driving school) candidates for employment.

UNDERLYING TRENDS HAVE SUPPORTED THE NEED FOR CHANGE

The CTHRC has established Canada's first national trucking industry supported driver training curriculum and set of national accreditation standards for driver

training institutions in Canada. Several underlying factors have made this a timely initiative:

1) Changing social patterns have affected the numbers and sources of good candidates for entry (as drivers) into the trucking industry.
2) The responsibility profile for a driver has evolved compared to previous needs.
3) There is an enhanced awareness of learnings to be included as part of the training process.
4) Concerns have arisen about the significant historical variations in driver training standards among driving schools in Canada.
5) There is a desire to enhance the employment prospects for graduates of training programs, especially programs partially funded by government.
6) There is a need for a mechanism to equitably/broadly involve the industry in contributing to training and obtaining access to employable trainees.

The following discussion seeks to elaborate on the foregoing major issues. Comments provided are based on viewpoints and concerns expressed broadly by over 100 trucking firms in Canada -- in two separate surveys, conducted in late 1994 and early 1995. The first survey was undertaken by Trimac to over 50 for hire carriers and private fleet managers representing relatively larger trucking enterprises across Canada. The second survey was directed to understanding the opinions and concerns of 50 trucking enterprises operating in Alberta -- which consequently included more medium sized and small family enterprises.

CHANGING SOCIAL PATTERNS

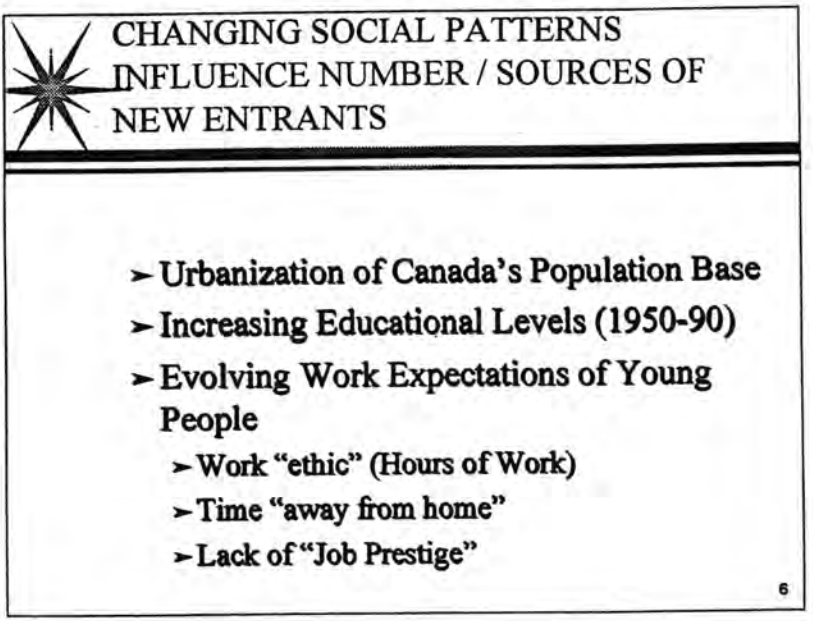
When considering the impacts of changes to Canadian society on truck driver availability (post World War II), several factors have worked to reduce the number of qualified/interested candidates for entry to the business.

Firstly, significant urbanization of the population base has taken place. In agricultural communities, young people often grow up gaining experience operating trucks and other large implements as part of the family farm business. Family farm life often instills a strong "work ethic" in family members and there are often times when there is a need to work alone and for long hours (eg. During harvest time). At the same time, agricultural earnings are not as "steady" as industrial earnings (due to seasonal and other factors), which gives

incentive for young people to seek out a career, such as driving a truck, to supplement and regularize their earning capability. During recent interviews with trucking firms across Canada, we found that those firms that were based in agricultural communities (a major national carrier headquartered in Manitoba, for example, as well as the smaller rural truckers interviewed in the Alberta study) indicated that they still experience minor difficulty locating suitable new driver applicants.

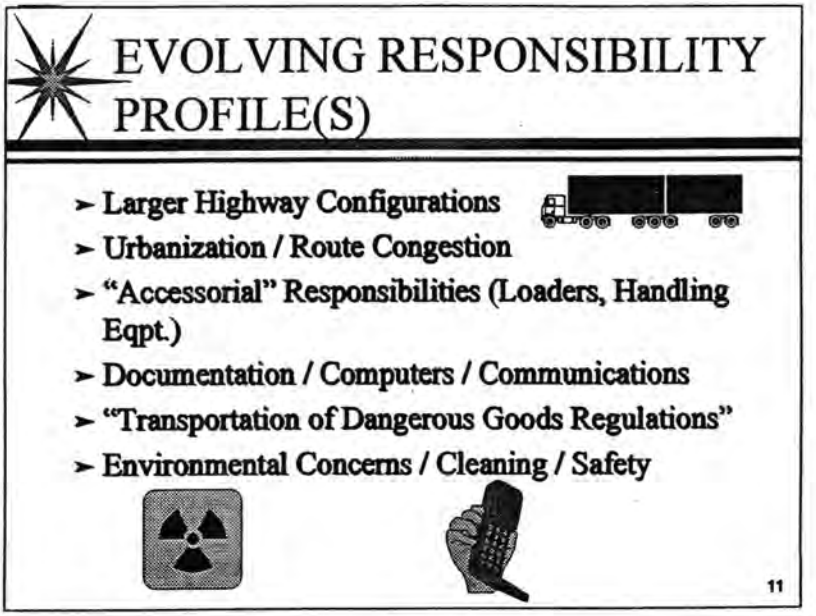
These same firms' counterparts in larger Canadian cities almost universally reported an increasing "driver shortage" and complained about an "aging workforce" of qualified drivers to send out on the road, with fewer young people interested or qualified for the task.

General educational expectations have increased, especially in terms of numbers of persons completing some form of post secondary education in Canada. Often, the "career expectations" of young people have been directed toward the "professions" or toward personal services industries -- as opposed to goods producing or transportation operations sectors. Driving a truck, especially if it involves a line haul operation, over longer distances often implies longer hours worked, significant time "away from home and family". Much of the urban "pick up and delivery" driving duty involves physical effort handling freight. Also, the career of "truck driver" has not been generally promoted to young people as one having significant "job prestige".



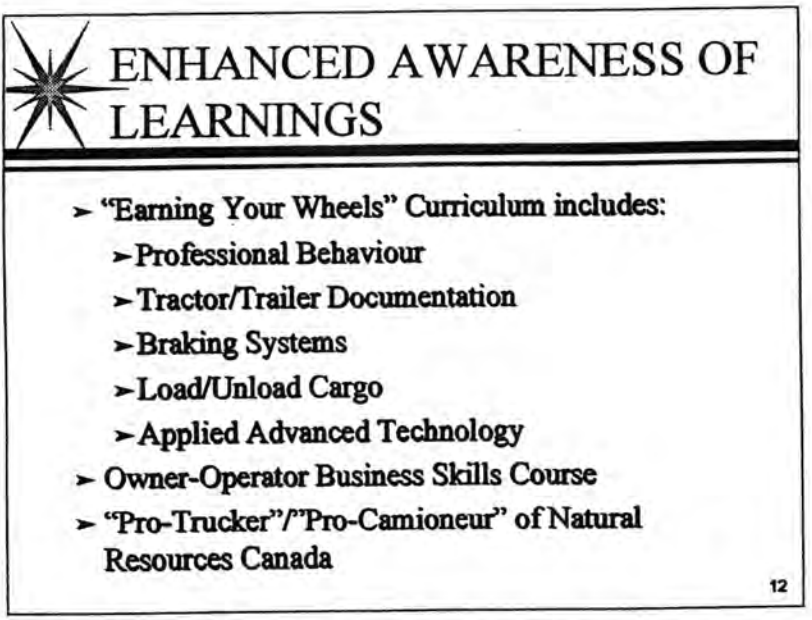
EVOLVING RESPONSIBILITIES FOR DRIVERS

At the same time as the foregoing social changes occur, industry requirements have evolved. Larger and more specialized vehicles are being operated. Traffic conditions reflect greater route congestion. There are frequently more specialized types of transport equipment and related goods handling functions involved in the job. Factors such as documentation for customs clearance, dangerous goods manifests/load acceptance factors, use of on board computers/satellite tracking/centralized dispatching/cellular communications have tended to change the work environment and the operator skills needed. Increased reliance of shippers on "just in time" deliveries and lower inventories impact driving schedules and critical "on-time"/"quality" service needs. The result of all these changes is the realization, by the industry, that higher standards of professional driver qualification are needed.



ENHANCED AWARENESS OF LEARNINGS

In this context, several initiatives directed to enhancing training materials that meet the required "learnings" for the industry have come to fruition in the past five years. In consultation with many of the agencies which sponsored the Canadian Trucking Human Resources Council, Employment and Immigration Canada supported the development of a very detailed "Earning Your Wheels" curriculum for driver instruction. Topics covered are very specific including all duties from pre-trip inspection, securing loads, brake inspections, tire inspections, driving skills, documentation, attitudes, loading, unloading, and through post trip responsibilities.



The Ontario Ministry of Transportation (Trucking Productivity Office) undertook to review and assist with the development of an Owner-Operator Business Skills Course which covers topics related to spec'ing a truck, life cycle analysis and management, financial management, assessment of operating costs/quoting rates, and other important issues for this sector of the industry.

Another very useful set of information, for operators and trainers of operators, is included within Natural Resources' Canada "Pro-Trucker" program. Although the prime focus of this program, which has existed since 1986, is on energy conservation driving techniques, benefits related to reduced maintenance costs and enhanced safety have been reported by fleets that were involved. This program is currently being re-designed by that agency for re-release in a new format that is directed to trucking fleets as well as non-trucking fleets such as municipal fleets, provincial fleets, service vehicle operators, etc.

Much of the foregoing material is incorporated in the CTHRC curriculum of driver studies. An additional emphasis, that is noted to be very important to the industry, is on the development of professional attitudes by truck operators. The aspects of customer service, public responsibility and safety, trusteeship for the assets of the trucking company and for the goods carried are to be instilled

in the new driver candidates. As a final component, there is a need for sufficient "behind the wheel" experience to ensure skilled operation of the vehicles.

CONCERNS ABOUT HISTORIC VARIATIONS IN DRIVER SCHOOL STANDARDS

Several of the industry managers contacted reported significant problems and variations in driving schools (non accredited). Some had interviewed candidates who did not have sufficient behind the wheel experience. Some of the poorer situations appeared to be geared toward enabling the candidate to "pass" the applicable provincial examination, only. At the same time, some of the schools had very good reputations and were noted to have excellent standards, but it was difficult for a specific trucking firm to know which type of driving school they were dealing with. Most of the trucking firms felt it was a good idea to integrate the learning experience at the school with some practical "behind the wheel" coaching in an actual trucking operation. Some of the driving schools owned by trucking companies were said to provide students with this sort of experience and the consultant was advised of a proposal for an "owner operator" type of driving school that would provide a tractor service, using qualified instructor/coaches, who could haul loads for trucking companies while undertaking instructional training.

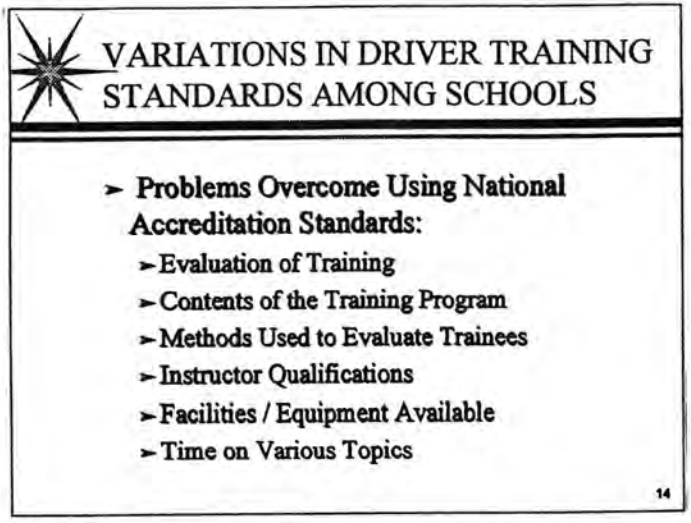
A key desire of trucking companies was to see "assessment" take place by the driving school throughout the candidate's training, and that assessment be directed to issues related to attitudes and skills, rather than simply a "road test" at the end of a driving course.

In this context, the CTHRC "driver curriculum" standard and "driver school accreditation" is aimed to ensure that uniform minimal standards are met by driving school programs in Canada.. By using this approach, differences in instructional techniques and existing programs are respected, yet industry standards become the "norm" for driver instruction in Canada.

ENHANCEMENT OF EMPLOYMENT PROSPECTS FOR GRADUATES

Many trucking industry officials were concerned, as are potential instruction candidates, that employment opportunities exist when the students graduate. Previously, there were instances of federal and provincial "re-training" efforts that have subsidized candidates to take training, which subsequently did not lead to employment in the industry.

Key components of effective training that would likely lead to employment were noted to include Classroom Training, Training in the Driver School Yard, Over the Road Training, and On the Job Training. The CTHRC curriculum prescribes specific topics to be trained as well as minimum standards for numbers of hours of instruction in each of the key training components. The industry broadly supported the establishment of these minimal training standards as a key to ensuring better acceptance of the graduates for entry level driver candidates.



In jurisdictions where pre-existing Provincial training standards exceeded the new CTHRC standards, it was decided that the "most stringent" standard will apply.



THE NEED FOR EQUITABLE/BROADER INVOLVEMENT OF INDUSTRY

Trucking firms and private fleet managers have a direct stake in assuring that their drivers are highly qualified and well trained. Most reported that they had (depending on need and availability of employment candidates) undertaken "driver instruction" activities either within the firm or, in some instances, by operating a driving school. Driving these efforts (even where a driving school situation existed) was a desire to assure qualified candidates for employment for the business.

In very specialized situations such as "hazardous goods", or special permit situations (eg. Overdimension programs or Heavy Hauling), only experienced drivers could directly apply for employment and even these drivers had a need for additional "re-training" for the specific application.

Problems were reported in relation to the knowledge that not all companies shared the same interest in training. This sometimes resulted in the company that made the initial investment in training incurring expenses to train drivers who ended up working for other firms -- sometimes a competitor who had avoided the expense of training.

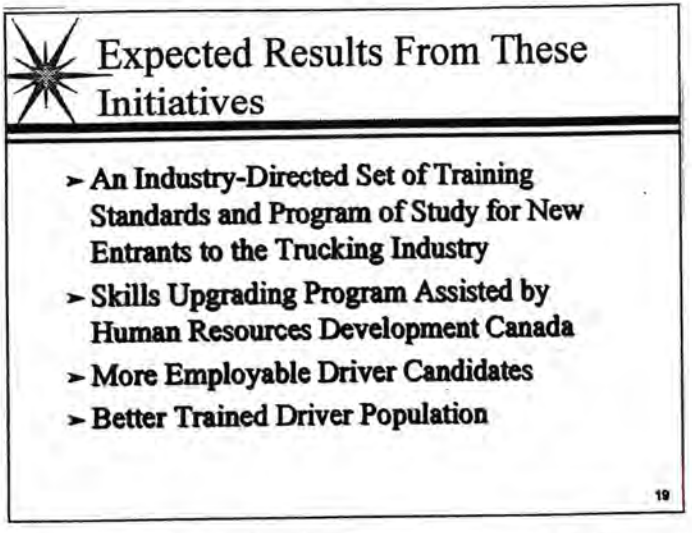
In the Alberta interviews, there was mixed support for a program that would see all members of the trucking industry (including private fleets) bear a proportionate share of costs for training -- that would then be undertaken on behalf of the entire trucking industry, according to industry standards.

EXPECTED RESULTS OF THE CTHRC PROGRAM

Current initiatives under active sponsorship of the Canadian Trucking Human Resources Council are threefold:

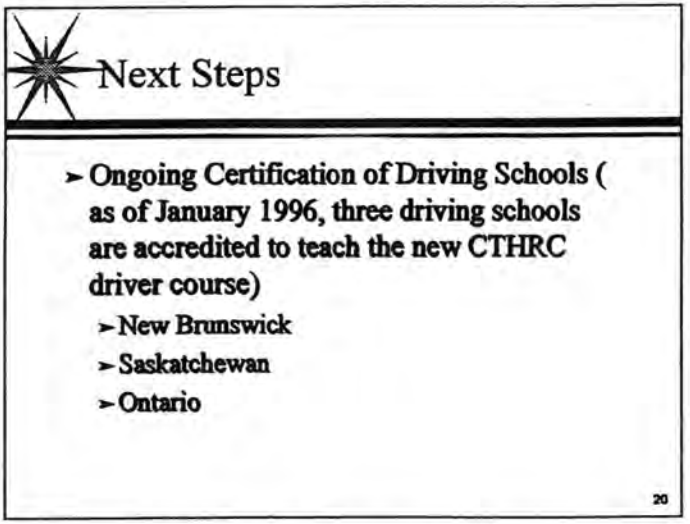
1) Driver School Accreditation
2) Driver Certification Process for New Drivers and Existing Drivers
3) Owner Operator Business Skills Training

The expected result from these initiatives is an industry directed set of training standards and a program of study for new entrants to the trucking industry, as well as attention to re-training / certification needs for existing drivers. This will result in more employable candidates for driver positions and a more professional, better trained, driver population in years to come.



Current efforts are focused primarily on promoting the accreditation process to the industry and to the driving schools across Canada. Beginning with the first

school in August of 1995, three schools have received accreditation as of January 1996.



The author hopes that this paper has enhanced the reader's awareness and understanding of these initiatives, the need for them, and the underlying issues of importance to truckers, shippers, regulators, and Canadian society as a whole.

RELATED REFERENCES

Canadian Trucking Human Resources Council, Driver Program: Earning Your Wheels, Instructor Guide, August 1994

Ontario Ministry of Transportation, Owner Operator Business Skills Survey: Summary of Survey Results, MTO Policy Division, October 1991.

Price Waterhouse, Canadian Trucking Industry Human Resource Challenges and Opportunities, Sponsored by Employment and Immigration Canada, 1990

Raney B, Rhone L, Siwak K, and Tharratt L, "Business Skills Requirements For Competitive Owner Operators", Proceedings of the 27th Annual Meeting of the Canadian Transportation Research Form, at Banff Alberta, 1992

Trimac Consulting Services, Partners for Productivity (Volume 1): Success Strategies for Owner Operators, prepared for the Ontario Ministry of Transportation, March 1992

Trimac Consulting Services, Trucking Industry Survey, prepared for the Canadian Trucking Human Resources Council, September 1994

Controlling Highway Safety – Commercial Driver Qualifications

Michèle Delaquis
17 Rosewarne Avenue
Winnipeg, Manitoba
Canada R2M 0V8

Fred P. Nix
307653 Hockley Valley Road
RR #1 Orangeville, Ontario
Canada L9W 2Y8

Introduction

Trucks – that is, trucks with a gross weight of 4,500 kilograms or more – accounted for 3.7% of vehicles involved in reportable traffic accidents in Canada in the period 1989 to 1993[1]. (A registered weight of 4,500 kilograms is the general dividing line between trucks falling under the National Safety Code and the smaller trucks, often pickups and vans that do not fall under the Code.) However, trucks accounted for 10.5% of the accidents where there was a fatality. This suggests that accidents involving trucks tend to be more serious than other accidents. But a more thorough exam of the statistics is needed to understand what is really going on. The purpose of this paper is to examine the relationship between driver qualifications and highway safety.

These accident statistics only show what vehicles were involved in accidents; they do not reveal anything about fault. Further, because of the absence of exposure measures, it is difficult to translate these into anything meaningful in the way of trends: is highway safety increasing or decreasing? how does the safety of large trucks compare with that of other vehicles? From the same source above, the truck accident rate declined from 3.9% of all reportable accidents in 1989 to 3.6% in 1993 – but this does not reveal any trends unless it is associated with changes in the relative proportion of registered vehicles. A few years ago, one federal Commission did suggest that "... according to tentative estimates made by Transport Canada for 1988, the fatal accident rate per 10,000 registered vehicles was 4.0 for passenger automobiles compared with 1.5 for commercial vehicles."[2] This, perhaps, provides a relative sense of truck safety, although the number of registered vehicles is a poor proxy for exposure. Distance driven would be preferable. Another source, using numbers from 1989-1991, suggests that, while trucks account for 3.7% of accidents and 3.9% of all registered vehicles, they account for 9% of total kilometres of travel in Canada.[3] These admittedly rough estimates seem to suggest that trucks have a better safety record than other vehicles. Whatever the relative safety of trucks, or

trends in this level of safety, the point remains that truck safety is a major policy issue. The public is demanding that something be done.

Available data suggest that drivers play a significant role in truck accidents. First, in the period 1989 - 1993, 18% of truck accidents were single-vehicle accidents. While this in itself is not proof that driver fault is the cause, it may be an indication that the driver's role is important. Single-vehicle accidents are often an indication of such things as a driver falling asleep at the wheel. Second, in nearly 29% of the truck accidents in this period "driver action" was a contributing factor. A further eight percent have "driver condition" as a factor in the accident. While these percentages are not additive, they are an indication that drivers play a key role in, perhaps, a third or more of all truck accidents.

What is not shown in these statistics is that, even where the driver is not directly responsible, driver action often plays a critical role in the severity of an accident. In fact, in trucking companies with better safety management practices, it is the concept of 'preventability' not 'fault' that is used in determining possible driver sanctions.

Numerous studies have suggested that the age, experience, or qualification of drivers plays a key role in determining which drivers are involved in accidents. At the Federal Highway Administration (FHWA) Truck and Bus Safety Summit[4], the topic "Education and Training of Drivers" was listed at the end of the conference as the third most important issue in improving commercial vehicle safety. At the Truck Safety conference held in Toronto in September 1995, the breakout working groups listed "driver qualification" as one of the top issues in improving highway safety. To quote from the summary of this conference:

> There is a concern that drivers are poorly trained to handle the requirements of long distance trucking, that licensing does not reflect the skills and experience that are necessary to drive large rigs. There is a parallel concern that drivers are not adequately compensated financially for their work, encouraging actions that could lead to unsafe situations.[5]

This paper looks at driver qualifications and at the attempts to manage these attributes.

Background

There are developments that help explain why the issue of driver qualification is both important and complex. For one thing, the trucking industry claims there is a

driver shortage. While this is difficult to document, if pay rates or working conditions in the industry are unattractive, then it may be more likely that drivers who are less qualified than desirable will be hired. Note the finding from the Toronto Safety Conference above linking the issue of driver pay and highway safety. Indeed, there is partial evidence that truck driver salaries have not kept up to the rate of inflation. Between 1987 and 1993 the average salary of a highway driver has increased from $28,348 to $34,701, an increase of 22.4%. During the same period, the consumer price index has increased by 24.9%. This is 'partial' evidence in the sense that it is simply the average annual salary of highway drivers taken from Statistics Canada's survey of for-hire motor carriers; it does not, therefore, account for all truck drivers; nor does it account for such things as the number of hours worked, the distance driven, overtime payments, etc. Nevertheless, and even if it is partial, it at least points to the possibility that driver wages are not keeping up to the levels necessary to attract highly qualified drivers to the industry.

Another issue is that the long delayed testing of Canadian drivers for drug and alcohol use has to be in place by July 1, 1996 if these drivers are going to operate in the United States. Most would agree that drug and alcohol use relate to driver performance in a rather direct manner. While some Canadian companies now do have programs, (for example, random testing of drivers on a volunteer basis), all companies with 50 or more drivers operating into the United States will be required to have a formal program in place by July. Since the announcement in January 1995 that Transport Canada would not be proceeding with plans to introduce substance use legislation, this has left it up to the industry to develop its own program. This will lead to inconsistency within the industry as only companies going into the US and having more than 50 drivers will be required to have programs.

The contentious issue of hours of work continues to frustrate both regulators and the industry. Much research has been carried out on both hours of work and fatigue. The general understanding is that safety decreases with fatigue. What is not so clear is exactly how to regulate hours of work in a manner that actually controls fatigue. The regulations currently in place are complex and difficult to enforce. They must often be applied between jurisdictions and they depend largely on the driver's log book. Log book cheating is not an uncommon problem: trade magazines quote figures such as "industry experts estimate that up to 80% of long distance truck drivers frequently make false entries to their log books".[6] Regulators, however, are reluctant to change current hours-of-work rules until the long-awaited joint Canadian-United States fatigue study is available.

There are a series of other issues touching on driver qualifications and/or the manner in which drivers are regulated:

- the continuing effort of Canadian jurisdictions to harmonize aspects of the National Safety Code that deal with drivers;
- the pending development of a "North American Driver Licence Agreement";
- the possible review of graduated licence systems and licence categories by the provinces[7];
- proposals in the United States to amend truck weight and dimension regulations (rescind the 80,000 lb cap) conditional upon new driver qualification categories[8];
- and (undoubtedly) other suggestions and/or possibilities.

The manner in which the state regulates commercial drivers is a much debated topic these days.

General Driver Qualifications

When trucking was deregulated in Canada, governments brought in "The National Safety Code" (January 1, 1988) in an attempt to ensure that a more competitive industry remained a safe one. The fear was that increased competition might result in companies taking short cuts on things such as safety practices. Several aspects of the National Safety Code are directed at driver training and/or qualification.

- Classified licence: Drivers must be licensed for the equipment they operate. There are 7 classes in total, classes 1 to 5 may be endorsed for the operation of equipment with air brakes. While exact configuration definition (by number of axles, GVW, length...) for each of the vehicle classes is left up to the provinces, trucks fall into either Class 1 (articulated vehicles) or Class 3 (heavy duty trucks with a RGVW greater than 4,500 kilograms)

- Knowledge test: There are standards for the tests for each class of drivers. Driver testing falls under provincial jurisdiction, but, by agreement, testing for each license class remains very similar between jurisdictions. Further, these knowledge tests are understood to be quite similar to those used in the United States.

- Examiner training: There are standards for driver examiners. Again, this falls under provincial jurisdictions so there are small differences in training requirements in each of the provinces.

- Hours of service: There is a limitation on the number of hours a driver may drive. The rules in Canada are complex (there are various cycles that may be used) but, to oversimplify, a driver can drive for 13 hours between two 8 hour rest periods; one of these rests may be taken in 2 periods, each of which is not less than 2 hours. There are further restrictions on the number of hours of

driving time per week. By comparison, in the United States, a driver can only drive for 10 hours between 8 hour rest periods.

- Self-certification: Carriers may train their own drivers. Larger governmental or municipal agencies are more likely to train and test their own drivers. As far as known, few (if any) motor carriers in Canada certify their own drivers.

- Single licence: A driver may not hold more than one licence. This prevents a driver with a bad record or a suspended licence from obtaining a licence in another jurisdiction.

- Interprovincial record exchange: Provinces can share data on drivers, their licences and their records.

- Medical standards for drivers: Drivers have to have a medical test every few years.

Without describing any more details on all of the NSC standards, the main observation here is that there is a great deal of regulatory effort aimed at drivers and their qualifications.

Licensing is another area of concern in driver qualification. While Canadian and US licenses are accepted reciprocally in both countries, there are some differences in licensing protocol. In order to legally drive an articulated vehicle in either Canada (Class 1 licence) or the United States (CDL) operators must pass a written knowledge test and a skills test. As most articulated vehicles have air brakes, an air brake endorsement is also required. However, drivers need only be 18 years of age in Canada whereas they must be 21 in the US in order to operate a vehicle interstate (in some states drivers may be 18 and operate intrastate only). A CDL in the US has further endorsements for carrying hazardous materials (H), hauling double or triple trailers (T), operating tank vehicles (N), and hazardous materials in tank vehicles (X). These endorsements require only a written knowledge test and have no skills testing requirements.

While evidence on how well these testing/classification procedures are working in Canada has not been found, one researcher from Michigan suggests that "persons who did better on various CDL tests were, in fact, more likely to be safer drivers..."[9] This is important as the purpose of the CDL test procedure is to improve highway safety. Apparently, it is working. This Michigan study related CDL test scores to driver records (driving 'points' and number of accidents) and found that "persons who pass CDL tests are significantly 'safer' than those who do not pass."

Specialized Driver Qualifications

As trucks have increased in size and as regulations have developed in other areas (eg, dangerous goods), there has been a tendency to require these drivers to be more highly qualified. This demand for greater driver qualification is evident in many of the 26 Canadian and American jurisdictions that allow long trucks under permit (longer combination vehicles or "LCVs" as they are known in the United States). Table 1 shows the additional requirements these jurisdictions place on drivers.

Some jurisdictions are developing agreements whereby LCV drivers will be recognized by more than one jurisdiction, eliminating the need for multiple testing. In the US, the Multi-State Agreement between Idaho, Oregon, and Utah (IOU) is in the draft stage. It would require a minimum age of 23, two years linehaul driving experience with double trailers, a clean driving record, a medical exam, and a road test prior to operating a long truck. In Canada, the Canadian Trucking Association has published a "LCV Driver's Manual" and has proposed a common testing procedure. To date the proposal has not been adopted in any province which currently allows LCV's but it is understood that Alberta and Saskatchewan are close to adopting a reciprocal license based on it.

The presumption in all of this is that long trucks operating under permit require more highly qualified drivers than other trucks. This presumption is not just limited to vehicle length. There is at least one province (Saskatchewan) where trucks operating under permit at higher-than-normal weights also have to be driven by more highly qualified drivers. This may be an isolated instance of where a jurisdiction requires special driving skills for overweight trucks; no survey of state or provincial practices on this matter is known.

The interesting point is that, while there may be a general presumption that extra long (or extra heavy) trucks require some sort of state regulation to ensure more highly qualified drivers, in practice these don't always exist or are not always enforced. That, at least, was the finding of a special investigation by the General Accounting Office (GAO) in the United States. "Many states, particularly in the West, have made little or no effort to establish special requirements for LCV drivers." [10] In a recent visit to one of these western LCV states by one of the authors, a motor carrier administrator agreed with the GAO findings. His attitude (a paraphrase of his comment) was: "Yes, it is true that we don't enforce the LCV driver requirements that we have. But, why should we? We don't have an LCV safety problem. The carriers make sure they only employ drivers capable of handling those trucks."[11]

Table 1: Driver Qualifications for LCV Operation

PROVINCE OR STATE[12]	SPECIAL DRIVER REQUIREMENTS
Alaska	No special requirements.
Alberta	2 years or 150,000 km experience with tractor-semitrailer. Skills & knowledge test.
Arizona	Triples: training by experienced driver of triples.
Colorado	No driver suspensions for 3 years.
Florida	Must be registered with Turn-pike Authority.
Idaho	No special requirements.
Indiana	26 years old. 5 years experience with tractor-semitrailer or double trailers. Experience driving in 4 seasons.
Kansas	Triples: 2 years experience with tractor-semitrailers and 1 year with double trailers.
Manitoba	Provincial Department of Transport must approve drivers.
Massachusetts	5 years experience with tractor-semitrailers. Registered with Turn-pike Authority.
Missouri	No special requirements.
Montana	Triple trailers: drivers must be certified (skill & knowledge tests) by operating company.
Nebraska	No accidents while driving LCV's.
Nevada	25 years old. Medical exam within the last 24 months.
New York	26 years old. 5 years experience with tractor-semitrailers. Tandem Trailer Driver Permit.
North Dakota	No special requirements.
Northwest Territory	No special requirements.
Ohio	26 years old. Driving experience in 4 seasons. Doubles: 5 years experience with tractor-semitrailers or short doubles. Triples: 5 years experience with permit doubles and special training.
Oklahoma	Triples: 2 years experience with tractor-semitrailers.
Oregon	No special requirements.
Saskatchewan	3 years experience with tractor-semitrailers. 2 years with non-permit doubles. Special course. Annual review of driving record. 'Points system' of merits/demerits.
South Dakota	No special requirements.
Québec	5 years experience with tractor-semitrailers.
Utah	Carriers must certify that drivers have a safe driving record and passed a road test with LCV's.
Washington	No special requirements.
Wyoming	No special requirements.

Proposed Policy Changes in Driver Qualifications

There are a number of policy changes regarding driver qualifications being proposed across the United States and Canada. These proposed changes range from graduated licences, to driver course accreditation, to common LCV licences. What follows are some examples of proposed policy changes.

Statistics show that younger drivers are more likely to be involved in accidents. "Young drivers (under 30), who make up less than 15 percent of truck drivers, accounted for 30 percent of the drivers of trucks involved in accidents."[13] Some jurisdictions are reviewing truck driver training procedures and examining the possibility of graduated licences for younger drivers. (eg, Ontario). The recent inquest in Ontario into deaths caused by runaway wheels resulted in, among other things, a recommendations for a graduated licensing system for Class A drivers (Ontario's 'Class A' is the NSC Class 1) and a recommendation for a special demerit-point system for commercial drivers. The new licensing system would control the size of vehicle driven by new drivers – both by GVW and number of trailers – and the distance and/or time on the road. Advocates of this measure believe that requiring younger drivers to gradually acquire experience at the wheel would increase their safety record; some also feel that the measure would encourage carriers to hire new drivers and train them to meet the needs of their specific operations. Opponents state that such a measure would add another layer of bureaucracy and red tape to already cash-strapped government agencies. It is understood that the provinces are thinking of extending this concept to commercial drivers. For example, it might be that in order to qualify for a Class 1 licence (articulated vehicle) a driver might have to already have a Class 3 (heavy truck) licence and a certain number of years experience.

Numerous agencies have put forth a recommendation for increased, higher quality, and more uniform training for drivers. The recommendations vary from training requirements for entry level drivers to those for special permit vehicles (such as long trucks). Driver training is not mandatory prior to receiving a CDL and the courses that do exist vary greatly – from 10 hours to 30 weeks. Efforts are being made to standardize training. In Canada, the Canadian Trucking Human Resources Council is offering accreditation to driving schools across Canada. To date, seven schools have been given accreditation. The program involves a set curriculum (4 weeks in class and 8 weeks of on-the-job training), and standards for instructors, classrooms, and vehicles. Reaction to the effort has been varied: some industry officials welcome the effort, others cannot see why it should be recognized. The program is geared to the needs of the industry in the hope that it will become a hiring condition for new drivers. So far, no mention of government imposed mandatory training has been made. As one official put it, "Unless industry grabs this thing and runs with

it, it's not going to fly." Further efforts are being made by: truck training schools (Ontario training schools have formed an association to better regulate their own industry); provinces (Saskatchewan recognizes training schools and certifies instructors); and the trucking industry (certain carriers operate their own training schools, eg. Reimer's). However, not all agree with the inadequacy of current driver training standards or that they are a safety issue. The Task Force on National Standardized Graduated Commercial Licensing found that "The general review did not reveal any major safety concerns related to industry training or quality concerns and the stratifying of existing Class 1 licences".[14]

Training requirements for special permit trucks such as long trucks (LCVs) vary by jurisdiction. Currently, efforts are being made to harmonize the process for jurisdictions within Canada and those within the United States. In the US, a state committee has recommended "the setting of standards...for entry level requirements and minimum driver training and experience necessary for operation 53 ft tractor semitrailer combinations, STAA twin combination vehicles, and all LCV's".[15] According to an FHWA official, the United States is currently in the process of publishing a "notice of rule making" on the topic. The drafted bill, if passed, would regulate the training requirements for LCV drivers on a federal level in the US. The specifics of the bill are not known at this time.

In Canada, the provinces have stated that "...the issue of a special licence class for Longer Combination Vehicles (LCV) be dealt with as a licence endorsement by those jurisdictions allowing those configurations on their highways."[16] This seems to fly in the face of an earlier statement advising a "training and endorsement proposal for LCV's".[17] The CTA, in the meantime, has published a Driver Manual for LCV's as reported above.

Finally, within the context of the current "Comprehensive Truck Size and Weight Study" now being conducted by the US Department of Transportation, there has been a suggestion that new driver qualification and training standards be developed for all drivers of straight trucks over 54,000 lbs. and all combination vehicles over 80,000 lbs. This is just a suggestion on the part of one person participating in a workshop held in Ann Arbor in June, 1995 and, therefore, has no official sanction or (likely) any possibility of being translated into policy. Nevertheless, it is instructive as it is indicative of the "political mood" in some quarters.

Addressing Driver Qualifications: Policy Issues

In order to understand the complexities of driver qualification regulations there are several policy issues that must be addressed. First, what are the existing qualifications for drivers? How good are they? How much would improved

qualifications add to road safety? In the absence of any quantitative answers to questions about the value of present driver qualifications, perhaps policy makers should heed the views of an industry observer. Stan Kimberly, a long-time industry observer, writes in the January/February issue of *Today's Trucking* that, "Inadequate training of drivers in Canada over the past eight or so years... means we have thousands of neophyte drivers on the highway without the benefit of training under the supervision of a fleet-management program."

The second issue to address is the scope of the problem. Highway safety, truck safety and driver qualifications need to be put in perspective. The perspective in Table 2 is that, whatever the nature of 'the problem', it is not a growing crisis. No one can make light of 44,486 trucks involved in reportable accidents in a year; and certainly, 554 accidents involving a fatality is something that needs to be addressed. But, to counter the impressions gained from a perusal of the press, it has to be said that truck safety appears to be improving. Between 1989 and 1993, the number of trucks involved in reportable accidents declined by 23%. This decline, of course, should be related to a measure of exposure so that real trends could be known more precisely. It is possible, if highly unlikely, that the recession in 1990-91 resulted in a 23% decline in trucking activity in Canada. More likely, truck activity has grown over the five years – commercial vehicle registrations grew by about 3%[18] – meaning that the accident rate has probably declined by more than the amount suggested in Table 2.

Table 2: Trucks Involved in Reportable Traffic Accidents

	Fatal Accidents	Total Accidents
1989	638	57,637
1990	616	53,285
1991	531	46,564
1992	479	45,783
1993	554	44,486

Source: P. Gutoskie, *Commercial Vehicle Safety in Canada* (The numbers shown refer to all straight trucks and all tractor-trailer combinations with a RGVW of 4.5 tonne of more.)

The most important issue to address is the effectiveness of all the effort that goes into programs aimed at controlling driver qualifications. There is a lot of effort: licence classes and all the mechanisms needed to manage this system; attempts at regulating/controlling/certifying driving training; hours of work regulations and the enormous industry associated with log books; and new substance abuse initiatives (for carriers operating into the United States). While all this effort may be necessary there is little documented evidence in Canada about the effectiveness of

these programs, and so it may be that the most effective manner of ensuring qualified drivers is to make the industry itself more responsible.

As an example of the effectiveness of making the industry responsible for safety, there is good evidence emerging from research at Northwestern University into the US federal government's auditing and safety-rating program.[19] The US federal government has been auditing carriers for some time to determine compliance with safety regulations which are similar to many features of the regulations the provinces agreed to under the NSC. Based on the audits, carriers are assigned a safety rating of "satisfactory," "conditional," or "unsatisfactory." Carriers with unsatisfactory ratings are re-audited at some stage in the future.

Several interesting findings emerge from a large-scale analysis of the results of these US carrier audits. Of most significance, carriers with satisfactory safety ratings have an accident rate less than half the rate for carriers with an unsatisfactory rating (0.59 accidents per million miles versus 1.29, where "accident" is all reportable accidents except those involving small amounts of property damage). Second, the research has found that carriers with an unsatisfactory rating, who are subject to further audits and various possible sanctions, improve their behaviour remarkably after initially receiving such a rating: the accident rate for these firms declines by 43% after they are initially detected and then subjected to further audits. Third, the research has been able to link accident rates to some aspects of management practices. Here are the findings:

> The second group of findings deals with those safety practices, investigated in federal safety audits, that are strongly related to accident experience. The keeping of records on accidents and using these data to take disciplinary [action] against or educational action for the drivers involved, appears to be a safety practice that is effective in reducing accident rates. Such actions may be a clear signal from management to drivers that the company takes accidents very seriously. There also appears to be a strong link between compliance with hours-of-work rules and accident rates. Firms that monitor and keep records of hours-of-service of drivers and use this information in dispatching have lower accident rates. When compared with the poor showing of [audit] questions on maintenance procedures, these findings are consistent with the well-documented fact that the vast majority of truck accidents are caused by driver error rather than mechanical failures. [20]

The 'lesson' here seems to be that truck safety can be managed if governments concentrate on, in the first instance, detecting good/bad management practices and, in the second instance, rewarding/penalizing such behaviour. To the extent driver

qualifications are a factor in truck safety (and they clearly are), motor carrier management can do much to control this factor.

In some ways, Canada is not nearly as advanced in the areas of safety ratings and audit procedures as is the United States – certainly not in the number of years or the scope of such programs. In Canada, recent revisions or additions to the National Safety Code have seen "Compliance Review-Safety Rating" and "Facility Audit" standards agreed upon. However, with the exception of a few provinces such as Ontario, there has not been much experience with auditing/rating so far. (Ontario has conducted something in the order of 10,000 audits since 1988.) Furthermore, as far as is known, there is no research in Canada on the effectiveness of such programs. Most provinces are planning to have their carrier rating systems in place sometime in 1996.

In other ways, Canada may be slightly more advanced than the United States. For one thing, the whole US auditing/rating program evaluated in the research referred to above has undergone significant change recently. Both countries are now working on new ways of auditing and rating, and there is some feeling that Canada's development of a carrier safety rating system is more advanced than United States' developments. Further, plans are proceeding in Canada to develop premier carrier programs – sort of a 'satisfactory plus' rating system – which will offer even more incentives to carriers with good safety practices. In the United States, such plans have either been cancelled or are on hold.

Knowing that there is a problem (though not a growing crisis) is one thing; knowing what policies are required to solve it is quite another. One difficulty here, is the absence of definitive model of accident causation. Without this, it is almost impossible to know how much effort should be devoted to the various factors that can either cause accidents or be a contributing factor: highway design, traffic regulations, weight and dimensions regulations, highway maintenance, driver controls, driver training, substance abuse regulations, training, education, audits/safety ratings, etc. The truth may well be that all of these programs have to have highway safety as one of their goals. That said, in a time of growing government cutbacks, there seems to be a compelling reason to find out which of these efforts is more or less effective than others. The emerging evidence from the United States on the effectiveness of safety-rating systems should be taken seriously. Perhaps the most effective manner of ensuring qualified drivers and improving highway safety is to make the industry itself more responsible.

NOTES AND REFERENCES

1. Gutoskie, Paul. 1995. *Commercial Vehicle Safety in Canada, 1989 - 1993*, Road Safety and Motor Vehicle Regulation Directorate, Transport Canada, Ottawa.

2. National Transportation Act Review Commission. 1993. *Competition in Transportation: Policy and Legislation in Review*, Supply & Services Canada, Ottawa, Vol 2, p 157.

3. Nix, F.P., 1994, *The Trucking Industry in Canada*, Canadian Trucking Research Institute. The numbers quoted use quite rough estimates of VMT from the Royal Commission on National Passenger Transportation.

4. Kansas City, March 12-15, 1995

5. Saccomanno, F.F., J.H. Shortreed and L. Craig. 1995. *Perspectives on Truck Safety: Issues and Recommendations emerging from the Truck Safety Conference: Perceptions and Reality, Toronto, September 11-13, 1995*. Draft, University of Waterloo.

6. Barry Holmes, *Motor Truck*, January 1996. It is not known where the figure of 80% comes from. In Ontario, based on their experience with auditing carriers, it is understood that these audits have determined that 41% of drivers falsify logs.

7. CCMTA, *Record of Decision, Standing Committee on Drivers & Vehicles*, Nov 10-11, 1995, p 10.

8. This was a component of a proposal discussed at a workshop organized by the US DoT at a workshop in Ann Arbor, Michigan on June 30, 1995.

9. Lyles, R.W., S. Narupiti and S. Johar. 1995. "Relationship Between CDL Test Scores and Driver Safety" paper #950074, Transportation Research Board annual meeting, Washington, DC.

10. United States General Accounting Office. 1993. *Longer Combination Trucks: Driver Controls and Equipment Inspection Should be Improved*, GAO/RCED-94-12, Washington, DC.

11. We are not reporting the state or the individual as the comment was made in a casual conversation. We did not ask permission to use this quote in a public document.

12. All states require, at a minimum, CDL with appropriate endorsement. From: *Federal Register, Vol 59, No.112 June 13,1994.*

13. TRB Special Report 227, *New Trucks for Greater Productivity and Less Road Wear*, Transportation Research Board, National Research Council, Washington, D.C. 1990.

14. CCMTA, *Record of Decisions, Standing Committee on Drivers and Vehicles*, Nov 5-6, 1994, item #16 (Task Force on National Standardized Graduated Commercial Licensing).

15. American Association of State Highway and Transportation Officials. 1995. *Report of the Subcommittee on Truck Size and Weight of the AASHTO Joint Committee on Domestic Freight Policy*. Washington, DC. p 7 "STAA twin combinations" are double 28 (in some cases 28.5) foot doubles which, under the federal Surface Transportation Assistance Act, all states have to allow on a designated network (the "National Network") of highways.

16. CCMTA op. cit. (Nov 5-6, 1994)

17. CCMTA, *Record of Decisions, Standing Committee on Drivers and Vehicles*, May 15-19, 1994, item #27 (Task Force on National Standardized Graduated Commercial Licensing).

18. Gutoskie, op. cit.

19. (i) Moses, L.N., and I. Savage. 1994. "The effect of firm characteristics on Truck Accidents," *Accident Analysis and Prevention*, Vol 26, #2, pp 173-179; (ii) Moses, L.N. and I. Savage. 1994. "A Cost-Benefit Analysis of United States Motor Carrier Safety Programs," working paper; (iii) Moses, L.N. and I Savage. 1995. "A Strategy for Identifying Dangerous Trucking Firms," working paper. Ian Savage attended the September 1995 Truck Safety Conference in Toronto and summarized this work.

20. Ibid. ("The effect of Firm Characteristics . . ") p 179.

Estimating O-D Travel from Link Traffic Counts: Can We Really Get There From Here?

by Bob Leore*

I. Introduction

Ask 100 transport practitioners what is the biggest gap in their knowledge and the vast majority would likely single out motor vehicle travel. The road motor vehicle is the dominant mode of travel in North America and yet we know virtually nothing about the patterns and structures of this activity. Passenger car travel exceeds the total of all the other public modes combined by a huge margin (5-10 to 1 depending on the measure). Trucking is not as dominant a mode for freight movement, but still accounts for 2/3 of the revenues earned by freight carriers and between 1/4 to 2/5 of the tonne-kilometres. By all accounts road transport is a major-league mode, yet we have at best bush-league data with which to analyse it. The basic sources should be the Trucking Origin-Destination survey for freight, and the Canadian Travel Survey for passengers, but their sample sizes are so small that it effectively precludes detailed city-by-city investigation. This has been the almost universal experience: to estimate travel for small cities with sufficient accuracy it is necessary to take such a large sample that it becomes prohibitively expensive to carry out a nationwide O-D survey. (Witness the 1993 U.S. Commodity Flow Survey which cost a whopping US$20 million and yet was still unable to estimate even some *national* level variables to within ±10%.)[1]

Understanding where vehicles travel is an important issue, so we cannot simply shrug and say ignorance is bliss. We should explore other, cheaper substitutes for obtaining data of acceptable quality. One neglected approach is to infer O-D trips from highway link traffic counts. Counts are routinely collected by provincial highways departments and are sufficiently consistent and accurate to constitute a valuable source of information on motor vehicle travel. Not only can important statistics such as aggregate vehicle-kilometres of travel for cars and trucks be calculated for a given highway network, but with the aid of trip distribution theory we can infer where vehicles are most likely to travel from and to throughout this network. The methods behind O-D estimation are surprisingly simple and the basic approach will be elucidated here. Although detailed empirical results must await a later date, preliminary work suggests that an O-D estimation model explains observed traffic patterns very well and yields a reasonable O-D trip table.

II. Can we really infer trips from traffic counts?

Link traffic counts are an aggregation of the myriad of O-D travel that is conducted everyday. When a vehicle passes over a counting sensor, it has been intercepted in its movement from origin to destination. A fundamental equation relates O-D travel with what we observe on our highways:

$$V_a = \sum_i \sum_j \rho_{ija} T_{ij} \qquad (1)$$

where V_a represents the observed traffic flow as represented by the counting sensor at a particular location in a highway network (many links, each with an assigned count, will be included in a vector **V** to define our road network), T_{ij} is the amount of travel that takes place between a given city-pair in this network (a very large number of city-pairs will appear in a vector **T**) and ρ_{ija} represents whether traffic between a given city-pair moves over a given highway link. The structure of this routing assignment matrix ρ can be modelled in different ways depending on the type of problem considered. Where congestion is an issue, traffic between a city-pair is likely to spread itself out over alternate paths and the cells of ρ will contain proportions of the total travel devoted to each path. In the intercity case where congestion is unimportant, O-D traffic will be fairly represented by single path or all-or-nothing assignment (i.e. the cells of ρ will contain a 1 if an (i,j) pair uses link "a" and a zero if the pair does not).

Usually, the number of highway links is vastly outnumbered by the number of city-pairs we want to study because many city-pairs will use the same road link to make a connection.[2] Although the unknown vector **T** could be estimated if we could invert ρ, this is normally not feasible owing to the dependencies that exist in ρ. We would need a separate network link joining each city-pair for this to work, an impossible requirement for an actual network. This effectively means that without further structure we are left with an infinite number of possible O-D combinations that can produce the observed set of traffic counts.

Since we are dealing with an interconnected highway network we need some other means for inferring the O-D trips from the observed traffic counts. This is where the principles of trip distribution theory can be brought to bear. If a generic structure can be imposed on the likely pattern of trip-making it could then be possible to estimate a unique O-D trip matrix with only a few parameters. We could then set up an optimisation routine to obtain parameter values that produce an O-D matrix that replicates the observed traffic counts when assigned to our highway network.

III. Trip distribution theory

1. Updating methods[3]

The use of counts in O-D estimation is an active field in transport research. The conventional approach is to work with an existing O-D matrix from surveys and then use mathematical programming techniques to update this structure using a current set of traffic counts. Although this may work in the urban travel setting where occasional O-D surveys have been conducted for some big cities, this method cannot be trusted to estimate an O-D matrix that does not exist. In the intercity case for Canada, we do not have an existing O-D matrix and must create one from scratch.

2. The Wills approach

a) The gravity model as centre-piece

Very little attention has been devoted in the transport literature to detailing how we can go about estimating an unknown intercity O-D matrix from traffic counts. In a pathbreaking Ph.D. thesis, Wills (1978) pioneered the essential tools for producing O-Ds directly from traffic counts. The centre-piece is the use of the well-known multiplicative gravity model to explain travel behaviour. Gravity models have been widely used in transport modelling, but no one, it appears, apart from Wills has used them to estimate O-D travel. Gravity models in transport employ an analogy with the Newtonian model of gravitational attraction in physics to explain travel behaviour. The general model is the following:

$$T_{ij} = \alpha \, (P_i P_j)^{\beta} \, D_{ij}^{\ \delta} \qquad (2)$$

where T_{ij} is the travel or attraction between two cities, P_i P_j is the product of the two cities' populations (a substitute for mass), D_{ij} is the distance separating the two cities, and α, β, and δ are parameters to be estimated. The basic structure can be generalised with the addition of further variables which are thought to affect the decision to travel such as employment, income, sales, and language or culture.

Only a small number of variables are required in the gravity model structure to explain travel behaviour between selected cities. This parametric approach has all sorts of applications not the least of which is the ability to make forecasts of future traffic levels, a critical input in highway planning.

b) The mechanics of Willsian O-D estimation

The trip distribution function says nothing by itself about what travel relationships are in reality. It is therefore necessary to embed the gravity model in a general optimisation program that relates it to the observed traffic counts. Wills has achieved this using fundamental equation (1). As this relation is intrinsically nonlinear it is not possible to transform the gravity model by logarithms and use standard regression methods. Instead, sophisticated nonlinear optimisation routines must be substituted. The ones that perform best in real-life problems seem to be of the quasi-Newton or variable metric family.

Employing the travel structure defined by (2) for T_{ij}, Wills uses the following algorithm to estimate the O-D matrix:

i) select initial values for the gravity model parameters (i.e. α, β, δ);
ii) compute the vector of O-D trips using the gravity model;
iii) assign this traffic to the links in the network using routing matrix ρ;
iv) compute the estimated traffic on links in the network (V^{est});
v) Compare this estimate against the observed link flows (V^{obs}) using a measure of distance such as least squares;
vi) Adjust the parameter set in the desired direction through repeated iteration until you cannot improve the sum of squares or other metric beyond a preset tolerance level.

At the conclusion of this optimisation routine, we are said to have, under certain conditions, a maximum likelihood estimate of the O-D matrix. In other words, this set of parameters and the accompanying O-D matrix is the most likely one that could have generated the observed traffic counts. This is a very powerful result, because, furthermore, if we can satisfy ourselves that the gravity model is the appropriate functional form for travel behaviour, then we have an O-D matrix that is *uniquely* consistent with the observed traffic. In short, we will have estimated the best O-D matrix for our set of cities without needing to conduct expensive surveys.

IV. Requirements for empirical estimation

Very little data are actually required to conduct an O-D estimation project. In essence, all we need is a network of road links with the observed traffic and distance of each, explanatory variables defined by a set of socio-economic data for the cities we want to study, and a set of computer programs for calculating the routing matrix, ρ, of shortest paths and for doing the actual nonlinear optimisation.

1. The network developed by Transport Canada

a) The centroid nodes

Since the summer of 1995, work has proceeded on building a detailed network linking all major urban areas in Canada and large neighbouring cities in the U.S. A total of 208 urban centres were ultimately selected for the analysis, of which 148 were Canadian and 60 were American. The nodes used to generate travel in the network are referred to as centroids to distinguish them from dummy nodes which do not generate traffic. All regions of Canada were covered including the Territories and a number of isolated areas. The geographic unit for Canada has largely been the set of Census Metropolitan Areas (i.e. urban core population exceeding 100,000) and Census Agglomerations (i.e. urban core population exceeding 10,000) although a number of exceptions were made to increase the coverage.[4] U.S. zones were predominantly Primary Metropolitan Statistical Areas (PMSA). Table 1 summarises the breakdown of centroids by province.

Table 1: Urban Areas Included in O-D Estimation, 1993

	Centroid nodes	('000s) Population	Average	Share of provincial population
Newfoundland	5	255	51	44%
P.E.I.	2	73	37	56%
Nova Scotia	7	574	82	62%
New Brunswick	7	400	57	53%
Quebec	31	5,239	169	73%
Ontario	43	9,041	210	84%
Manitoba	6	741	123	66%
Saskatchewan	9	553	61	55%
Alberta	12	1,972	164	74%
British Columbia	24	2,916	121	82%
Yukon	1	19	19	61%
N.W.T.	1	16	16	25%
Canada	148	21,800	147	76%
U.S.	60	87,265	1,454	34%
Total	208	109,065	524	38%

Three-fourths of the Canadian population is covered in the model although a lot of provincial variation exists. This is caused by the different patterns of settlement across provinces from the dispersed, largely rural patterns in Newfoundland, New Brunswick, and Saskatchewan, to the dense concentrations in Ontario and B.C., where, for example, 50% of the population of B.C. is contained within the Vancouver CMA.

b) The road network

The Canadian and northern U.S. highway system is represented as a network of links connected together by a large number of nodes. In order to join all the 208 centroids in the most direct manner, it was necessary to define a Canadian network of about 44,500 route-km and an American system of almost 38,000 route-km. The road network was laid out in the following manner: a road link extends from a given centroid until it hits an intersection with a road link used by other centroids. Nodes (either dummy or centroid) are then inserted at these intersections and the process is continued until all the centroid nodes have been linked together. For example, the links related to Ottawa include the following: from the junction of Highways 17 & 7 to Ottawa (about 25 km in length); from Ottawa to the junction of Hwys 17 & 417 farther east (about 10 km); from Ottawa to Hwy 401 along Hwy 16 (about 80 km); from Ottawa to Grenville, Que. using Hwy 148 (about 100 km); and Ottawa to Grand Remous, Que. using Hwy 105 (155 km).

Considerable effort is required to draw up the relevant network, especially in dense areas like southern Ontario and southern Quebec where numerous connections are needed to join the various centroids. Table 2 breaks down the extent of the network by province. As can be seen, Ontario and Quebec account for over half of the Canadian links but less than 40% of the route-km, a testament to the many centroids clustered fairly closely together. Their average link distances are significantly lower than the rest of the country.

Attached to each of these network links are three things: the length of the link, the observed traffic volume represented by average annual daily traffic (AADT), and the per cent of the traffic stream that is commercial vehicle (a broadly defined "truck" measure). Road lengths were largely obtained from provincial highway maps, while traffic data and vehicle splits were provided by provincial highways departments.[5]

The traffic volumes were applied to the links in two different ways. The first was to build an estimate of all of the traffic moving over the defined network. This was expressed in terms of total vehicle-kilometres, an aggregate measure of all vehicle traffic using the network. This measure includes the traffic generated by all of the hundreds of towns and cities located along our defined network and not just our centroids. The estimate of aggregate vehicle-km, totalling some 100 billion vehicle-km, is also shown in Table 2 and has been constructed by applying the hundreds of counts supplied by the provincial authorities to a finely detailed digital map created in the software program MapInfo (counts were attached to over 4,500 individual road sections in this map).

Table 2: Road Network Defined for O-D Estimation

Province	One-way links	Share	Route-km	Share	Average link distance	Total vehicle-km (billions)	Share
Ont.	219	33%	9,500	21%	40	38.4	38%
Que.	144	21%	7,600	17%	50	24.7	24%
B.C.	88	13%	7,900	18%	90	15.0	15%
Alta.	63	9%	6,100	14%	100	9.7	10%
Sask.	59	9%	4,700	11%	80	3.9	4%
N.B.	25	4%	1,800	4%	70	3.1	3%
N.S.	20	3%	1,400	3%	70	2.9	3%
Man.	37	5%	3,000	7%	80	2.2	2%
Nfld.	5	1%	1,100	2%	210	0.8	1%
P.E.I.	8	1%	200	0.4%	30	0.3	0%
Yuk.	4	1%	900	2%	230	0.2	0%
N.W.T.	1	0%	500	1%	540	0.04	0%
Canada	673	100%	44,700	100%	70	101.4	100%
U.S.	214		37,900		180		
Total	887		82,600		90		

Ontario and Quebec account for less than 40% of the network, but generate over 60% of the vehicle traffic. Contrast this with the Prairie provinces which have over 30% of the network but only 16% of the traffic.

A second measure of travel, which is used in the O-D estimation, attempts to exclude all the traffic that is generated by urban areas below the size of a Census Agglomeration and which does not represent intercity travel between the centroid nodes in our network. We do not want to try to explain all of the flow we observe on the network but only that flow which our 208 nodes have generated. This traffic subset is computed by taking the *minimum* observed traffic count for a defined link. Thus, for each of the 673 links in Canada, a scan was made of all the counts applying to each link and the smallest one was taken to represent the traffic generated by our centroid nodes. This process reduces the aggregate traffic flow by about 1/4 to around 73 billion vehicle-km.

The final bit of traffic data is the vehicle classifications. Although this is still a rather primitive practice, the technology for classifying types of vehicles in the traffic stream has accelerated greatly since the 1980s. Sophisticated machines are now capable of classifying many different types of vehicle from motorcycles to huge multi-trailer trucks. For our purposes it was sufficient to breakout passenger vehicles from commercial vehicles. Passenger vehicles include recreational vehicles along with cars and light trucks but would exclude multi-axle freight vehicles and passenger buses. Thus we are able to measure, roughly,

how much traffic is generated by "cars" and how much by "trucks". These amounts are depicted in Table 3.

Table 3: Total Travel Generated on the Network, Passenger and Commercial Vehicles, 1993

	Car		"Truck"		Truck % in	Average annual daily traffic		
Province	Vehicle-km (bill.)	Share	Vehicle-km (bill.)	Share	stream	Car	"Truck"	Total
Ont.	33.0	37%	5.4	42%	14%	9,500	1,600	11,100
Que.	21.8	25%	2.9	23%	12%	7,900	1,100	9,000
B.C.	13.7	15%	1.3	10%	9%	4,800	400	5,200
Alta.	8.4	9%	1.3	10%	14%	3,800	600	4,400
Sask.	3.4	4%	0.6	4%	14%	2,000	300	2,300
N.B.	2.6	3%	0.5	4%	16%	4,000	800	4,800
N.S.	2.6	3%	0.3	3%	12%	5,200	700	5,800
Man.	2.0	2%	0.3	2%	12%	1,800	300	2,100
Nfld.	0.7	1%	0.1	1%	15%	1,700	300	2,000
P.E.I.	0.3	0%	0.0	0%	9%	3,600	400	3,900
Yuk.	0.2	0%	0.0	0%	4%	600	23	600
N.W.T.	0.02	0%	0.0	0%	38%	100	100	200
Canada	88.5	100%	12.8	100%	13%	5,500	800	6,200

Commercial vehicles make up about 13% of the traffic stream overall, but this traffic is heavily concentrated in Ontario and Quebec. While the car shares are similar to the overall shares, Ontario and Quebec make up almost 2/3 of the truck traffic. The average annual daily traffic (AADT) figures show clearly how busy Ontario and Quebec roads are. Their AADTs are at least double that of any other province, averaging 9,000-11,000 vehicles per day.

2. The trip distribution model actually used in the O-D estimation

Once a network has been defined and the number of centroid nodes selected, we can proceed to do an actual O-D estimation. Since the trip distribution model is the key ingredient, it is necessary to specify the functional form and variables that will enter into it. The generalised form of the gravity model used for this inquiry is:

$$T_{ij} = \alpha\,(P_i P_j)^{\beta}\; D_{ij}^{\delta}\; L_{ij}^{\gamma}\; I_{ij}^{\theta}\; C_{ij}^{\mu} \qquad (3)$$

where $P_i P_j$ is the population product as before, D_{ij} the time-distance separating the nodes, L_{ij} is a linguistic pairing index defined as the difference in the percentages of each city's population with an English mother tongue, I_{ij} is the weighted average per capita personal income of the city-pair, and C_{ij} is a network centrality variable expressing the relative location of the nodes in the network. α, β, δ, γ, θ, μ are parameters to be estimated. The inclusion of the centrality variable takes account of the different propensities to travel for nodes located near the geographic centre of the network (defined roughly by highway 400

through central Ontario) compared with those on the periphery (i.e. a heartland-hinterland distinction).

3. Sources of socio-economic data for the explanatory variables

Canadian information on population and income for CMAs and CAs were taken from data generated by the regional model developed by the Conference Board of Canada. Data on language and for seven other urban areas were based on the demographic Census. Data for the 60 U.S. nodes were taken from the U.S. Bureau of the Census' Regional Economic Information System (REIS) CD-ROM. The centrality data were produced internally by the custom computer program designed for this project.

4. Structure of the optimisation routine used in the empirical application

A Fortran computer program (ODESSA) was developed by Michael Wills under contract to Transport Canada to carry out the actual O-D estimation. This program contains a number of interesting features. The assignment procedure by which the estimated O-Ds are loaded onto the links is based on the algorithm of Dijkstra (Dijkstra, 1959). It is an all-or-nothing or single path method in which all the travel of a given pair is routed uniquely down the path in the network with the shortest distance. This algorithm is so efficient that calculation of the 208^2 shortest paths literally takes seconds to complete on an 80486 computer.

The assessment of how well the estimated flows compare with the actual flows is accomplished by a simple log-likelihood function whose negative is minimised. This minimisation is the most time-consuming part of the estimation as a 40,000+ cell trip table needs to be assigned repeatedly to the links each time an evaluation (i.e. new set of parameter values) of the likelihood function is made. Runs usually require about 120-130 function evaluations before the parameter set stabilises. A full optimisation is normally completed in about 20-30 minutes.

V. Output

1. Typical parameter values[6]

Since the gravity model's form is multiplicative, like the Cobb-Douglas function in economics, the values of the parameters can be interpreted as constant elasticities defining the percentage change in O-D travel for a given city-pair resulting from a percentage change in the explanatory variables. Based on very early results, the parameter values seem to fall within the following ranges:

- Population products: between 0.3-0.5, implying a doubling of population will not quadruple trips but only double them at best;
- Distance: seems to cluster around the Newtonian value of -2.0 (i.e. the inverse square law), implying a very strong negative influence on trip making. A halving of the distance separating a pair, for example, will quadruple the number of trips.
- Language: usually below 0.7. This implies that the more similar is the linguistic makeup of a city-pair, the more travel will be generated but only at a less than proportional rate;
- Income: seems to lie between 0.7 and 1.0 implying that travel is relatively unresponsive to income; while this may be a counter-intuitive result it is unfortunately confounded by high variance in the estimate which frequently makes this parameter fail tests of statistical significance;
- Centrality: clusters around a value of -2.0, implying that relatively less travel takes place between nodes located in the centre of the network (i.e. high centrality index) compared with nodes located on the periphery. This might also seem counter-intuitive since the network's centre includes the big southern Ontario cities; however, what it might be picking up is that these heartland cities are normally self-sufficient and thus tend to have lower propensities to travel than peripheral hinterland nodes that need to travel to satisfy important functions they cannot supply themselves.

2. *Other output from ODESSA*

In addition to the basic parameter estimates, ODESSA also produces a great deal of information including the following:

- An Origin-Destination trip table for each city-pair defined;
- The goodness of fit of the optimisation as revealed by the pattern of the traffic flow residuals on each relevant link in the network;
- A distance matrix showing the length of the shortest-path connecting each city-pair;
- A routing matrix showing the itinerary of how each shortest-path route has been built up to link a given pair of cities;
- The number of shortest-path routes that use each link in the network;
- Measures of relative nodal centrality;
- Measures of accessibility for each node defined as either the total number of routes that pass through a node or the total traffic flow that passes through a node.

A couple of selected tables will help illustrate the kinds of data we can expect from this program.

a) O-D estimates

The O-D trip estimates are the most important piece of information produced by ODESSA. These data can be tabulated not only by city-pair, but for each city, province, and region as well. Table 4 ranks the centroids generating the most O-D travel in Canada. Not surprisingly, traffic is concentrated around the three largest Canadian cities, Toronto, Montreal, and Vancouver.

Table 4: Total Vehicle Trips Generated by Largest Centroid Nodes

Centroid	Province	Vehicle trips per day	Share of total trips generated
Toronto	Ont.	67,000	8%
Montreal	Que.	56,000	7%
Hamilton	Ont.	45,000	5%
Kitchener	Ont.	33,000	4%
Vancouver	B.C.	28,000	3%
Matsqui	B.C.	26,000	3%
Guelph	Ont.	23,000	3%
Oshawa	Ont.	21,000	2%
Brantford	Ont.	19,000	2%
St. Catharines	Ont.	19,000	2%
Chilliwack	B.C.	16,000	2%
Iberville	Que.	16,000	2%
London	Ont.	15,000	2%
Victoria	B.C.	13,000	2%
Saint-Jerome	Que.	13,000	2%

Overall, the O-D matrix is highly skewed with a few city-pairs accounting for the bulk of the trips generated. About 1/4 of the total trips is accounted for by the 10 largest pairs, and 3/5 is accounted for by the 100 largest pairs.

Table 5 reveals the pattern of intra- and inter-regional O-D flows. As with the vehicle-km data presented earlier, Ontario and Quebec account for almost 2/3 of the total O-D traffic, with Ontario at 42% and Quebec at 23%. It is startling as well to witness how little traffic is generated between regions. All regions' internal shares are in the high 80% to low 90% range, with travel to an adjacent region and travel to the U.S. making up most of the difference. About 8% of the total travel generated is between Canada and the U.S. The asymmetry in the inter-regional shares is also interesting. For example, while Ontario destinations represent 8% of Quebec's total travel, 5% of the Ontario-originating traffic is to Quebec. Similarly, while 6% of Atlantic Canada's travel is destined to Quebec, only 1% of the Quebec traffic is destined for Atlantic Canada.

Table 5: Preliminary Estimates of Regional O-D Flows, All Vehicles

	Destination						
	Atlantic	Quebec	Ontario	Prairies	B.C./Terr.	U.S.	Total
Share of each region's total O-D travel (%)							
Origin							
Atlantic	85	6	3	0	0	6	100
Quebec	1	88	8	0	0	3	100
Ontario	0	5	90	0	0	5	100
Prairies	0	0	1	87	6	5	100
B.C./Terr.	0	0	0	2	93	4	100
U.S.	2	7	26	4	8	53	100
Total	4	23	42	6	17	8	100

b) Overall goodness of fit

In conjunction with examining the O-D matrix to see if it produces a sensible trip distribution pattern, we can also examine the goodness of fit of the overall optimisation to see how well the estimated traffic flows fit the observed traffic counts. This is summarised by both the R^2 value and the percentage of flow explained. In the preliminary work that has been conducted with the simple model presented above, R^2 of at least .80 have been achieved. This means that over 80% of the variation in the traffic counts is explained by the model. Moreover, an aggregate error residual of about 15-20% has been achieved. Even though many links possess large positive (i.e. not enough traffic is assigned to a link) or negative (i.e. too much traffic is assigned) residuals, for most of the major Canadian highways such as Highways 401 and 417 in Ontario and Highways 20 and 40 in Quebec, residuals of under 10% are routinely produced. This is an encouraging result given that more elaborate modifications are currently being developed to boost further the model's explanatory power. That five basic variables could provide this much explanation for a complex phenomenon can only instil confidence that it is possible in principle to model intercity travel behaviour for the whole country.

c) Strategic factors

ODESSA also allows us to analyse the strategic importance of the links and nodes in the network. This is accomplished by the accessibility measures. If we find that certain links and nodes are associated with lots of traffic and O-D routes we can identify these strategic areas and use this as an input into the planning of future capacity improvements. For example, we find that if a node is high in terms of route accessibility (i.e. lots of O-D pairs pass through or terminate), it is usually low in terms of flow accessibility (i.e. lots of traffic

passes through or terminates). This is owing to the shape of the network where most of the traffic is located along Highway 401 and Highway 20 in Ontario and Quebec but most of the possible O-D routes travel along the Trans-Canada highway. A notable exception, though, is Montreal which ranks high under both accessibility measures. By identifying and ranking key areas for the whole country, one could aid greatly the conventional benefit-cost evaluations which tend to focus too narrowly on the total traffic moving over the link under examination and make little attempt to explain from where this traffic is being generated.

VI. Conclusion and next steps

Our objective has been to explain how we can estimate unknown road motor vehicle O-D travel using link traffic counts, and whether inexpensive practical methods can yield reasonable results. Based on preliminary work, we can say that a simple gravity model explains observed traffic patterns remarkably well. Over 80% of the intercity traffic is currently being explained by our simple model and the computed O-D matrix seems realistic. If we are satisfied that the model is appropriate then we will no longer need to conduct expensive O-D surveys to obtain this important information. What is most encouraging about the Wills approach used here is that more sophisticated methods currently being developed promise even better results. Two refinements deserve a brief mention:

- Trip distribution theory has inadequately recognised that distance is *not* the only measure of travel impedance. The size and location of cities that intervene between an O-D pair also affect the travel decision. The presence of intervening opportunities is a potentially potent determinant of the pattern of trip making since it is clear that persons will not travel as much to a city if a larger one comes between them and the destination (Wills, 1986). This effect is to be incorporated in a revised model and is expected to increase appreciably the goodness of fit.

- The forecasting ability of the current model is hampered by its restriction as a cross-sectional model fit to the traffic counts for a single year. The observed trip patterns will be influenced by the vagaries of the particular year that is selected and will make any forecasts risky. One solution is to pool several cross-sections and estimate as a panel accounting for changes to the network over space and time. This will be the first time a pooled cross-sectional O-D estimation model will have been attempted and this approach not only should improve estimates for each year but should also greatly enhance the ability to forecast future traffic levels.

Notes

* Transport Canada. The views expressed here are those of the author and do not necessarily reflect the views of Transport Canada.

I should like to thank Marc Gaudry for his wise recommendation that helped steer us in the right direction from the outset. I should also like to thank David Stambrook for his helpful comments on the paper and for providing much support during the course of this project. Lastly, and most importantly, I should like to thank Michael Wills for devoting considerable effort to rehabilitating and later extending, in new and creative ways, his empirical estimation software. This research project could not have been conducted without his active participation.

[1] 95% confidence intervals, approximated by two standard errors (See U.S. Bureau of the Census (1995), Appendix B).

[2] This is a severely under-determined system of equations where the number of variables (O-D pairs) greatly exceeds the number of equations (network links).

[3] One can find numerous examples of this approach in the major journals, but some especially useful ones are: Van Zuylen and Willumsen (1980), Bell (1983), Cascetta and Nguyen (1988), and Tamin and Willumsen (1989).

[4] The following seven non-CA nodes were added to increase the geographic coverage: Yarmouth, N.S., Newcastle/Chatham, N.B., Gaspé, Que., Kapuskasing, Ont., Flin Flon, Man., Banff/Canmore, Alta., and Castlegar/Nelson/Trail, B.C. In addition, the CMA Sydney, N.S. was divided into two smaller areas to aid in the assignment of traffic on Cape Breton Island.

[5] Since we are concerned with explaining Canadian travel patterns, we have not obtained traffic counts from the U.S. Although a U.S. network is defined in order to connect the U.S. nodes to the border, no traffic is attached. The inclusion of U.S. traffic will be considered in future work.

[6] Specific parameter estimates from the latest calibrations will be presented at the CTRF meeting.

References

Bell, M.G.H. (1983) The Estimation of an Origin-Destination Matrix from Traffic Counts. *Transportation Science*, Vol. 17, No. 2, pp. 198-217.

Cascetta, E. and Nguyen, S. (1988) A Unified Framework for Estimating or Updating Origin/Destination Matrices from Traffic Counts. *Transportation Research*, Vol. 22B, No. 6, pp. 437-455.

Dijkstra, E.W. (1959) A Note on Two Problems in Connection with Graphs, *Numerische Mathematik*, Vol. 1, pp. 269-271.

Tamin, O. and Willumsen, L. (1989) Transport Demand Model Estimation from Traffic Counts. *Transportation*, Vol. 16, pp. 3-26.

U.S. Bureau of the Census (1995) *1993 Commodity Flow Survey: United States, Preliminary.* TC92-CF-52(P)

Van Zuylen, H. and Willumsen, L. (1980) The Most Likely Trip Matrix Estimated from Traffic Counts. *Transportation Research*. Vol. 14B, pp. 281-293.

Wills, M.J. (1978) *Linear and Nonlinear Estimators of the O-D Matrix*, Unpublished Ph.D. thesis, Department of Geography, University of British Columbia.

Wills, M.J. (1986) A Flexible Gravity-Opportunities Model for Trip Distribution. *Transportation Research*. Vol. 20B, No. 2, pp. 89-111.

Aggregate Air Passenger Forecasting - Winnipeg International Airport Activity Forecasts

by Darryl Howery, Ellen Taylor, Drina Culo *

This paper provides an overview of an air passenger forecasting model developed in 1994 by Applications Management Consulting Ltd. for the Winnipeg Airport Authority. The model describes aggregate air service demand and supply for the top eight airports in Canada. The demand equation estimates total origin/destination passenger activity using a variety of economic and aviation industry variables, and includes air service supply (aircraft movements) utilizing the instrumental variable estimation technique. Results are presented separately for domestic and transborder/international sectors.

Background

The transfer of management responsibility and control of the Winnipeg International Airport from Transport Canada to the Winnipeg Airport Authority has involved the negotiation of a complex lease agreement. A Base Case financial analysis prepared by Transport Canada underlies the proposed lease arrangements. The key drivers of the major revenue categories in the Base Case financial analysis are Transport Canada's projections of enplaned/deplaned (E/D) passengers, origin/destination (O/D) passengers and itinerant aircraft movements. The aviation projections for the Winnipeg International Airport initially provided by Transport Canada appeared to be overly optimistic, and these effects were reflected in the overall results of the Base Case financial analysis. Recognizing the potential impacts of alternative projections of aviation activity in their negotiation of a lease agreement with Transport Canada, the Winnipeg Airport Authority retained Applications Management Consulting in 1994 to develop a customized forecasting model of air passenger activity for the airport.

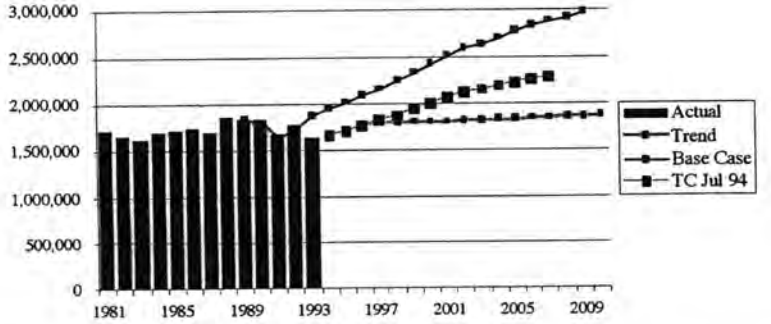
Transport Canada's Base Case: Up, Up and Away

Transport Canada's Base Case projections are provided for a forecast period of 20 years, between 1990 and 2010. Over this period, E/D and O/D passengers are projected to increase substantially, at an average annual growth rate exceeding

* Applications Management Consulting Ltd.

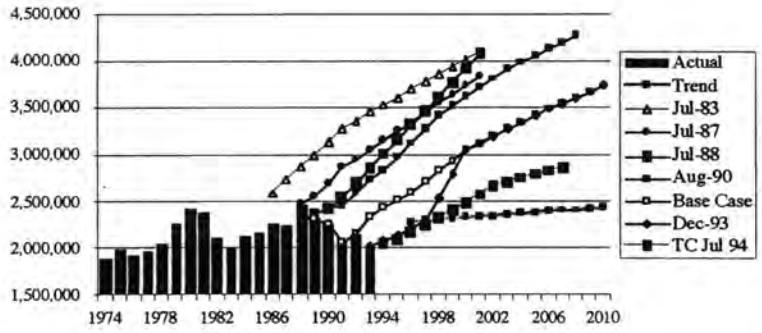
2.5 percent.[1] In light of the historic growth of air passenger activity at the Winnipeg International Airport, the Base Case projected levels of aviation activity appeared to be optimistic. Even revised "preliminary" projections by Transport Canada (July 1994)[2] exceeded projections of historic trends by almost 20% by 2007. The revised forecasts are more conservative than those provided in the Base Case as a result of downward adjustments to the growth rates of Winnipeg population and per capita personal disposable income. However, Winnipeg Gross Domestic Product is assumed to grow at a faster rate compared to the assumption for this variable in the Base Case forecasts.

Total O/D Passenger Activity - Actual vs Base Case
Winnipeg International Airport



Over a 16 year period, enplaned/deplaned passenger activity had increased from approximately 1.96 million (1977) to 2.14 million (1992). This represents a total increase of approximately 9 percent. During this same period (1977 to 1992), E/D passenger totals ranged between 2.0 million and 2.46 million, never exceeding 2.5 million passenger per year. A trend analysis using the 16 years of historical data imply modest growth in E/D passengers over the forecast period, increasing to almost 2.4 million passengers by 2007. This is significantly lower than the 3.5 million projected in the Base Case and the 2.87 million projected in the revised "preliminary" projections (July 1994).

Various E/D Passenger Activity Forecasts
Winnipeg International Airport



Origin/destination passenger volumes had not changed significantly since 1981. In fact, domestic and transborder O/D passenger volumes on major and regional/local carriers have declined since 1981. Over the 12 year historical period for which data was analyzed, O/D passenger volumes ranged between 1.6 million passengers per year to almost 1.86 million passengers, and never exceeded 1.9 million per year. A trend analysis using the 12 years of historical data imply minimal growth over the forecast period, reaching just over 1.85 million passengers by 2007. Once again, this is in direct contrast to the significant growth projected in the Base Case, which projects O/D passenger volumes to reach almost 2.88 million by 2007. Similarly, the revised "preliminary" projections are for O/D passenger traffic at Winnipeg to reach 2.28 million by 2007, which is significantly greater than the historical trend would imply.

Overview of the Forecasting Model

Because of the importance of activity forecasts in the developing financial projections for an airport, their reliability is of significant concern. Variability, especially the risk that the Base Case forecasts, and to a lesser extent the revised forecasts, are optimistic, can have a significant impact on the financial viability of the Winnipeg Airport Authority. To test the validity of the Transport Canada forecasts employed in the Base Case (and revised), and the sensitivity of the financial analysis results to this variability, alternate forecasts were developed utilizing an air passenger forecasting model prepared by Applications Management Consulting.

The air passenger forecasting model describes air service demand and supply for the top eight airports in Canada.[3] It utilizes site specific data for a variety of economic and aviation industry parameters, which are discussed in more detail below. The demand equation estimates total origin/destination passenger activity separately for domestic and transborder/international sectors. It includes air service supply (Aircraft Movements) utilizing the instrumental variable estimation technique.

Projections of the Independent Variables

The projections of origin/destination passengers at the Winnipeg International Airport are formulated on the basis of a multiple regression forecasting model. The demand equation of the forecasting model utilizes 12 "independent" variables (with origin/destination passengers specified as the "dependent" variable). The independent variables are listed below.

- **Employment (Winnipeg):** Related to the Toll Calls per employment variable.
- **Population (Manitoba):** Related to the provincial GDP per capita variable.
- **Urban Income:** The number of persons in Manitoba urban areas (i.e. population 100,000 and over) with real incomes (in $1986 constant dollars) greater than $60,000 per year.
- **Domestic Fares:** The average domestic fare index.
- **International Fares:** The average international fare index.
- **Real GDP Per Capita:** Real provincial Gross Domestic Product per capita (in constant $1986).
- **Toll Calls Per Employment**: The number of long distance toll calls from Manitoba, per employment.
- **Domestic Aircraft Movements (Instrumental Variable):** The number of movements, at the Winnipeg International Airport, of aircraft departing to or arriving from another point in Canada. For the forecast years, this variable is estimated using the "Instrumental Variable" technique, which involves estimating a separate equation with the following independent variables:
  - Aircraft: number of major carrier and charter operator aircraft in operation
  - Fuel Costs: index of fuel costs
  - Hours Flown: by major carrier and charter operators, adjusted for sites
  - City Dummy Variables

- **Transborder and International Aircraft Movements (Instrumental Variable):** The number of transborder aircraft movements (aircraft departing to or arriving from a point in the U.S.) and international movements (aircraft departing to or arriving from points in countries other than Canada and the U.S.) at the Winnipeg International Airport. For the forecast years, this variable is estimated using the "Instrumental Variable" technique, which involves estimating a separate equation with the following independent variables:
  - Aircraft: number of major carrier and charter operator aircraft in operation
  - Fuel Costs: index of fuel costs
  - Hours Flown: by major carrier and charter operators, adjusted for sites
  - Real Total Exports: total real provincial exports ($1986)
  - City Dummy Variables
- **City Dummy Variables:** Dummy variables were included for each city to capture site specific demand factors.

Alternate Activity Scenarios

Alternate projections for each of the independent variables have been developed to capture a possible range of results, given the high degree of complexity and uncertainty related to the prediction of air passenger volumes over the forecast period (1994 to 2010). The alternate forecasts include the following:

- **Medium Scenario:** Represents the expected, or most likely projection.
- **Low Scenario:** Represents an estimate of the downside variation in activity which may realistically be expected.
- **High Scenario:** Represents the most optimistic activity projection.
- **Trend Scenario:** The trend scenario has been developed utilizing the "trend" projections calculated for each of the independent variables in the forecasting model. The 'trend' projections are a straight line extrapolation of recent historic information for each variable. Obviously, in many instances, our view of the future deviates from the historic trend. This scenario is provided for information purposes only.

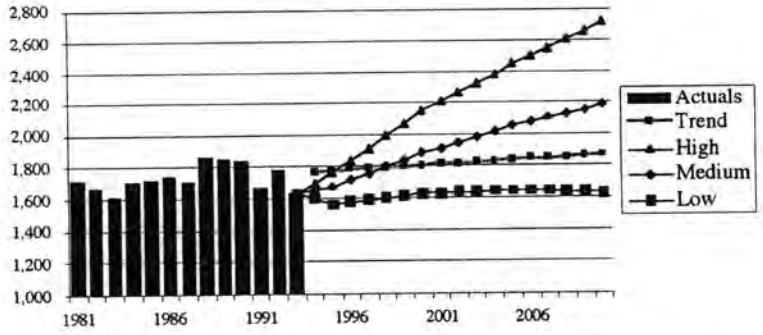
Alternate Scenario Origin/Destination Passenger Forecasts

The Medium Scenario forecasts project a modest increase in annual O/D passenger activity, increasing to approximately 2.2 million passengers in 2010. This represents a total increase of 34 percent from 1992, and an average annual growth rate of approximately 2.0 percent. By the end of the forecast period, the Trend Scenario projects annual O/D passengers to be approximately 1.9 million passengers, versus 2.2 million in the Medium Scenario.

The Low Scenario forecast projects O/D passenger volumes to initially decline, due in large part to a continued adjustment in domestic air service capacity. Thereafter, growth for Winnipeg is slow due to very modest rates of population and economic growth. By the end of the forecast period, Low Scenario O/D passengers total approximately 1.6 million.

The High Scenario forecast mirrors the growth projected in the Base Case, increasing at a 3.1 percent annual rate of growth and reaching just over 2.7 million O/D passengers by 2010.

Alternate Total O/D Passenger Activity Forecast
Winnipeg International Airport (000's)

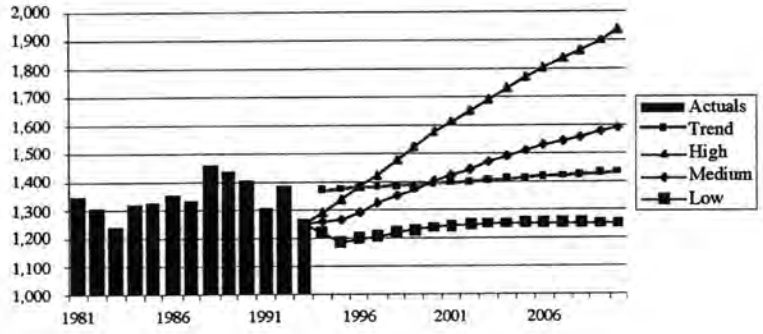


Domestic O/D Passenger Activity Forecasts

The Medium Scenario forecast projections reflect modest growth in the domestic sector. This is evidenced by the Medium Scenario lagging behind the Trend Scenario until 1997. Total Medium Scenario domestic O/D passenger volumes are projected to reach approximately 1.6 million by 2010. This reflects an annual growth rate of 1.4 percent.

The Low Scenario is characterized by a decline in O/D passenger volumes, which increase modestly to just under 1.25 million by 2010. In contrast, the High Scenario forecast projects total O/D passenger forecasts to reach over 1.9 million by 2010.

Domestic O/D Passenger Activity Forecast
Winnipeg International Airport (000's)



International O/D Passenger Activity Forecasts

Consistent with the view expressed in the Base Case, our Medium Scenario projects significantly greater growth in International O/D passengers (including both transborder and other international traffic). This is consistent with the general trend for transborder and international business and leisure trips to comprise a larger share of total trips in Canada. It is expected that this trend will be reinforced with the globalization of the world economies as evidenced by the passage of NAFTA and other possible trade agreements. It is also expected that there will be liberalization of existing air bilateral agreements with the US and other countries which will allow market forces to play a greater role in determining which international markets are served.

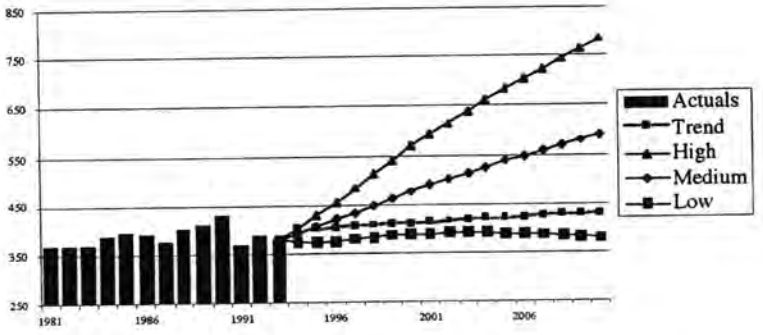
The Medium Scenario forecasts project international O/D activity to increase to approximately 590,000 passengers per year by 2010. This reflects a total increase of 56 percent from the 1992 estimate, and an average annual growth rate of 2.6 percent.

The Low Scenario is less optimistic about the potential for economic growth on the Winnipeg economy due to globalization, and the extent to which international service is provided from Winnipeg (either because of slower

liberalization of bilaterals or the desire for carriers to provide transborder and other international services from higher priority markets). As a result, the Low Scenario projects little change in total transborder/international O/D for Winnipeg.

The High Scenario projects significant growth in international OD traffic, increasing to over 780,000 passengers per year. This represents a doubling of transborder/international traffic, or an average annual growth rate of approximately 4.4 percent over the forecast period.

Transborder/International O/D Passenger Activity Forecast
Winnipeg International Airport (000's)



Alternate Scenario Enplaned/Deplaned Passenger Forecasts

Enplaned/Deplaned passenger forecasts have been based upon the O/D forecasts presented above, and assumptions regarding average ratio of E/D passengers to O/D passengers. To simplify the analysis, the ratios employed in these forecasts have been taken from "Transport Canada Aviation Forecasts 'Top 30' Airports" (December 1993). Historically, Winnipeg's ratio of E/D passengers to O/D passengers has varied between 1.23 and 1.32. Transport Canada has projected (Dec. 1993) that to the year 2000, this ratio will be 1.25, and between 2001 and 2005 it will be 1.24. In 1993, the E/D to O/D ratio was 1.32 for Domestic traffic and 0.99 for Transborder/International traffic. These assumptions have been employed in the development of the Alternate E/D Forecasts presented below.

The Medium Scenario forecasts project E/D passengers to increase to almost 2.7 million by 2010. This reflects a total increase of 33 percent over 1992 and an average annual rate of change of approximately 1.7 percent.

The Low Scenario forecast projects total E/D passengers to decline somewhat initially then increase modestly to approximately 2.0 million passengers by 2010. The High Scenario (Base Case simulation) projects total ED passengers to total approximately 3.3 million by 2010.

Alternate Aircraft Movement Forecasts

Alternate projections of Aircraft Movements (not including General Aviation) have been developed based upon our E/D passenger forecasts and assumptions regarding the aircraft capacity and load factors for each the carrier types (e.g. Major, Regional/Local and Charter). The updated Transport Canada "preliminary" projections (July 1994) have assumed increases in aircraft utilization over the forecast period. Where possible, those assumptions have been incorporated in the development of the Alternate Aircraft Movement forecasts presented below.

The Medium Scenario aircraft movement forecast projects total non-General Aviation traffic to increase to approximately 63,500 movements by 2010. This represents an average annual rate of increase of approximately 0.7 percent. This relatively low growth rate reflects a short term domestic capacity adjustment and an increase in aircraft utilization over the short term which continues to be enjoyed over the forecast period.

The Low Scenario projects total non-General Aviation traffic to decline to approximately 57,000 movements. This reflects the capacity adjustment made by carriers on domestic service and the relatively slow rate of total passenger traffic growth in this scenario. The High Scenario projects total non-General Aviation traffic to increase to approximately 90,000 movements by 2010.

Alternate Scenario Forecasts Versus Transport Canada (July 94)

In general, the revised Transport Canada "preliminary" projections are between the Medium and High Alternate Scenario forecasts. Typical of this are the E/D passenger forecasts. The July 1994 "preliminary" Transport Canada forecasts project total E/D passenger traffic to increase by approximately 41% between 1994 and 2007. This is only marginally lower than the original Base Case forecasts, which projected an increase in E/D passengers over the same period to be approximately 45%.

Conversely, the Medium Scenario Alternate Forecasts project total E/D traffic at the Winnipeg International Airport to increase by approximately 28% over the period 1994 to 2007. While modest, compared to the Transport Canada forecasts, this rate of increase is significantly greater than the Trend line. The most optimistic Alternate Forecasts, project total E/D activity to increase by approximately 55% over the same period. Thus, while the July 1994 "preliminary" forecasts appear to be within the range of potential activity levels, they clearly are on the optimistic side of this range.

Most of the disparity between the July 1994 "preliminary" Transport Canada projections and the Alternate Forecasts presented in this report is due to Domestic traffic projections. Projections of Transborder/International traffic are very similar over the period 1994 to 2007. Transport Canada's July 1994 "preliminary" forecasts project domestic E/D traffic at Winnipeg to increase by approximately 40% to 2007. The Alternate Forecasts project Domestic E/D traffic to increase by approximately 23% over the same period.

Conclusion

The revisions to the Transport Canada forecasts clearly indicate that the figures utilized in the Base Case are optimistic. This is evidenced by the downward revisions indicated by the "preliminary" July 1994 forecasts, as well as the recent declines in activity at the Winnipeg International Airport. The July 1994 "preliminary" Transport Canada projections typically fall between the Alternate Forecast Medium and High Scenarios. As a result, it can be expected that the recently revised "preliminary" forecasts for Winnipeg are also optimistic.

Due to these factors, the conclusion can be drawn that the revised activity forecasts for the Winnipeg International Airport should be deemed as optimistic projections of the range within which activity at the Winnipeg International Airport can reasonably be expected to vary. The Medium Scenario and Low Scenario alternate activity forecasts presented in this report are appropriate projections, based on the robustness of the forecasting model, to test the sensitivity of the financial viability of the Authority for reasonable variations in future airport activity.

---

1 E/D passengers are projected to increase to almost 3.7 million by 2010, which represents an average annual growth rate of approximately 2.6%. Similarly, O/D passengers are projected to increase to nearly 3.0 million by the end of the forecast period, which represents an average annual rate of increase of 2.7%.

2 Transport Canada, *Preliminary Aviation Passenger Forecasts - Winnipeg International Airport*, July 1994.

3 These include L.B. Pearson International Airport, Vancouver International,

Dorval International Airport, Calgary International Airport, Ottawa International Airport, Winnipeg International Airport, Halifax International Airport and Mirabel International Airport.

COMMERCIAL VENTURES AT PARK+RIDE TERMINI

David M. Cape,
Ryerson Polytechnic University and Lancaster University

1. This paper examines commercial ventures at Park+Ride termini in the context of Transit Transfer Centres (TTC's) and Multi-Purpose Centres (MPC's). Both are logical extensions of the pure Park+Ride suburban terminus.

2. DESCRIPTIONS

a. An early use of TTC's happened in Vancouver after the election of the NDP government in 1974, with the application of the hub-and-spoke concept by B.C. Transit. There, buses from local areas converged at regular intervals and passengers transferred to the through bus to downtown with no lay-over often using a limited stop service. These TTC's were basically located on cheap land remote from any other facilities in the fare-paid zone. Several operators use this concept.

b. MPC's differ from TTC's in that other activities occur at transit connection points, most such activities are commercial in nature. An "infant industry", not many MPC's are yet operational.

c. Park+Ride suburban termini.

Park+Ride (P+R) systems also originated to provide cheap, efficient access to congested urban centres. The first and most commonly occurring systems were the rail based systems. A parking lot proximate to a rail station allowed commuters and shoppers easy access to their destinations, particularly where either local transit was either unnecessary or readily available at destination. The commuting to London from the "stock-broker belt", south-west of London, and the GO Lakeshore service to Toronto exemplify. In some station areas, there are commercial ventures either *in situ* or adjacent.

A more recent development has been the bus-based systems. Oxford typifies such. It is a city where rivers meander, blocking bridge and artery construction, while the density of "listed" (non-removable) buildings hamper construction of more roads. One slight variation on the theme occurs at Maidstone, where Park+Sail is practiced using a local waterway.

3. DEMAND FACTORS

Urban congestion has been a common phenomenon from the days of the horse, through those of the horse-drawn carriage, and the horseless ones. It is also

common knowledge that road speeds have remained constant at 12 miles per hour throughout, and with spreading cities this led to longer commute times and the pressing need for a rapid transit alternative.

Declines in downtown manufacturing, rises in suburbanized office activity, the separation of "office work from a fixed office, and an aging population have spurred new needs and new opportunities for the urban transit community. With changing urban land use including also the replacement of low cost housing with gentrification, recent trends show falling transit ridership and "re-engineering" the downtown area as land owners attempt to stem declines in land values by creating supplies of new attractions. The 1950's suburban commute although declining in many places is being augmented with other purpose trips.

4. SUPPLY FACTORS

Let us now look at the factors which led to the growth of P+R and the future prospects.

i. cheap land

The first leg of the introduction of successful P+R services has been the presence of relatively cheap land to act as a suburban collector parking lot. As time progresses with a developing successful service, land contiguous to the suburban terminus appreciates in value as some commuters want to live within walking distance to the P+R terminus. Over time, P+R termini will alter the land use and pricing in a given area. Unfortunately, some security problems often arise when large numbers of cars are left unattended throughout a working day. This problem is one for which the presence of commercial ventures is helpful.

ii. fast service

The second leg is a more economic service, either faster line haul, or shorter walk to ultimate destination, or cheaper than using downtown commercial parking lots. In some cases, reserved right of way for buses (Ottawa), or frequent, convenient downtown pick up points (Oxford) provide a *raison d'etre.*

iii. convenient connections

In order to attract ridership away from the personal automobile, attractive services were necessary, convenience being paramount. The personal automobile is convenience *par excellence*, it provides the third support leg. Own choice of arrival/departure times and routings, to smoke or not, to listen to radio/music or not, to avoid close contact with other individuals with perhaps unsavoury personal traits all lure us away from efficient public transit carriers.

5. P+R SUBURBAN TERMINI

a. Facilities *in situ*

Facilities at P+R termini are frequently desirable in terms of toilets, shelters, ticket sales points, and telephones. Where present these sites are frequently unattended, In Chester (UK) a security guard is on site at open hours. The rise

of multi-purpose centres (MPC's) as transit termini may well change monitoring processes.

b. Problem areas and attempts to obviate crime.

From a potential user's perspective, P+R suburban termini have several requisites. Among these: ease of use, including access, egress and transfer from car to carrier; personal safety and security; and safety and security for the auto from vandals and purloiners. Adequate lighting and (monitored) closed circuit television or a security guard *in situ* help tremendously. But "help" is all they can do. As (potential) miscreants increasingly realize the vulnerability of both a late returning rider and a car parked for almost pre-defined hours, the likelihood of crime increases.

One way to discourage crime is to keep a steady flow of traffic, both pedestrian and road. In England, some P+R facilities have recycling points located there, so casual passers-by may visit the site, adding to the presence of legitimate users. In Canada, where home collection of recycling has burgeoned, this opportunity is missing.

c. Commercial ventures as a natural security device.

Although commercial ventures sometimes lend themselves to crime victimization, generally hold-ups, the largest impact is to attract business and custom. This has the dual effects of attracting P+R clientele and discouraging crime. The economics of commercial ventures at multi-purpose centres pose a dilemma: demand is severely peaked, but the supply of commerce is needed over a longer working day.

i. as a convenience to users

Besides the facilities above (shelters, toilets, and telephones), several convenience store types have potential for P+R termini. Although it might be nice to have these stores open for the full day, opening when the services start and closing after the last bus leaves, even four hour shifts, may prove economic. The operative word is "may", however. In Toronto, from the 1960's to the 1980's many subway stations had newspaper kiosks inside, open only at rush hours. Many of these closed in the 1980's, with perhaps only the two Yonge-Bloor stores remaining, albeit all day long, and a rather craftily built "McDonalds" at the Dundas West Station which caters to both fare-paid passengers and walk-ins from Dundas Street. This too, is an all-day facility but caters to both transit and non-transit users.

Besides newsstands, however, other possibilities emerge at P+R termini. These would generally be either impulse purchases or fast transaction purchases. They include candy and card shops, stamp machines and mail boxes, banks and booksellers, shoe repairers, laundry and dry cleaners, fast food take-out stores and florists, bakeries and milk-marts, to say nothing of video rental concessions. The latter is a "no-brainer", but are there any at P+R termini? A rising

phenomenon is the child care centre. As these would be generally outside the "fare paid zone", they could attract non-P+R clients - a simple re-admission pass to transit issued by the merchant could when accompanied by a valid transfer[1] ensure the P+R carrier no lost revenue. In Canada ATM's (Instant Teller machines -basically all from the Canadian Imperial Bank of Commerce) have largely disappeared from the fare paid zone on the Toronto Transit Commission property, although at least one site has been taken over by an independent at the Spadina station. Even with Toronto's very large passenger base, something was obviously wrong with the economics, costs too high (rents?) or profits too small. For small cities and towns, possibly the banking concept may need some work, but other opportunities look more promising. We will see below some initiatives in the above areas.

d. as a convenience to locals

As areas contiguous to P+R termini expand with residences, so too expands the logical location for service outlets. On a far grander scale Toronto's Yorkdale Shopping Centre first built in 1952 has expanded to include a P+R terminus since the subway system extended to serve Yorkdale. A pricing policy has been put in place to ensure both that commuters do not park free and do not encroach on all the centre's parking space. What space has not been occupied by mall-opening time is available free of charge to shoppers or commuters. The commerce-transit marriage can work!

6. SOME PRESENT INITIATIVES

Here follows a selection of commercial initiatives that reflect some current thinking.

a. Canada

In Union Station, Toronto, the waiting area for GO (commuter) trains as well as VIA (intercity) trains, includes upwards of twenty merchants.

b. Germany

In Germany, at least one high-rise P+R terminus has been built, providing undercover parking and convenience stores.

c. United Kingdom

In York, Marks and Spencer Stores has linked itself with the P+R services, although in the city centre, rather than in the suburbs. M&S has contributed financially to the P+R operation and has provided a waiting space for Park and Riders, along with some coin operated lockers, as a *quid pro quo* for having the P+R services terminate outside the store.

---

[1] Obviously short stop-overs would need to be condoned.

d. United States

A selection of recent notes from the U.S. transit industry reveals the following quotes.[2]

i. Houston TX:

Allows certain concessions to lease space at selected P&R lots and transit centres: daycare, dry cleaning, shoeshining & repair, daily newspapers (etc.) [3]

ii. Seattle WN:

a) Seattle Metro leased spaces for espresso stands at "half a dozen" of its largest P&R lots (Aug'93). Multi-Service termini are planned. [4]

b) plans rental outlets for books on tapes, videotapes, and portable tape players.[5]

iii. Bellevue WN:

The Bellevue Transit Center (a downtown facility) has had a positive impact on the Bellevue CBD supporting land use policy that has enabled development to occur by increasing the modal split share accruing to transit. 458 bus trips/day; 5362 passengers [6].

iv. Oakland CA

The National Station Car Association announced that three demonstration vehicles have been approved to begin service this year. Station cars are battery powered transit vehicles used by transit riders for improved access to mass transit. Demonstrations planned for Miami, Long Island, Northeastern Illinois, Sacramento CA, and the San Francisco Bay Area. Suppliers are Solectria (Massachusetts), Suntera (Hawaii), and U.S. Electricar (California).[7]

---

[2] The author is indebted to Brendon Hemily of the Canadian Urban Transit Association for material for this section.

[3] The Urban Transportation Monitor May 13, 1994 p.5

[4] The Urban Transportation Monitor May 13, 1994 p.5 Esther Alley, Project Manager

[5] The Urban Transportation Monitor Oct 15, 1993 p.5

[6] Seattle Metro promotional brochure (undated) -opened September 1985

[7] The Urban Transportation Monitor May 13, 1994 p.5

v. Eight systems sell fare media at transit parking facilities.[8] [A bigger issue at unmanned location is "which ones do not require exact cash?" -ed.]

vi. Chicago IL:
"KinderCare at Work" established child care centre at Lombard Metra Station, Chicago IL providing care for 70 children from 6 weeks to 12 years. [9] (No cost given).

vii. Los Angeles CA:
Day-care centres are scheduled to open early next year (1994) at two Metrolink commuter rail stations in suburban Los Angeles. Run by the Los Angeles County Metropolitan Transportation Authority, one facility will be 12,000 square feet of building and outdoor space, costing $779,000 US catering to 90 children up to age five; the other 10,000 square feet, costs $750,000 US and caters for 75 children. Outdoor space is larger than indoor.[10] (Costs are US$8,700 and $10,000 per child space respectively)

viii. Baltimore, MD:
(1995) Maryland Mass Transit Administration has received $1.5 mln US to build a 100-child care facility and a police substation at its Reistertown Road Plaza Metro Station. Total cost $1.9 mln [11] (US$19,000 per child space)

ix. Santa Clara County CA:
a) (1994) The Santa Clara County Transportation Agency is constructing a childcare center at the Tamien Multimodal Station in San Jose. The facility will occupy 28,000 square feet (9,600 building; 15,500 play yard) displace 145 parking stalls, accommodate 150 children (45 infants/toddlers, 89 pre-school, 16 school-age), and be proximate to the University education faculty as a teaching lab. Cost $2.3 mln US.[12] (US$15,350 per child space)

b) The Santa Clara County Transportation Agency has established a Joint Development Program to identify and implement joint development opportunities at Transit District owned park-and-ride lots. The goals are: revenue production, increased transit ridership, and environmental

---

[8] APTA 1994

[9] Press Release, The Bromley Group, Mimi Field, 212-807-0878 Undated

[10] Metro Magazine, May/June 1993 p.6

[11] Elana Mazile, "Baltimore Metro Station to Add Day Care, Police Substation" Passenger Transit, July 10, 1995

[12] Press Release, Santa Clara County Transportation Agency, August 1994.

enhancements around the transit facility. One outcome is the Almaden Lake Village, a high density, 250 unit multi residential 2 and 3 storey complex plus exercise facilities, fully secured main grounds, sub-grade parking. Completion March 1996. [13]

x. St. Paul MN:

The Minnesota Valley Transit Authority is expanding park-and-ride facilities and intermodal facilities at the Burnsville Park and Ride location. Convenience retail, office development, day-care facilities, expanded parking, movie theaters, seniors housing and a passenger service centre/civic information centre are planned. It will provide access to numerous modes of transit including express commuter services with direct HOV lane access, local circulator buses, handicap and senior citizen transit services, van-pool and car-pool facilities.[14]

Elsewhere in the US the following have been noted.

xi. Quantico VA

In Quantico VA, a suburb of Washington DC, a bakery has located close to fare paid zone for 40 years and, as reported in 1996, continues to be profitable. Its management provide service to both sets of rush hour traffic. The following quote was taken from the Internet:

About 200 feet east of the Amtrak/VRE station at Quantico VA ("Crossroads of the Marine Corps") is "Harry-O's" bakery. They [sic] have a full line of baked goods plus a light food menu (soup, sandwiches, chili, beverages, etc.). ... [many] VRE passengers stop in for coffee & donuts before boarding their northbound train in the morning. Now there is nothing unusual about that. However, I was surprised to learn that they take phone orders from VRE passengers who want to pickup something on their way home to one of the stations further down the line. He or one of his staff meet evening southbound VRE Trains when they stop at Quantico; passengers hit the platform, he hands them their baked goods, they hand him cash/check and they quickly re-board their southbound train. Seems like an unusual service provided by an enterprising businessman. I suppose it can be done without delaying the train as long [as there] are not too many customers picking up orders.[15]

---

[13]. Press Release, Santa Clara County Transportation Agency, August 1994.

[14] The Urban Transportation Monitor, March 31, 1993 p.6

[15] Jack Deasy
> Alexandria VA
> jhdeasy@tmn.com
> john_deasy@hq.dla.mil
Reposted from the railroad list:

Surely such services could readily be provided by either full-time commercial ventures proximate to P+R termini. Alternately, mobile sales facilities, not unlike the industrial coffee trucks, could gain regular clientele. One similar example occurs in Catterick Garrison, England, where Graham Wharton is a mobile purveyor of fresh sandwiches and bakery goods, with regular clientele placing orders either same day or day before or simply selecting off the truck as happens with the standard Canadian coffee-truck hawking its wares around industrial complexes. Mobile canteens have not yet perceived a P+R market.

In summary, the above provide a selection of initiatives that will increasingly appeal to the park and ride community, and should encourage its growth. A small survey of ten facilities noted below, does not provide as yet much re-inforcement for the traffic-growing effects of MPC's. Vending machines were reported at three locations and newsagents at two locations. A broader commercial spectrum appears unavailable at any of the ten.

7. SURVEY OF EFFECTIVENESS OF TRANSIT TRANSFER CENTRES (MPC's).

The ITE (Institute of Transportation Engineers) Technical Council Committee 6A-40[16] lists the following:

a. Objectives:

i.Operator's

(a)Rest area
(b)Image
(c)civic facility
(d)aid to downtown development and revitalization [no comments provided]
(e)improve ridership [yes:3/10 locations - 5/10 no impact]

ii.User's

(a)Weather protection [yes -9/10]
(b)reduce accidents - vehicle/rider interface
(c)passenger convenience

b. Problems: It identifies two major problems

i. non-transit users (youths/homeless) [This could be reduced significantly by keeping it in a fare-paid zone - ed.]

ii. vehicular circulation problems. [Planners: "To horse!" -ed.]

Date: Tue, 6 Feb 1996 06:17:46 GMT

[16] "Effectiveness of Downtown Transit Centers" ITE Journal, Sept 1990 pp.34-38

8. JOINT DEVELOPMENTS AT TRANSIT STATIONS

King Cushman[17] as long ago as 1988 surveyed some developments in transit-commercial ventures. He noted that in Pierce County, WN two of the four centres were located in regional shopping centres, each replacing 110-130 parking spaces. To secure co-operation it was demonstrated that of 8,000 transit riders, "...nearly 1,500 people stopped in to shop, the value of a permanent transit centre as a means to increase shopping center patronage while minimizing parking space development became clear to the land owner"[18].

In San Francisco, BART "...found private market interest that parallels the pace of economic development in general in its heavy rail corridors."[19] "In Pleasant Hill with several acres in use for a park-and-ride lot, it is exploring proposals...and may even consider an equity position."[20]
In Atlanta, MARTA in 1987 received lease income of US$700,000.[21] And in Washington WMATA by that date received US$14 million in capital contributions, and in the 1988 fiscal year received $3.6 million in rental income, even with no commercial enterprises allowed within their operating shell areas.[22]

9. EXPANSION CONSIDERATIONS for MPC's

From the above, it is clear that commercial-transit ventures can have a symbiotic relationship. One must keep in mind some basics, however.

a. economics

Any commercial venture expects to make at least normal profits. To encourage a commercial node to develop, it may be necessary to grant a tax holiday or even a subsidy. Some of the initiatives in the U.S. (day-care particularly) draw from non-transit budgets. If the presence of commercial establishments could be shown to diminish the need for other security measures, then such subsidies could also be netted out from the foregone expenditures. Markets take some time to grow, both sides should work to this end.

---

[17] King Cushman, , "Joint Development at Transit Stations" in Wayne Attoe (ed.) Transit Land Use and Urban Form, Austin TX, Center for the Study of American Architecture. 1988.

[18] ibid. p.182

[19] ibid. p.183

[20] ibid. p.184

[21] ibid. p.185

[22] ibid. p.186

b. Physical access

i. auto access for non-P+R users

The more non-transit traffic is attracted to MPC's, the more vehicular conflicts will arise, unless carefully planned. The direct bus access to the HOV lanes in St. Paul is an example of good planning.

ii. foot/cycle access

Environmentally preferable modes of travel need also to be integrated in MPC's. Again, planners need to recognize this.

c. socio-political considerations

MPC's must become integrated with the social fabric of the city. Cities, in an effort to avoid ghettoization are increasingly trying to mix society's economic strata. The MPC does not depend on auto ownership/operation - a symbol of wealth - and hence can be of socio/political use.

10. THE FUTURE

Several initiatives and suggestions have been noted above. If MPC's have the beneficial impacts suggested then their development should be encouraged by land-use planners and economists. The above demonstration projects should be eagerly watched for actual and potential benefits and costs.

11. BIBLIOGRAPHY

Cape, David M. "Bus-based park and ride: a tool to help the urban core economy", CTRF Proceedings, 1995

Cushman, King, "Joint Development at Transit Stations" in Wayne Attoe (ed.) Transit Land Use and Urban Form, Austin TX, Center for the Study of American Architecture. 1988.

International Transportation Engineers Technical Council Committee, "Report 6A-40, Survey of Effectiveness of Transit Transfer Centers",ITE Journal, September 1990

Mazile, Elana, "Baltimore Metro Station to Add Day Care", Passenger Transit, July 10, 1995

Metro Magazine, May/June 1993

Santa Clara County Transportation Agency, Press Release August 1994

Seattle Metro, Promotional Brochure, September 1985

The Bromley Group, Press Release, Undated

The Urban Transportation Monitor, select issues 1993, 1994

Application of Diesel Rail Car Technology

Darrell Richards, Transport Concepts

Summary

Commuter rail in small cities or low volume markets has been difficult to set up because of the cost of providing service with conventional train technology. Self-propelled rail diesel cars are less expensive to operate than locomotive hauled trains in low volume markets, but no new rail cars have been produced in North America for a number of years. Several manufacturers are developing or have developed specifications for new self-propelled cars that comply with North American crash and accessibility standards.

Diesel light rail vehicles have been developed in Germany which operate on conventional intercity rail tracks. Vehicle maintenance costs are reduced by using standard bus and truck parts. The vehicle accelerates and decelerates faster than a conventional rail car and has performance similar to a bus. Station time is reduced by low floor boarding. A diesel light rail vehicle will be tested for five months on a CP freight track in Calgary in 1996.

Because it can operate on existing rail tracks and avoids the cost of electrification, diesel light rail can reduce the start up costs of a light rail system. But according to current North American safety standards, operation of this light vehicle must be separated by time of day from heavy conventional trains. Thus, use of diesel light rail technology in the short-term is constrained where there is frequent use of lines by conventional trains.

A case study analysis of potential application in Ottawa finds that diesel light rail is compatible with the CP freight line which has only two freight trains, one in each direction, daily. Compared to a conventional train, diesel light rail on this route would reduce equipment and crew requirements in half due to its superior performance.

Introduction

I would like to acknowledge information and ideas provided by Brian Sullivan, Gilles Hébert, Oliver Bowan, Tim Lane, John Schnoblegger, Barbara Ramsay, John Baaker, Rainer Ibowski, Henri-Louis Batty, Frank Guzzo, Dhruv Sharma, Donald Hunt, Jerome Pier, Bill Middleton, Clifford A. Woodbury, and Thomas Holmqvist. Any omissions or errors however are mine.

Local and provincial governments are finding it increasingly difficult to finance new urban transportation infrastructure. Developing new infrastructure in urban areas is often in conflict with neighbourhood, land-use, and environmental concerns. And it can be difficult to find available corridors for new highways or rapid transit lines. In some urban centres existing railway tracks could be used, but the cost of operating conventional commuter train equipment can only be justified where there are sufficient volumes of passengers. Thus, it has been difficult to use rail infrastructure in small and medium sized cities for urban passenger transportation, that is commuter rail service. The challenge then is to exploit the existing rail infrastructure by developing less expensive ways to operate commuter rail transit in small and medium volume markets.

The purpose of this paper is to examine new railway technologies and operating methods that could be used on low and medium volume routes. It compares the characteristics and suitability of locomotive hauled trains, diesel multiple units, and diesel light rail vehicles for different applications and markets. An example of a locomotive hauled commuter train is Toronto's GO Transit. Diesel multiple units are self-propelled commuter coaches meeting railway crash standards that can operate singly or in multiple. Diesel light rail vehicles are bus-like rail vehicles that weigh about one-half as much as conventional rail cars and do no not meet conventional rail crash standards.

History of Diesel Rail Cars

Early in this century some railroads developed self-propelled "doodlebugs" or "motor cars" to provide branch line service or in some cases commuter service. The most popular diesel multiple unit (DMU) rail car was introduced by the BUDD company in 1949. The power train had two diesel engines of 210-230 kilowatts each. Cars could be run singly or in multiples. By the end of production in 1962, more than 2,000 units were sold worldwide with the majority in North America.[1] These BUDD Rail Diesel Cars (RDCs) are still being used in North America by B.C. Rail, VIA, Cape May Seashore Lines, Delaware and Otsego, and the Québec North Shore and Labrador. Dallas Area Rapid Transit will open a 17 kilometre commuter rail line in 1997 using 13 RDCs purchased from VIA Rail Canada and rebuilt by AMF Industries in Montréal. (VIA Rail Canada still has some used BUDD cars for sale through Canac International Inc.)

BUDD tried to update the RDC in 1978 with introduction of the SVP 2000 which was neither a marketing nor operational success. Meanwhile, development and expansion of the DMU fleet overseas continued in England, Israel, Australia, Norway, Germany, France, Finland, and Sweden.

New Developments in DMUs

Based on the success of diesel multiple units overseas, and the renewed interest in commuter rail, manufacturers are marketing DMUs for the North American market. These new vehicles will comply with North American crash standards and the Americans with Disability Act accessibility standards.

Bombardier is marketing a DMU based on the car shell of its electric multiple unit used on the Deux Montagnes commuter line in Montréal. Compared to the older BUDD car, this vehicle has higher acceleration and horsepower. Siemens Transportation is modifying its model 628/928 from Germany for the North American market. Nippon Sharyo is marketing a diesel rail car based on its rail car design used in Maryland. ABB will be demonstrating an IC-3 flexliner on Amtrak inter-city routes in May 1996. This equipment, or its derivative the IC-2, would also be suitable for commuter operations. GEC Alsthom markets its TER self-propelled train, of which the French railway has purchased 100.

Diesel Light Rail Cars

During the post World War II period, Europe's electric streetcar systems were modernized and upgraded into light rail transit systems. This technology is capable of operation on tracks in city streets or on exclusive transit lines. A number of North American cities introduced this technology: Calgary, Edmonton, Sacramento, Portland (ORE), San Jose, San Diego, St. Louis, Denver, San Francisco, Dallas, Los Angeles, Long Beach, Baltimore, and Buffalo (although this was built more like a heavy rail métro). Older streetcar or light rail systems in Toronto, Cleveland, Boston, and Pittsburgh have been partly or completely modernized. The cost of new trackage and electrification is sometimes a barrier to the development of new light rail systems.

The requirement to build new tracks for light rail systems is being avoided in Germany which has begun to integrate light rail and conventional rail systems. For the past 10 years in Karlsrhue Germany electric light rail vehicles have operated both on city streets and on the regular inter-city rail tracks even though the light vehicles do not meet crash protection standards for conventional rail cars.[2][3][4] To compensate for lower crash standards, the light vehicles have shorter braking distances and run under automated train protection.

Light rail cars will also be introduced to conventional rail infrastructure in Saarbrüken, Germany in 1997. Trams (light rail streetcar technology) in Saarbrüken were abandoned in 1965, but will be re-introduced and extended beyond the city on the German railway. Sharing of tracks by light and heavy trains is also under consideration in Nottingham, England.[5] In North America, electric light rail trains operate on freight railroad tracks in San Diego, California. Unlike the German example, however, light and heavy trains in San Diego are separated by time of day. The freight operates at night; light vehicles during the day.

Germany has also developed a diesel light rail vehicle for use on existing railway tracks. Maintenance costs are reduced by using off-the-shelf bus, streetcar, and truck parts. It accelerates and decelerates faster than a heavy rail car and has performance similar to a bus (accelerates 0-80 kph in 38 seconds and 520 metres with seated load). The vehicle's low floor permits level boarding. Station time is reduced because passengers do not have to climb or descend stairs. Bicycle racks are provided on board, but can be folded out of the way when not in use. The total weight is about 30 tonnes compared to 55 tonnes for a conventional diesel rail car of similar seating capacity. Lower fuel, maintenance, and operating costs reduce the life cycle cost about 40% compared to a conventional rail car.[6]

Siemens Transportation has produced 16 "RegioSprinter" diesel light rail trains for use on conventional railway infrastructure between Aachen and Köln Germany. Calgary Transit will demonstrate use a RegioSprinter on a 6 kilometre stretch of CP freight track for five months beginning April 1996. The Calgary train will connect a suburban area to the Anderson light rail station where passengers will transfer to the existing electric light rail system. The City of Folsom, California is negotiating with Siemens to purchase five RegioSprinters which would be used on a 16 kilometre freight track near Sacramento as a feeder service to the Sacramento light rail system. Denmark's Lyngby-Naerum Jernbane has ordered a RegioSprinter and has an option for 50 more.[7] GEC Alsthom markets a diesel light rail vehicle, the LINT light motor car developed by Linke-Hosmann-Busch GmbH. Its performance and characteristics are very similar to our description of the RegioSprinter. Bombardier is producing 25 diesel light rail vehicles in Europe.

Both the RegioSprinter and LINT diesel light rail cars are currently unable to run on city streets. The width of the carbody (3 metres) is more than the 2.65 maximum acceptable for street running. Also, the minimum turn radius of 80-90 metres would prevent the car from making tight turns at city street intersections. Reducing the carbody to 2.65 metres and using a tight turn radius would then permit a truly diesel light rail vehicle that can operate both on railway tracks and on tracks in city streets.

Analysis and Comparison

There is no such thing as a good or bad mode of transport and no one mode can do everything. The bus, train, bicycle, pedestrian, and car have different advantages and each have a role to play in urban transport. Similarly, within the rail mode different technologies have different advantages in different applications. Locomotive hauled commuter trains are the most cost-effective option where there are large volumes of passengers and stations are spaced four kilometres or more apart. Diesel multiple units are suited in smaller markets or on secondary lines in major cities where commuter trains must share tracks with freight and conventional passenger trains. They could also be used to reduce the cost of off-peak service on main routes. Diesel light rail cars are advantageous on low volume routes where there are light volumes of freight traffic that can be scheduled to avoid conflicts with light rail cars.

The West Coast Express in Vancouver recently selected locomotive hauled trains for its commuter rail service which opened in 1995. This technology was appropriate because of the long distances and large volumes of passengers per train movement. Raleigh, North Carolina is selecting diesel multiple units for its future commuter service because commuter trains must share tracks with heavy trains and there are insufficient volumes for locomotive hauled trains. Folsom, California is selecting diesel light rail technology because it has exclusive use of tracks during commuter hours.

Case Study

CP Rail System has proposed a commuter train service on a 30 kilometre section of underused freight track linking Gatineau and Hull Québec to Carleton University and the Confederation Heights employment centre. Dedicated shuttle buses would transfer passengers to downtown Ottawa and Hull. The service plan for the 10 station system called for four conventional train sets plus a spare, a 40 minute transit time, and five peak period trains each direction for a total 20 trains per day. (A study prepared for la Societé de Transport de l'Outaouais recommends a busway along the Québec portion of the corridor instead of a conventional commuter train.)

Diesel light rail is a feasible alternative on CP's freight track in the national capital because freight traffic does not operate during commuter hours. The superior performance of a diesel light rail car (faster acceleration/deceleration and shorter boarding time) would reduce the transit time to 30 minutes. If the end station in Gatineau were deferred, and if Walkley Station were substituted for Hunt Club, the light rail train would offer a transit time of about 25 minutes. This would reduce the required number of active train sets from four to two. The number of crews would also be cut in half. Material and labour requirements would be reduced by half, but ridership would decline by only 10%.[8]

An alternative service option for a demonstration project or starter system would be to use a 7 kilometre route for a diesel light rail service linking two existing transitways to Carleton University. The light rail shuttle could link the east-west transitway near LeBreton flats to City Centre, Carling Avenue, Carleton University, Confederation Heights, and the southeast transitway. Diesel light rail vehicles are suited to this line because of the short average station spacing (1.5 kilometres).

Description of Corridor Between LeBreton Flats and Walkley

The section of line in question is mostly grade separated, except for a level crossing of a recreational path near Heron Road that is controlled by flashing lights and a level pedestrian crossing at Carleton University. From LeBreton Flats to Dow's Lake it runs through a 7-8 metre wide rock cut below cross streets. It goes under Dow's Lake via a tunnel 580 metres long, passes Carleton University at grade, crosses the Rideau River on a bridge, runs to Heron Road in a slight depression, and runs to Walkley road and beyond at grade.

Service Plan

There are three options to connect the diesel light rail line to the east-west main bus transitway at LeBreton Flats:

a) Build a new bus transitway station at the crossing of the rail line. A bus transitway at this location is already in the region's long term plan, but may be delayed until development of City Centre proceeds. This would permit transfers, with a change of elevation, from busway to train. It would require transitway buses to make an additional stop.

b) Provide shuttle buses from the rail line to Tunney's Pasture and downtown Ottawa.

c) Build a one kilometre rail line from the City Centre area to the existing Booth St. bus transitway station. This would permit level transfers to all buses on the transitway in both directions.

A tentative service plan for an initial starter system from Booth St. to Walkley Road would be 11 minutes one-way transit time and 13-15 minute headways. This would require two light trains plus a spare. The Booth St. station would permit transfers to and from buses on the spine east-west transitway which serves downtown Ottawa as well as middle and outer suburbs east and west. The City Centre Station would serve an employment centre for LeBreton Flats designated in municipal plans but not yet built (There is however already one office building at City Centre). The Carling Avenue station would serve the Dow's Lake tourism area, Natural Resources Canada, Agriculture Canada, and the Civic Hospital (this hospital which employees 4,500 people already has a shuttle bus from an overflow parking lot next to the proposed rail station).

The Carleton Station would serve the University campus which has 17,000 full-time students, 5,000 part-term students, and 3,600 workers. People in the student age bracket of 18-24 are almost twice as likely to use transit as the average population.[9] The Confederation Heights station would serve 7,500 employees. There are 140,000 residents and 165,000 jobs within three kilometres of this section of line.

The connection to the southeast transitway could be:

a) at the existing Walkley transitway station which is at the same level and beside the rail line

b) by building a light rail bridge across Heron Road to beside the existing Heron bus transitway station. Later on the light rail line could be extended to Billings Bridge, possibly on the transitway.

Start Up Capital Costs

There are only two daily freight trains on this line. Thus, it is feasible to use existing rail capacity for the light rail service, and this substantially reduces the capital costs compared to new infrastructure. Start up costs would include a ballast program, relaying rail, and replacing one-third of the existing wooden ties. New track of 1 kilometre could be built from City Centre to Booth St (further investigation would determine if there is any contaminated soil to be removed along the trajectory). This new construction could be deferred if feeder buses were provided from City Centre to downtown and Tunney's Pasture. Because the route is single track, a passing track would be required near the mid-point at Carleton University

The new track to Booth St. could also be avoided if the planned transitway station at City Centre were built. A City Centre transitway station would provide a better train operating plan than a Booth St. transfer station. The one-way train run time would be reduced about two minutes, thus permitting 11-12 minute frequencies rather than 13-15 minutes. But high cost of the City Centre transitway station (about $6 million) may delay construction until development of LeBreton flats proceeds. It is assumed here that the 1 kilometre track extension to Booth St. on relatively level ground would be less expensive that building the City Centre transitway station.

No automobile parking is required at any of the six stations because this is an inner city shuttle system. A completely new station at Walkley or Heron is not required, since the existing transitway stations and elevators could be connected to the rail line by a short pathway. To contain costs for a start-up operation, stations should comprise basic platforms, pathways, bicycle parking, shelter, passenger drop off area, lighting and emergency phone. Since the rail line is depressed at Confederation Heights, the station there would also require a sloping ramp to provide access for wheelchairs. The Carling Station would be the most expensive as it requires widening the rock cut and installing of an elevator and staircase. The cost of the additional work compared to a standard station has been estimated at $1.6 million.

Operating and Equipment Costs

An initial service plan of 13-15 minute headways and about 11 minute one-way transit time would require two train sets plus a spare. The incremental cost of providing off-peak service is very low: fuel and maintenance. Because of the method of assigning railway crews, off-peak service can be run without additional labour. The marginal costs and revenues of off-peak service should be analyzed.

On this route the light rail vehicle would achieve a transit time 30% faster than conventional locomotive train technology due to fast acceleration/deceleration and low floor boarding. To operate 15 minute headways would require 4 conventional trains, but only 2 light rail trains. In addition to these savings in materials and crews, the light rail train would have lower fuel and maintenance costs. The light rail cars cost about $2.1 million $Canadian each, or about $200,000 annually each for a lease. On a per seat per year of life basis, the cost is similar to that of a new bus.

Ridership

An initial ridership forecast can be made by isolating the Ontario origin destinations from the inter-provincial trip matrix prepared for the Interprovincial Commuter Rail Appraisal Study. The base ridership for Ontario trips for a 15 minute headway was 1800-2050 daily. This ridership forecast was a preliminary examination at the overview level of the national capital region. The forecast could be reviewed, targeted, and updated for the service plan proposed here:

a) The previous study did not provide full transfer integration at the transitway stations because feeder buses only went east to downtown and west to Tunney's Pasture. Thus, the ridership matrix selected by the study projected 0 morning trips from Orleans, Cumberland, Ottawa east of the downtown Rideau Canal, Ottawa west of Island Park Drive, Nepean, or Kanata to Carling, Carleton, or Confederation Heights. The Booth St and Walkley (or Heron) Road stations proposed here would provide full connections from transitways to the activity centres along the CP corridor.

b) The previous study may have under-estimated student trip travel to Carleton because at the overview level it did not examine student travel in much detail. In the a.m. peak period it projected 60 employees (out of 3642 total employment) and 46 students (out of 17,000 full and 5,000 part time students) would arrive at Carleton University by train. The imputed trip rate of students appears to be one-sixth that of employees. Yet other surveys find that the student age population is almost twice as likely to use transit as the average population.

c) substituting Walkley station for Hunt Club station. Walkley has less residential population nearby, but provides a better transitway connection east to Billings Bridge and Alta Vista.

d) attractiveness of 30% faster transit time compared to conventional train schedule

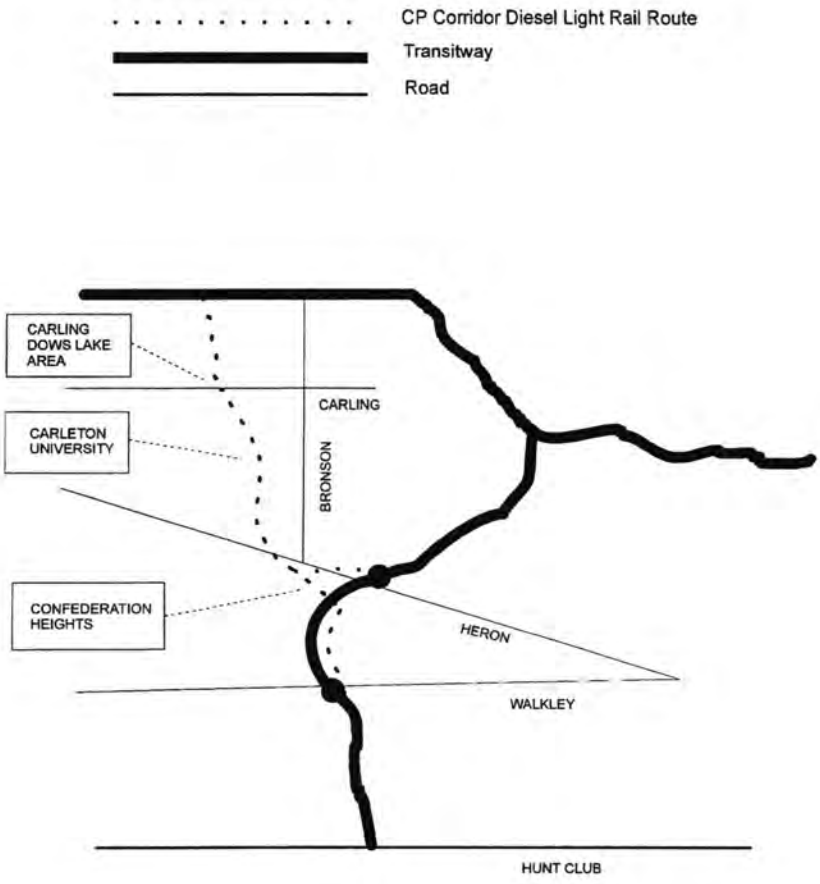
e) potential tourism travel and development at the Dow's Lake area and Experimental Farm.

Potential Staging of Development

A light rail service along the CP corridor linking the two transitways to Carleton University could be implemented in stages. An example of staging is provided:

Stage One	3-5 Month Demonstration with a single light rail car. Temporary Platforms at City Centre, Carleton University, and possibly Confederation Heights and Walkley. Shuttle bus from City Centre to downtown Ottawa (and possibly Tunney's Pasture) via temporary access near the rail line. No Carling Station. Temporary vehicle storage and maintenance at Walkley Yard shops.
Stage Two	Tie, rail and ballasting program. Build 1 kilometre new track from City Centre to Booth St. Build Carling Station. Build more permanent station facilities. Build passing track near Carleton University. Purchase or lease equipment sufficient to provide two trains of adequate capacity plus spare requirements. Upgrade Walkley shop or construct a new one. Heron Road station is an alternative to Walkley, but would require a $2-3 million light rail bridge over Heron Road.
Stage Three	The line could be extended south and north on CP's existing track. Development of a 2.65 metre wide light rail car and tight turn radius would permit extension of trackage to downtown Ottawa. This creates an interesting opportunity to link Carleton University to the University of Ottawa.
Long-Term	Electrification and double tracking if required.

852


CP Corridor Diesel Light Rail Route
Transitway
Road
CARLING DOWS LAKE AREA
CARLETON UNIVERSITY
CONFEDERATION HEIGHTS
CARLING
BRONSON
HERON
WALKLEY
HUNT CLUB

Endnotes

[1]Jerome R. Pier, Allied International Corporation, A Short History of the Rail Diesel Car in North America, presented to the CONEG Policy Research Centre's DMU Workshop in Springfield, Massachusetts October 1995
[2]W.J. Wyse, "Light Rail through the valleys; Karlsruhe's Product of the Year," Light Rail and Modern Tramway, November 1994
[3]Dip-Ing Georg Drechsler, "Karlsruhe trams cross more city boundaries," Railway Gazette International, November 1994
[4]"Dual-mode light rail offers competition to the car," Railway Gazette International
[5]David Jenkins, "Track-sharing in Britain: Is Karlsruhe's work valid here?" Light Rail and Modern Tramway, July 1994
[6]Peter Jacques, "Lightweight railcar bids to rescue regional routes," Railway Gazette International
[7]"Folsom backs commuter rail," Railway Gazette International
[8]This does not account for the impact of 25% faster transit time which would tend to increase ridership.
[9]OC Transpo Attitude Survey 1994 reported that 18-24 year olds were 27% of frequent transit riders, but made up only 17% of the population.

Urban Transit Ridership in Canada: Nowhere to Go But UP!

by Mebs S. Kanji

Introduction

In advanced industrial countries, most modal choice analysis typically centers on two basic alternatives: private versus public transportation.[1] Much of the recent empirical evidence from both Western Europe as well as the United States clearly suggests that publics increasingly prefer the former over the latter: both private car ownership and usage continue to rise, while urban transit ridership steadily declines (OECD, 1995; Pucher and Kurth, 1995; Pucher, 1995; Turton, 1992).[2] In addition, many transportation analysts also claim that expanding sub-urbanization (or sub-urban "sprawl") coupled with increasing urban transit fares, creates a "vicious downward spiral" — one that threatens to virtually eliminate public transportation over time (Turton and Knowles, 1992; Webster et al., 1985). So far, however, much of the preliminary Canadian evidence has been somewhat inconsistent. On the one hand, a straightforward comparison of basic modal split figures presents a relatively optimistic picture: the average Canadian uses urban transit much more frequently than his/her American counterparts, and only slightly less than some leading Western Europeans (Newman and Kenworthy, 1989). But on the other hand, recent crosstime evidence suggests that Canadians are not unlike publics in other advanced industrial countries; that over time, Canadians too have become less supportive of public transportation (Perl and Pucher, 1995; Gratwick, 1993). Inspired mostly by the latter findings, policy analysts have been quick to table certain fiscal and regulatory initiatives which, they argue, must be implemented in order to regain lost urban transit ridership. The problem, however, is that such measures often rely, to a large extent, on increased government subsidies. Furthermore, very few recommendations even consider the important catalytic role that both rising public concern for the environment, and increased environmental activism (Nevitte and Kanji, 1995; Kanji and Nevitte, 1995) may play in stimulating additional urban transit ridership.

The purpose of this paper is to examine urban transit ridership in Canada over the last two decades (1975 to 1993). This investigation differs from previous explorations, however, in many important ways. For example, most preliminary empirical examinations of this sort rely heavily on crosstime comparisons of unstandardized ridership statistics — evidence which basically ignores any simultaneous expansion or contraction in urban transit services. Part I of this analysis accounts for such fluctuations in service levels, and evaluates urban transit ridership on a per kilometer basis. Another important difference between this analysis and those preceding it has to do with the basic approach used to explain the crosstime variation in urban transit usage. Most preliminary examinations which investigate the predictors of Canadian urban transit patronage remain, for the most part, highly anecdotal; very few studies actually mount any direct empirical

[1]Acknowledgements: I would like to thank both Anthony Perl and Stan Drabeck for their comments on earlier versions of this paper. I would also like to express my gratitude to Judi Powell for her assistance in compiling and editing this manuscript. None of the above are responsible for the interpretations contained in the paper.

[2]Some possible exceptions include West Germany, Switzerland, Austria, the Netherlands, and France. These countries have recently expanded and improved the quality of urban transit services. As a result, they have been increasingly successful at reversing the recent broad-gauged decline in public transport patronage (Pucher and Kurth, 1995; Pucher, 1995).

explorations, and as such, most prospective explanations for the so-called decline in urban transit ridership have yet to be significantly linked to urban transit usage. This analysis attempts to move closer toward filling that "gap" by directly examining two plausible explanations. The first explanation concerns the economics (or affordability) of urban transit. Many transportation analysts propose that in order to regain lost urban transit ridership, increased government subsidies are needed to make transit more affordable. Part II of this paper examines that proposition by asking: Have increased government subsidies been effective at stimulating additional urban transit usage in Canada? Following that, the second candidate explanation considers the Canadian public's changing environmental orientations, and how they may affect levels of urban transit patronage. More specifically, Part III of this analysis explores how effective both increasing public concern for the environment, as well as increased public willingness to actively pursue the environmental cause are at motivating additional urban transit usage.

Data

The data for this investigation come from two basic sources. Parts I and II examine aggregate level data which are revised and reported annually by Statistics Canada.[3] These data include figures for urban transit usage and total transit services (in kilometers). In addition, there is also detailed information on transit's total operating revenues (including subsidies), as well as transit's total operating costs. Next, Part III of this investigation turns to evaluate data from the 1990 World Values Surveys (WVS) — a large and unique body of individual level data collected in over forty different countries, one of which is Canada. The WVS ask publics about their general concern for the environment, as well as their willingness to actively pursue the environmental cause. Moreover, the dataset also includes indicators of several prominent explanations for increasing environmental concern and activism (i.e. rising levels of public support for the environmental movement and its basic goal of "environmental protection;" the rise of postmaterialist value orientations; rising levels of cognitive mobilization; and various background factors), all of which are required for the latter part of this analysis.

Literature Review: Automobiles, sub-urbanization, and transit ridership

Many recent crosstime studies suggest that urban transit ridership across multiple advanced industrial countries is in decline (OECD, 1995; Pucher and Kurth, 1995; Pucher, 1995; Turton, 1992). In most instances, the drop in urban transit patronage is linked to two growing trends: (1) rising rates of automobile ownership and usage; and (2) growing sub-urbanization. For instance, Turton (1992) argues that in both Europe and North America, "the car is now the leading mode for all categories of journey and with the growing trend towards decentralization (or sub-urbanization) of many urban activities which have traditionally been located at the city core, private transportation is becoming of even greater significance" (1992:74).[4] Similarly, Pucher (1995) claims that both increased car ownership and sub-urbanization place public transport at a competitive disadvantage which gets worse with each year (p.222). In fact "no mode can really compete," he argues, "with the private car in low-density suburbs" (p.219).[5]

But what accounts for the recent expansion in both private transportation, as well as, low-density sub-urbanization? According to a recent report published by the Organization of Economic Cooperation and Development — *Reconciling Transportation,*

---

[3]The specific report referred to here is: Passenger Bus and Urban Transit (1975 through to 1993).

[4]In brackets is mine.

[5]It is important to note here, that according to a recent study conducted by Charles Lave (1991), private vehicle use is expected to continue growing, while ownership is likely to taper off. The reason, Lave argues, is simple: the market for private vehicles is essentially saturated. "This means that two thirds of the previously expected increase in vehicle-caused problems will go away, all by itself, without any intervention..." (p.229).

Energy and Environmental Issues: The Role of Public Transport (1995) — the increasing public demand for cars may be directly linked to economic growth. More precisely, the report argues that "increased prosperity leads to the purchase of more cars and larger cars, and to greater use of cars" (p.16). Furthermore, the report also finds that while the first family car is generally acquired for group-related leisure activities, "one member of the household appropriates it for travelling to work, stimulating the purchase of a second or even a third car" (p.16). In a similar vein, Pucher (1995) contends that rising private car ownership and growing sub-urbanization are the direct result of both economic as well as technological advancements: "Increasing per capita incomes have made private car ownership and sub-urban home-ownership more affordable," while advances in communication and transportation have greatly increased the locational flexibility of firms (p.105).

Also important to note here is the tremendous flexibility and personal convenience offered by the private car — both of which allow the user to execute daily travel patterns with minimal constraints (Turton, 1992). Because the low density of sub-urban community development makes the provision of adequate public transport difficult (OECD 1995:16), most of the preliminary evidence indicates that there is a strong inverse relationship between the density of residential development and private transportation.[6] The real advantages associated with private car usage are unlikely to be matched by any other form of transport — except in congested urban areas; "the car has a unique flexibility...it is always available, it offers door to door transport, it has a high level of privacy and comfort, and it forms an extension to the home" (Banister and Pickup, 1989:8). Moreover, the private car also offers time savings, and satisfies the need for transportation at irregular hours (OECD, 1995).

In Canada, however, much of the preliminary evidence regarding the effects of increased private car ownership and usage, as well as growing sub-urbanization on urban transit patronage are far from being clear-cut. According to basic cross-sectional modal split figures from the early 1980s, Canadian urban transit usage is much higher than in the United States, and only slightly lower than in some leading Western European cities (Newman and Kenworthy, 1989). For example, an average Torontonian makes over 170 more transit trips than the average American and about 121 less trips than the average Western European (p.36). On the other hand, the most recent crosstime evidence suggests that levels of urban transportation patronage in Canada have declined, and as a result, the urban transit system may soon be headed for serious trouble. Perl and Pucher (1995), for instance, argue that during the 1990s alone, urban transit ridership in Canada has fallen by slightly more than 11%. Furthermore, "the Canadian pattern," according to Perl and Pucher, "...suggests the possible emergence of a secular decline in which demand continues to erode despite a modest upturn in the overall economy" (p.269). Consequently, Perl and Pucher warn that "transit managers and public officials who assume that today's downturn is cyclical, and who wait for a revival of demand from their traditional clientele, could wind up presiding over Canadian transit's decline into irrelevance" (p.269).

But consider more carefully, the data employed in Perl and Pucher's analysis. Although the authors attempt to account only for "linked" passenger trips, as opposed to "unlinked" trips,[7] their entire evaluation relies on a crosstime comparison of unstandardized transit ridership figures; statistics that may fluctuate across multiple timepoints for various different reasons, thereby making them inherently difficult to compare. As a result, basing conclusions solely on a crosstime evaluation of these sorts of data can be extremely risky. Furthermore, note too, that according to Perl and

[6]For a more detailed examination of this relationship, also see Pushkarev and Zupan (1977).

[7]See Perl and Pucher (1995) for a detailed explanation of "linked" as opposed to "unlinked" transit trips.

Pucher's analysis, urban transit ridership in 1990 (1.5 billion riders) actually surpasses the previous high of 1.4 billion riders achieved in 1950 — Perl and Pucher's original baseline point of comparison — by 7%. When compared to the 1950s, then, Canadian urban transit ridership in the early 1990s has only fallen by little more than 4% — a shift that is much less alarming than the proposed 11%. In fact, when placed in this long-term context, it is even conceivable that the 4% decline in urban transit ridership may simply be indicative of a temporary period effect.

Part I: A standardized crosstime evaluation of Canadian urban transit ridership

A more dependable approach for examining the state of urban transit patronage over time is to compare standardized ridership figures. Table 1 reports both urban transit ridership and total transit services (in kilometers) in Canada over the last two decades. To begin with, notice the extreme variation in the basic unstandardized data; between 1975 and 1990, both ridership and transit services expanded significantly — the former by nearly 36% and the latter by 56%. By comparison, however, both ridership and transit services in the '90s appear relatively stable, although admittedly, ridership does appear, at least at face value, to be steadily declining. But now consider the standardized figures (urban transit ridership on a per kilometer basis); they suggest a remarkably continuous and stable pattern: since 1975, urban transit has consistently carried an average of two passengers per each kilometer of service. According to these data, then, urban transit ridership in the 1990s, remains virtually identical to ridership in the 1970s. In fact, the sheer consistency of the above results, identifies another, more pressing, problem that currently faces urban transit services in Canada. The problem is one of efficiency: despite a near doubling of urban transit services over the last two decades, ridership, on average, continues to remain strikingly stagnant. The critical question, then, is not how to regain **lost** urban transit ridership, but rather how can Canada's urban transit system attract **"new"** ridership?

Table 1
Urban transit ridership, total transit services, and urban transit ridership per kilometer of service (1975 - 1993)

Year	Urban transit ridership (total fare passengers carried)	Total transit services (in kilometers)	Urban transit ridership per kilometer of service (nearest whole person)
1975	1,118,032,050	493,362,255	2.0
1980	1,304,309,000	653,276,772	2.0
1985	1,447,528,000	725,250,533	2.0
1990	1,528,400,000	769,326,000	2.0
1991	1,450,057,000	780,825,000	2.0
1992	1,432,105,000	754,399,000	2.0
1993	1,396,451,000	756,634,000	2.0

Source: *Passenger Bus and Urban Transit* (Statistics Canada 1975 to 1993).

Part II: Transit fares, government subsidies, and urban transit ridership

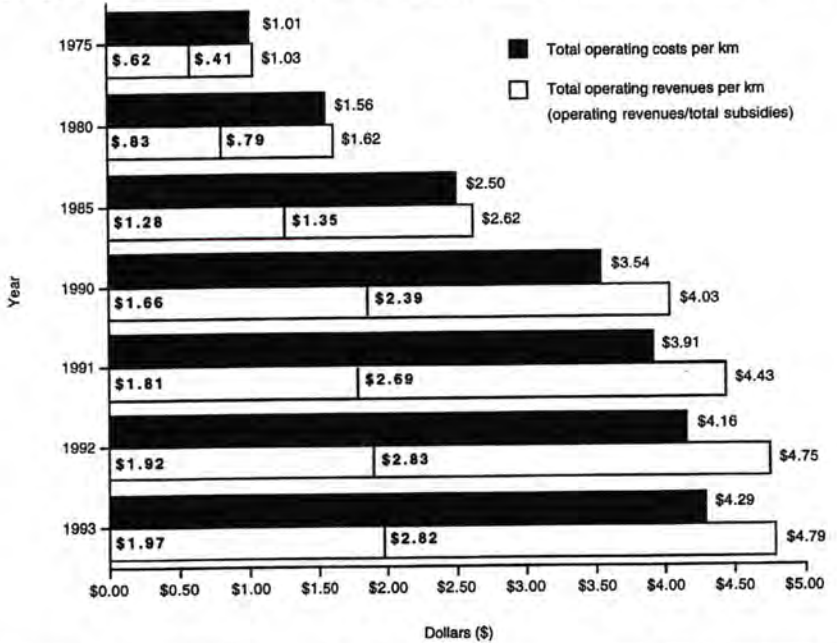
There is no doubt that both private and public modes of transport have their respective advantages. But because most publics typically prefer the convenience and flexibility of private modes of transport, public systems of transport tend to develop large scale inefficiencies and (in some instances, may even) experience declining levels of patronage (Jones et al., 1983; Sheppard, 1986). In such cases, the general consensus among many transportation analysts is that appropriate public policy initiatives are required to lure publics away from using their private cars and on to urban transit instead (Pucher, 1995; Pucher and Kurth, 1995; Perl and Pucher, 1995). To date, the most popular policy measure has consistently been to rely on increased government subsidies. Take the problem of rising transit fares, for instance; many policy analysts believe that increased government subsidies have the direct effect of making transit more affordable and thereby increase public transit ridership. Pucher (1995), for example, argues that "...the sharp decline in public transport use in Britain is partly the result of subsidy reductions and fare increases," whereas in other European countries like West Germany, Switzerland, Austria, the Netherlands, and France, massive increases in government subsidies have helped expand and improve public transport systems, as well as reduce fares (p.103; Pucher and Kurth, 1995). In Canada, Perl and Pucher (1995) claim that increasing fiscal austerity has made urban transit all the more dependent on both provincial and municipal government subsidies, and that this dependency, in turn, has enabled urban transit systems to become more effective in terms of offering more transit service.[8]

But just how effective are increased government subsidies at generating additional urban transit usage in Canada? Figure 1 breaks down the costs and revenues (including subsidies) of Canada's urban transit system on a per kilometer basis. Notice that in 1975, the total operating costs amounted to slightly more than $1.00 per kilometer. Yet less than just two decades later, those same costs increased by more than four times (1993: $4.29). Undoubtedly, at least a portion of the jump in operating costs likely reflects both the recent expansion of transit services (see Table 1), as well as basic inflation. But, there is also good reason to suspect that Canada's urban transit operation has become increasingly inefficient. Recall from Table 1, that the bulk of the recent expansion in transit services was achieved by 1985: at that point, the total kilometers covered by urban transit exceeded the 700 million mark, and operating costs per kilometer had already more than doubled to $2.50. Throughout the early '90s, transit services have remained basically consistent, while costs, on the other hand, have nearly doubled yet again. Moreover, the most recent increase in transit's total operating costs has not been sufficiently matched by a corresponding increase in operating revenues. In fact, over the last two decades, transit's operating revenues have steadily declined as a proportion of total operating costs: in 1975, operating revenues per kilometer covered 60% of the total costs, whereas by 1993 those same revenues accounted for only 40% of the total costs. In turn, the increasing inefficiency of urban transit has meant that a growing proportion of the total operating revenues (+20%) have been funded by government subsidies. In 1975, for instance, government subsidies amounted to .41 cents per kilometer, whereas by 1993, those same subsidies have only recently been "capped" off at close to $3.00 per kilometer.

---

[8]They also argue, however, that urban transit has become very inefficient when it comes to supplying services (also see Table 1). In an attempt to curb this inefficiency, governments have recently implemented a funding "cap" on Canadian transit subsidies. But according to Perl and Pucher, the recent "cap" has been at the root of declining transit ridership: the "...sudden shift in fiscal policy toward limiting subsidies at the same time that the automobile was becoming more price competitive goes far in explaining the resulting drop in transit usage" (1995:272).

Figure 1
Urban transit's total operating costs and total operating revenues (including subsidies)



Note: Total operating costs are comprised of: (1) transportation operation costs; (2) revenue vehicle maintenance and servicing costs; (3) non-revenue vehicle maintenance and other maintenance costs; and (4) general and administrative costs. The above figure reports the total operating costs per kilometer. Total operating revenues are comprised of: (1) operating revenues (including auxiliary service revenues); and (2) operating and capital subsidies (from municipal and provincial governments, as well as other sources). The above figure reports operating revenues; total subsidies; and total operating revenues per kilometer.
Source: Passenger Bus and Urban Transit (Statistics Canada 1975 to 1993).

The prognosis, then, according to this analysis, is not very encouraging. Despite a huge increase in government subsidies, Canadian urban transit ridership on a per kilometer basis has remained relatively consistent (see Table 1). In fact, over the last two decades, Canada's urban transit system has attained no "new" ridership (at least not on a per kilometer basis). As remarkable as this finding may be, however, it in no way implies that government subsidies are entirely ineffective; on the contrary, increased government subsidies have indeed been very effective at permitting urban transit in Canada to become increasingly inefficient. That is, government subsidies have kept the Canadian urban transit system viable despite its growing costs and declining operating revenues. Furthermore, the evidence also suggests that depending solely on increasing government subsidies to make transit more affordable, and thereby enticing private transport users to switch to using urban transit is not, in itself, sufficient.[9] In the first place, it is not very likely that in upcoming times of increasing fiscal austerity, governments will see fit to overturn the recent "cap" on transit subsidies and increase funding in order to increase ridership — particularly since the evidence clearly indicates that ridership on a per kilometer basis has remained consistent despite earlier increases in subsidies. And secondly, even though most behavioural modal choice analysts agree that the key tradeoff for private car users is between convenience and cost,[10] many also warn that merely improving the affordability of urban transit (the competing mode in this instance) may not be the best approach for attracting private car users (Daniels and Warns, 1980). "For car travellers there is the added consideration that few of them can adequately estimate the true money costs of, for example, their journey-to-work trips, though they can arrive at a time cost (a prospective measure of convenience)[11] rather more easily" (1980:205). But that being said, the basic question still remains: How can Canada's urban transit system successfully generate "new" ridership? The next section examines, empirically, one potential catalyst that only recently has begun to capture the attention of many transport analysts.

Part III: Achieving sustainable transportation: a function of controlling rising social costs

The emergence of the "New Environmental Paradigm (NEP)" (Yankelovich, 1972; Harman, 1977; Henderson, 1976; Dunlap and van Liere, 1978; Dunlap, 1980) in the early '70s has played an important role in persuading members of the social science community, transportation analysts included, to refrain from examining social phenomena using only the traditional *anthropocentric* worldview — one that emphasizes technological progress, economic growth, and overall prosperity — and to focus, instead, on a new *biocentric* worldview — one that places non-human nature ahead of human nature. Consequently, many transportation analysts (i.e. Himanen et al., 1990; Nijkamp et al., 1990; Santos and Braga, 1988) have recently begun to focus their research more closely on three important areas: (1) the link between transportation and environmental degradation; (2) public reaction toward the problem of deteriorating environmental quality; and (3) the resulting ramifications for sustainable transportation. This part of the analysis deals specifically with two related propositions that arise from this new paradigmatic approach to modal choice analysis. The first is that publics in advanced industrial

---

[9] In all fairness, however, it is important to note that Goodwin's (1992) recent study confirms that prices do have a powerful cumulative effect on travel demand patterns, but "that behavioural response to cost changes is a response that takes place over time" (1992:163). Moreover, pricing strategies continue to gain increasing popularity; they are currently being considered by many transportation analysts as a prospective method for dealing with road congestion (see Jones and Hervik, 1992; Giuliano, 1992; Evans, 1992; Jones, 1991).

[10] See Beesley (1965); Quarmby (1967); Lisco (1968); Stopher (1968); and Wilson (1967).

[11] In brackets is mine.

countries, Canada included, have become increasingly concerned about and/or actively involved in the environmental cause. And second, rising levels of environmental concern and/or activism, in turn, motivates publics to search for more sustainable forms of transport (i.e. public transport).

In relation to the first proposition, consider three recent, but fairly well-documented broad-gauged shifts that many scholars claim have taken place across most advanced industrial societies: *increasing public support for the environmental movement and its goal of "environmental protection;" the rise in postmaterialist value orientations; and rising levels of cognitive mobilization.* Each of these shifts, may be at least partly responsible for stimulating increased levels of both environmental concern and activism. The first shift, for example, indicates that an overwhelming number of people in multiple advanced industrial states, unequivocally approve of the environmental movement and its basic goal of "environmental protection" (Dunlap 1992; Dunlap, Gallup G., and Gallup, A., 1993; Dalton, 1988, 1994; Milbrath, 1984; 1989). In fact, one recent crossnational exploration even finds that public support for the goal of "environmental protection" in 17 European countries as well as the United States, is stronger than public support for six other new social movements (Mertig and Dunlap, 1995). Likewise, there is also an abundance of similar evidence which suggests that Canadians too have become more supportive of the environmental movement (Bakvis and Nevitte, 1989; OECD, 1991; Nevitte and Kanji, 1995; Kanji and Nevitte, 1995). Consequently, because publics increasingly support, at least in principle, both the environmental movement and its basic goal of "environmental protection," they are also more likely to be concerned about, and actively willing to pursue the environmental cause.

The second broad-gauged shift contends that the value systems of publics are changing. The most widely documented variant of this argument comes from the work of Ronald Inglehart (1977, 1988, 1990a, 1990b). Inglehart claims that "the values of Western publics have been shifting from an overwhelming emphasis on material well-being and physical security (materialism) toward a greater emphasis on the quality of life (postmaterialism)" (1990a:5).[12] Postwar prosperity, Inglehart argues, has meant that large proportions of Western publics have been socialized under conditions of improved economic and physical security. As a result, these publics now place less emphasis on material and physical security, and a higher priority on non-material needs like a sense of community and the quality of life. One particular quality of life issue that postmaterialists place a great deal of emphasis on is the state of the environment: postmaterialists are far more likely to be concerned and active supporters of the environmental movement than are materialists (Inglehart, 1988:1224).

A third broad-gauged shift considers the changing "cognitive mobility" of publics in advanced industrial states. Dalton (1984, 1988), for example, claims that certain structural transformations like the education explosion, the technological revolution, and the ability to rapidly disseminate information have jointly contributed to producing a new brand of citizen — one that is more cognitively mobile. More specifically, "cognitive mobilization means that citizens possess the level of political skills and resources necessary to become self-sufficient in politics" (1988:18). As such, publics may be more concerned and active when it comes to the environment because they are more interested, informed, and educated about the relevant issues (Dalton, 1984, 1988). Furthermore, cognitively mobile publics are also better positioned to make the important connections between pressing environmental issues and their potential causes and ramifications.[13]

---

[12]In brackets is mine.

[13]There is no denying that postmaterialism and cognitive mobilization may be closely allied; after all, each argument emphasizes different aspects of a common underlying phenomenon - the transition from smokestack industrialism to late industrialism. But despite their apparent similarity, however, there are two important justifications for examining the effects of each argument separately. First, each

The remainder of this paper employs direct empirical evidence from the 1990 WVS to examine the following three questions: (1) Do all three of the proposed broad-gauged shifts affect both environmental concern and activism in Canada? (2) Are those Canadians who are concerned about, and actively involved in the environmental cause, also more motivated to use urban transit? And (3), is the link between environmental concern, activism, and urban transit usage generationally driven? The latter question is particularly important because it explores both the long-term consequences of environmentalism and its effects on sustainable transportation. Many preceding studies have demonstrated that at least two of the three proposed broad-gauged shifts — postmaterialist value orientations and cognitive mobilization — may be linked to inter-generational change (Inglehart, 1990; Dalton, 1988). Consequently, if either of these two prospective explanations, together with "age," emerge as being significant predictors of environmental concern and/or action, then one would expect that the motivational importance of both environmental concern and activism for urban transit usage, if it is indeed significant, will likely increase over time.

Figure 2 measures each of the prospective links using basic gamma coefficients.[14] By conventional empirical standards, a gamma coefficient which ranges from 0 to $\pm$.1 generally signifies a weak relationship. Gamma coefficients that are greater than $\pm$.1 but less than $\pm$.2 reflect a relationship of moderate strength. And any gamma coefficients that have a magnitude of $\pm$.2 or greater, typically represent a strong relationship. Keeping those guidelines in mind, then, consider the impact of the three candidate broad-gauged shifts on environmental concern. At least two of the three shifts — public approval of the environmental movement and its basic goal of "environmental protection," and cognitive mobilization — turn out to have very strong effects. Both Canadians who approve in principle of the environmental movement and its basic goal of "environmental protection," as well as those who are cognitively mobile, tend to express high levels of concern for the environment. Postmaterialism, the third prospective broad-gauged shift, only has a moderate effect. As surprising as this finding may seem, it is not that inconsistent with Inglehart's own results (Inglehart, 1990b). Inglehart, himself, reports a relatively weak association between postmaterialist value orientations and attitudes toward environmentalism. Environmental concern, Inglehart argues, likely represents a "soft" attitudinal indicator, one that taps a large number of "spur of the moment" responses. Consequently, the true reservations that materialists are likely to have about the importance of environmentalism, ones that would set them far apart from postmaterialists, are not likely to become evident until more behavioural types of indicators (like environmental activism) are brought into play.

Age and environmental concern are also linked. In this case, age turns out to have a powerful, inverse effect: young Canadians, more so than older ones, tend to be highly concerned about the environment. This finding, when combined with the strong effect of

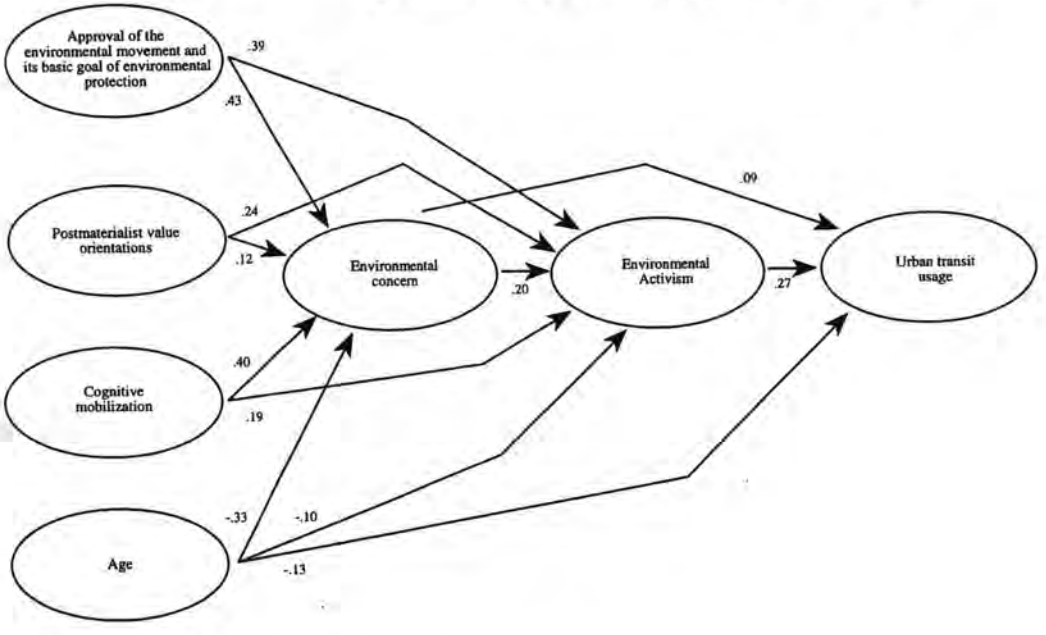
---

explanation is analytically distinct: cognitive mobility and values are not the same thing — they tap individual characteristics that are conceptually different. Secondly, because each explanation implicates different variables as lying at the root cause of changing environmental orientations, it should be possible to empirically evaluate the merits of each explanation in quite distinct ways.

[14]Each variable is operationalized as follows: "public support for the environmental movement and the goal of 'environmental protection'" (0=low level of support, 1=high level of support); "postmaterialist value orientations" (0=materialist, .5=mixed, 1=postmaterialist); "cognitive mobilization" (0=low level of cognitive mobilization, .5=moderate level of cognitive mobilization, 1=high level of cognitive mobilization); "age" (0=young, 1=old); "environmental concern" (0=low level of concern, .5=moderate level of concern, 1=high level of concern); and "environmental activism" (0=low level of activism, .5=moderate level of activism, 1=high level of activism). Data for urban transit usage are taken from the following Statistics Canada report: *Urban Transit Statistics* (provincial totals on a monthly basis for the year 1990). These ridership figures were then standardized on a per capita basis, and operationalized as follows: "urban transit usage" (0=low level of usage, .5=moderate level of usage, 1=high level of usage).

Figure 2

Environmental concern, environmental activism, and urban transit usage



Note: The above figures are gamma coefficients.
Sources: 1990 World Values Surveys, and Urban Transit Statistics (Statistics Canada 1975 to 1993).

cognitive mobilization (a prospective inter-generational shift), lends at least preliminary support for the claim that environmental concern in Canada may be generationally driven. Indeed, if that is the case, then environmental concern is likely to become a much more influential factor in future transportation planning.[15]

When it comes to environmental activism, again only two of the three broad-gauged shifts turn out to have powerful effects. Specifically, publics who approve, in principle, of the environmental movement and its basic goal of "environmental protection," as well as those publics who have postmaterialistic value orientations emerge as being more willing than those who are non-approving and/or materialistic to actively pursue the environmental cause. The latter finding, in this instance, while it is certainly somewhat weaker in magnitude than the former, adds grist to Inglehart's earlier contention that the differences between the level of concern that materialists, as opposed to postmaterialists, place on environmental issues are more likely to become manifest when examining behavioural indicators. The third broad-gauged shift — cognitive mobilization — has only a moderate effect on environmental activism, while age, the lone socio-demographic indicator, has a weak effect. Conceptually, however, both findings parallel the earlier findings for environmental concern: younger Canadians, and those who are cognitively mobile are more likely than those who are older, and less cognitively mobile, to be actively involved in the environmental cause.[16] Note too, that environmental concern also turns out to have a powerful effect on environmental activism. Not surprisingly, those people that are highly concerned about the environment are also most likely to be highly active in the environmental cause.

What, then, do these findings mean for urban transit usage In Canada? The prognosis, in this instance, appears to be much more promising than was the case for increased government subsidies. Even though environmental concern has only a weak effect on urban transit usage, it has a powerful impact on environmental activism. And environmental activism, in turn, has a powerful direct effect on urban transit usage. Simply put, those Canadians that actively pursue the environmental cause are also highly likely to ride urban transit. This latter finding probably suggests that for publics who are actively involved in the environmental movement, urban transit usage may serve as one plausible medium of activism. In addition, the evidence also indicates that age has a moderate effect on transit usage: younger Canadians are more likely than older Canadians to take urban transit. Altogether, then, the findings reported in Figure 2 lend at least preliminary support to the claim that sustainable transportation in Canada may, indeed, be a function of controlling rising social costs.

Conclusions

This paper began with an important modal choice problem that currently faces many advanced industrial countries: the bulk of recent crossnational evidence suggests that both private car ownership and usage are on the rise, while urban transit patronage, continues to steadily decline. The findings of this crosstime investigation indicate, however, that in Canada, urban transit usage, at least over the last two decades, has been remarkably stable; since 1975, Canadian urban transit services have consistently carried an average of two passengers per kilometer. Perhaps of more concern, though, is another finding which suggests that despite a significant expansion in urban transit services, as well as huge increases in government subsidies, Canada's urban transit system has been largely unsuccessful at generating any "new" ridership (at least on a per kilometer basis).

---

[15] Admittedly, however, much more detailed analysis would be required in order to convincingly substantiate this claim.

[16] The weak age effect in this case may be because environmental activism taps the public's willingness to make additional financial contributions to the environmental cause, and young Canadians simply do not have adequate disposable income with which to participate with more vigor.

Instead, what has occurred, is that transit's operating expenses have basically skyrocketed (by more than four times). While it certainly may be true that at least part of the recent increase in urban transit's operating expenses is likely attributable to the expansion of transit services and/or basic inflation, there is also strong evidence to suggest that the urban transit operation in Canada has become increasingly inefficient. On a more positive note, however, there is evidence to suggest that increased public awareness of the potential social (i.e. environmental) costs associated with rising automobile usage may serve as a potential catalyst for achieving sustainable transportation. In particular, both rising public concern for the environment as well as increased public willingness to actively pursue the environmental cause, emerge as being positively associated with high levels of urban transit usage. Furthermore, this finding appears, at least initially, to be generationally driven; that is, both environmental concern and activism are likely to become increasingly influential at motivating more publics to use more ecologically viable forms of public transport.

Based on these findings, then, this paper makes three basic suggestions for both ensuring the long-term survival of urban transit in Canada, as well as potentially increasing "new" ridership. The first suggestion calls for urban transit to move away from being largely dependent on government subsidies, and strive instead on becoming more efficient at supplying urban transportation services. The long-term prospects for attaining a viable urban transit system in Canada depends heavily on such a transition taking place. Government subsidies already account for more than half of transit's total operating revenues, and it seems highly unlikely that in times of increasing fiscal austerity that governments will see fit to overturn a recent "cap" on urban transit subsidies. Also, any proposed fiscal initiatives designed to make urban transit more affordable, and thereby increase ridership, should be reconsidered. Over the last two decades, increased government subsidies have not been successful at generating any "new" ridership; instead, they have simply helped to further subsidize a "captive" core of riders, and in turn make the entire operation increasingly inefficient.

The second suggestion is for governments to introduce stronger regulatory measures aimed directly at detouring unnecessary private car usage. According to many modal choice analysts, the basic tradeoff for most private car users is between convenience and cost. The evidence from this investigation clearly demonstrates, however, that increased subsidization of urban transit has not been very effective at attracting "new" transit ridership in Canada. The other alternative, then, is to implement more regulatory initiatives which focus on making private car transportation seem less convenient. Some examples include: introducing increased traffic priority for transit; limiting downtown parking space; and providing sub-urban private car users with increased access to park and ride stations. Much of the preliminary European evidence indicates that these sorts of regulatory initiatives have been very effective at generating increased urban transit ridership (Pucher and Kurth, 1995). More importantly, particularly for the Canadian urban transit system, is the fact that regulatory initiatives require little, if any, government subsidies.

The third suggestion is simply for transit marketers to continue promoting increased public awareness of the rising social costs associated with escalating private car usage. There is an emerging literature which suggests that publics are better able to make the connection between environmental problems, their sources, and their ramifications. Moreover, the findings from Part III of this analysis indicate that there is a definite link between environmental concern, environmental activism, and transit usage in Canada. Increased public concern for the environment and rising levels of environmental activism may be two important catalysts for achieving sustainable transportation, and urban transit marketers should exploit that opportunity with increased vigor. Moreover, because most initial indications suggest that both environmental concern and activism may be driven by

inter-generational change, the likelihood is that both prospective trends will likely become increasingly powerful at motivating publics to use urban transit.

Bibliography

Bakvis, H. and N. Nevitte
1992 "The Greening of the Canadian Electorate: Environmentalism, Ideology and Partisanship," in R. Broadman (ed.) *Canadian Environmentalism Policy: Ecosystems, Politics and Process.* Toronto: Oxford University Press.

Banister, D. and L. Pickup
1989 *Urban Transport and Planning: a bibliography with abstracts.* London and New York: Mansell Publishing Limited.

Beesley, M.
1965 "The value of time spent in travelling: some new evidence," *Economica,* 32:175-85.

Dalton, Russell J.
1984 "Cognitive Mobilization and Partisan Dealignment in Advanced Industrial Democracies," *Journal of Politics* 46:264-284.
1988 *Citizen Politics in Western Democracies: Public Opinion and Political Parties in the United States, Great Britain, West Germany and France.* Chatham, New Jersey: Chatham House Publishers, Inc.
1994 *The Green Rainbow: Environmental Groups in Western Europe.* New Haven & London: Yale University Press.

Daniels, P. and A. Warnes
1980 *Movement in Cities: Spatial perspectives on urban transport and travel.* London and New York: Methuen.

Dunlap, R.
1980 "Paradigmatic change in social science: from human exemptions to an ecological paradigm," *American Behavioural Scientist,* 24:5-14.
1992 "Trends in Public Opinion Toward Environmental Issues: 1960-1990" in Dunlap, R. and Mertig, A. (eds) *American Environmentalism: The U.S. Environmental Movement, 1970-1990.* Philadelphia, Taylor and Francis.

Dunlap, R., G. Gallup, and A. Gallup
1993 *Health of the Planet.* The George H. Gallup International Institute, Princeton, New Jersey.

Dunlap, R. and K. Van Liere
1978 "The new environmental paradigm: A proposed measuring instrument and preliminary results," *Journal of Environmental Education,* 9:10-19.

Evans, A.
1992 "Road Congestion Pricing: When Is It a Good Policy?" *Journal of Transport Economics and Policy,* 26:3:213-43.

Goodwin, P.
1992 "A Review of New Demand Elasticities with Special Reference to Short- and Long-Run Effects of Price Changes," *Journal of Transport Economics and Policy,* 26:2:155-70.

Giuliano, G.
1992 "An Assessment of the Political Acceptability of Congestion Pricing," *Transportation,* 19:335-58.

Gratwick, J.
1993 *Canadian Transportation: Origins, Perspectives and Prospects.* Canada: John Gratwick.

Harman, W.
1977 "The Coming Transformation," *The Futurist,* 11:106-112.

Henderson, H.
1976 "Ideologies, Paradigms, and Myths: Changes in our Operative Social Values," *Liberal Education*, 62:143-157.
Himanen, V., P. Nijkamp, and J. Padjen
1992 "Environmental Quality and Transport Policy in Europe," *Transportation Research*, 26A:2:147-157.
Inglehart, R.
1977 *The Silent Revolution*. Princeton NJ: Princeton University Press.
1988 "Cultural Change in Advanced Industrial Societies: Postmaterialist Values and Their Consequences," *International Review of Sociology*, 3, 77-100.
1990a *Culture Shift in Advanced Industrial Society*. Princeton: Princeton University Press.
1990b "Values, Ideology, and Cognitive Mobilization in New Social Movements" in R. J. Dalton and M. Keuchler (eds.) *Challenging the Political Order: New Social and Political Movements in Western Democracies*. New York: Oxford University Press: 43-66.
Jones, P.
1991 "Gaining public support for road pricing through a package approach," *Traffic Engineering and Control*, 32:4:194-6.
Jones, P. and A. Hervik
1992 "Restraining Car Traffic in Road Pricing," *Transportation Review*, 26A:2:133-45.
Jones, P., M. Dix., M. Clarke, and I. Heggie
1983 *Understanding travel behaviour*. Aldershot: Gower: Oxford Studies in Transport
Kanji, M. S. and N. Nevitte
1995 "North American Reactions to Environmentalism." Paper presented at Border Demographics and Regional Interdependency: A Trinational Symposium, Western Washington University, Bellingham, Washington, February 16-18.
Lave, C.
1991 "Future Growth of Auto Travel in the U.S.: A Non-Problem," in J. W. Tester et al. (eds.), *Energy and the Environment in the 21st Century*. Cambridge: M.I.T. Press).
Lisco, T.
1968 "The value of commuters' travel time: A study in urban transportation." *Highway Research Record*, 245.
Mertig, A. and R. Dunlap
1995 "Public Approval of Environmental Protection and Other New Social Movement Goals in Western Europe and the United States," *International Journal of Public Opinion Research*, 7:2:145-156.
Milbrath, L.
1984 *Environmentalists: Vanguard for a New Society*. Albany, SUNY Press.
1989 *Envisioning a sustainable society: Learning our way out*. Albany, State University of New York Press.
Nevitte, N. and M. S. Kanji
1995 "Explaining Environmental Concern And Action In Canada," *Applied Behavioral Science Review*, 3:1, 85-102.
Newman, P. and J. Kenworthy
1989 *Cities and Automobile Dependence: An International Sourcebook*. Aldershot: Grower Technical.
Nijkamp, P., S. Reichman, and M. Wegener
1990 *Euromobile: Transport, Communications and Mobility in Europe: A Cross-National Comparative Overview*. Avebury, Gower Publishing Company Limited.

OECD
1995 *Reconciliation Transportation, Energy and Environmental Issues: The Role of Public Transport.* Budapest, Hungary: European Conference of Ministers of Transport and European Commission.
1991 *The State of the Environment.* Paris: Organization of Economic Cooperation and Development.
Perl, A. and J. Pucher
1995 "Transit in Trouble?: The Policy Challenge Posed by Canada's Changing Urban Mobility," *Canadian Public Policy,* 261-283.
Pucher, J.
1995 "Urban Passenger Transport in the United States and Europe: a comparative analysis of public policies (Part 1. Travel behaviour, urban development and automobile use; and Part 2. Public transport, overall comparisons and recommendations)," *Transport Reviews,* 15:2&3:99-117/211-227.
Pucher, J. and S. Kurth
1995 "Making Transit Irresistible: Lessons from Europe," *Transportation Quarterly,* 49:1:117-128.
Pushkarev, B. and J. Zupan
1977 *Public Transportation and Land Use Policy.* Bloomington and London: Indiana University Press.
Quarmby, D.
1967 "Choice of Travel Mode for the Journey to Work," *Journal of Transport Economics and Policy,* 1:273-314.
Sampson, R., F. Martin, and D. Schrock
1990 *Domestic Transportation: Practice, Theory, and Policy (6th edition).* Boston: Houghton Mifflin Company.
Santos, M. and M. Braga
1988 "Research Trends in Urban Transport Science: Some Empirical Evidence from Academic Research, *Transport Research,* 22:1:57-70.
Sheppard, E.
1986 "Modelling and predicting aggregate flows," in Hanson, S. (ed.) *The Geography of Urban Transportation,* New York: Guildford.
Statistics Canada
1990 *Urban Transit Statistics.*
1975 to 1993 *Passenger Bus and Urban Transit*
Stopher, P.
1968 "Predicting travel mode choice for the journey to work," *Traffic Engineering and Control,* 9:436-9.
Turton, B.
1992 "Urban Transport Patterns," in Hoyle, B. and R. Knowles (eds.) *Modern Transport Geography.* London and New York: Belhaven Press.
Turton, B. and R. Knowles
1992 "Urban Transport Problems and Solutions," in Hoyle, B. and R. Knowles (eds.) *Modern Transport Geography.* London and New York: Belhaven Press.
Webster, F., N. Paulley, P. Bly, M. Dasgupta, R. Johnston
1985 *Changing Patterns of Urban Travel.* European Conference of Ministers of Transport, Paris.
Wilson, F.
1967 *Journey to Work: Modal Split.* London, MacLaren.
1990 *World Values Surveys*
Yankelovich, D.
1972 "The New Naturalism," *Saturday Review,* 55:32-37.

ASSESSING NON-CONFORMING VEHICLES FOR NET TRANSPORT AND SAFETY BENEFIT

by John Woodrooffe*, Doug Johnson**, John Palmer**, Dave Smith**

This paper puts forward a methodology for the objective assessment of non-conforming transport vehicle performance. It offers a rationale to examine vehicle stability and control, pavement and bridge impact, operational concerns and safety issues in order to determine if a special non conforming permit can be justified. This paper uses a case study approach to illustrate the practical mechanics of the process recently used by the province of New Brunswick.

The approach is particularly important for cases where divisible loads are considered. It uses an engineering method to establish the net impact on safety, the infrastructure and transport efficiency and serves as a mechanism for achieving improved safety performance from the carriers.

1. INTRODUCTION

The paper addresses the question of how to determine if a "non-conforming" vehicle should be permitted to operate within the general road network or on designated routes and if so, what conditions of operation should apply. New Brunswick has recently developed a new procedure that considers vehicle safety, performance, the influence of road and bridge life, road geometrics and traffic considerations. The procedure is based on the following principles:

(1) Any request from industry for use of a non-conforming vehicle, on all of or part of the highway system, will be considered and evaluated using sound engineering principles and analysis.
(2) If the theoretical engineering analysis proves that the vehicle can be operated safely, a pilot project using one or more vehicles will be carried out for up to one year to test the theoretical analysis and ensure that there are no serious unforeseen problems.
(3) If the pilot project proves that the theoretical analysis was correct and no further safety issues were raised during the pilot project, then a "non-conforming" permit will be issued.

---

* Roaduser Research International
** New Brunswick Department of Transportation

(4) The theoretical analysis will be comprised of the following components:

- Vehicle Safety Performance and Stability Assessment
- Load Security Assessment
- Bridge and Road Loading Impact Assessment
- Route Geometry and Traffic Assessment

(5) Weight considerations are not to exceed those in the current New Brunswick weight regulations.

(6) Other vehicle safety enhancements may be mandated for trial during the pilot project and beyond.

(7) All costs of pilot vehicles are the responsibility of the operator.

2. THE CASE STUDY VEHICLE

The particular case described in this paper involves a 26.5 meter 8 axle "B" Train designed to transport large rolls of tissue paper between two cities on a continuous year round basis. The vehicle exceeded the provincial overall length limit of 25 meters.

The case study vehicle is a 8-axle B-train with an overall length of 26.42 meters (1.42 meters longer than current regulations permit). The loaded gross vehicle weight of the vehicle was expected to be 46,500 kg. The current maximum gross vehicle weight for a 8-axle B-train in New Brunswick is 62,500 kg. The original vehicle proposed by the applicant was a 7-axle B-train, however analysis indicated that an 8-axle unit was significantly superior.

The vehicle has been designed to transport 10 large paper rolls (5 in each trailer). The rolls are 261.6 cm (103 inches) high. The width and weight of the roll can vary but nominal values are; width 213.4 cm (84 inches) and weight 2,500kg (5,500 lbs).

The benefits as outlined in the request associated with this proposal were as follows:

- Savings in transportation costs estimated to exceed $260,000 annually.
- Reduction in annual truck volumes from current procedures of 1,286 loads or 27%.

3. ASSESSMENT OF DYNAMIC PERFORMANCE

Using the procedures developed in the RTAC Heavy Vehicle Weights and Dimensions study report (1) and other measures, the following vehicle performance attributes were examined using the University of Michigan Transportation Research Institute (UMTRI) Yaw/Roll program:

- roll stability
- load transfer ratio
- rearward amplification
- low speed offtracking
- high-speed steady-state offtracking
- high speed dynamic offtracking
- friction demand

3.1 STEADY-STATE ROLLOVER

Heavy trucks can withstand limited lateral acceleration in a curve before rolling over. This rollover limit is expressed in terms of the lateral acceleration, and is given as a proportion of gravitational acceleration (g). Total rollover occurs when all the wheels on one side of the combination vehicle (on the inside of the turn) lift off the road surface, as shown in Figure 1.

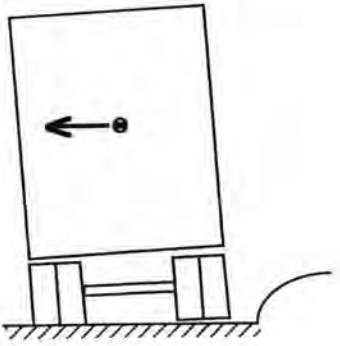
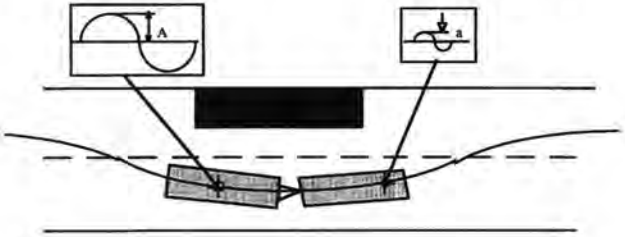


Figure 1. Onset of vehicle rollover

3.2 REARWARD AMPLIFICATION

When articulated vehicles undergo rapid steering, the steering effect at the trailer is magnified. This results in increased side force, or lateral acceleration, acting on the rear trailer. This in turn increases the likelihood of the trailer rolling over under some circumstances.

Rearward amplification is defined as the ratio of the lateral acceleration at COG of the rearmost unit to that at the hauling unit in a dynamic maneuver of a particular frequency. As shown in Figure 2, steering from side to side produces more lateral movement at the rear unit than at the hauling unit.



RA= A/a

Figure 2. Rearward amplification.

3.3 LOAD TRANSFER RATIO

Load transfer ratio is defined as the proportion of load on one side of a vehicle unit transferred to the other side of the vehicle in a transient maneuver. Where vehicle units are roll-coupled - as in B-doubles - the load transfer ratio is computed for all axles on the vehicle. When the vehicle units are not roll-coupled, as with truck-pony configurations, they are evaluated as separate units. When the load transfer ratio reaches a value of 1, rollover is about to occur. The LTR is the ultimate measure of rollover stability.

3.4 HIGH SPEED STEADY-STATE OFFTRACKING

High-speed offtracking is defined as the extent to which the rearmost tires of the vehicle track outboard of the tires of the hauling unit in a steady-turn at highway speed. High-speed offtracking relates closely to road width requirements for the travel of combination vehicles and is an important part of the total swept width of the combination vehicle (that is, the extent to which the lateral excursions of the rear of the vehicle exceed those of the hauling unit in normal operation).

3.5 HIGH-SPEED DYNAMIC OFFTRACKING

High speed dynamic offtracking is a measure of the lateral excursion of the rear of the vehicle with reference to the path taken by the front of the vehicle during a dynamic maneuver. This expresses the amount of additional road space used by the vehicle combination in an avoidance maneuver.

High speed dynamic offtracking is computed for the lane-change maneuver at a speed of 90 km/h. In each case, the vehicle was made to follow a path so that the lateral displacement of the vehicle's path remained constant.

3.6 LOW-SPEED OFFTRACKING

Low-speed offtracking represents a measure of the swept path of the vehicle and its lateral road space requirement when turning at intersections or when turning into loading areas. The offtracking of the vehicle depends on axle position, trailer tongue length and hitch location.

3.7 FRICTION DEMAND

Friction demand is the non-tractive friction levels between the tires and the road surface at the steer axle of the truck. It is a measure of the shear force between the tires and the road that results from the vehicle negotiating a curve in the road. As the aligning moment of an axle group increases, this measure becomes more significant. Friction demand at the steer axle and drive axles is evaluated at high and low speed. The RTAC measures focus only on low speed friction demand of the drive axles. From experience, Roaduser Research has found it to be more practical to examine the additional low-speed friction demand at the steer axle and high-speed friction demand at the drive axles.

3.8 RESULTS

Table 1 contains the results of the assessment of dynamic performance. In general, the vehicle demonstrated very good dynamic behavior.

Performance Measures	TAC Max. Values	Candidate Vehicle	Grade
Static roll threshold by axle group			
• Drive axles		0.425	Very good
• Lead trailer axles		0.422	Very good
• Pup trailer axles		0.330	Marginal
Net static roll threshold (est.)		0.370	Good
Load transfer ratio	0.60	0.375	Pass
Rearward amplification	2.00	1.35	Pass
Low speed offtracking	6.00 m	5.63 m	Pass
High speed offtracking	0.46 m	0.40 m	Pass
High speed dynamic offtracking	0.80 m	0.22 m	Pass
High speed friction demand (drive axles)		0.121	
Low speed friction demand			
• Steer axle		0.213	
• Drive axles	0.10	0.028	Pass

Table 1. Assessment of dynamic performance.

The 7-axle B-Train passed the RTAC performance measures, however the vehicle does exhibit non uniform roll resistance at the center axle group supporting both trailers. It was found that significant improvements in roll stability can be realized by converting the vehicle to an 8 axle B-Train. This is done by replacing the tandem axle group, at the junction between the lead and pup trailers, with a close spread tridem.

4. ASSESSMENT OF VEHICLE WEIGHT AND DIMENSIONS AS REQUIRED BY BRIDGE DESIGN CAPACITY AND LIVE LOADING

New Brunswick has developed an extensive bridge assessment model, which bridge engineering officials from provinces in the Eastern Task Force have endorsed. The following is a brief summary of how this model was developed and how it is used.

The majority of regulating authorities in Canada use the present CSA S6-88 Code for the design of highway bridges. The load model that is used by most highway authorities is the CS-600 truck model as shown in Figure 3, where 600 represents the GVW in kilo-Newtons. The CS-600 model is also often used for evaluating existing bridges and was chosen to simulate a B-Train configuration with a maximum GVW of 62,500 kg. It must be noted that the physical model for designing bridges may not resemble a "real" vehicle but is chosen to represent force effects in bridges that result from real vehicles.

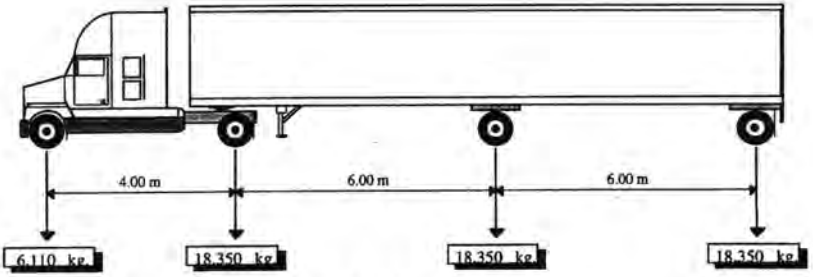


Figure 3 - CS 600 Design Truck

Truck weight and dimension regulations should be formulated such that the force effects of the entire spectrum of vehicles are at best equal to or less than those produced by the design vehicle, including any administrative tolerances.

It follows that the force effects from the vehicles when ratioed to the CS-600 model effects, must be less than or equal to 1.00. Conversely, the force effects of the design model must be greater than the corresponding force effects for the real vehicles; that is, CS-600 divided by the real force effects must be greater or equal than 1.00. This will then give a satisfactory safety index for bridges designed using the current code and model.

A major assessment of vehicles in the New Brunswick Motor Vehicle Regulation has been done based upon the methodology noted above. The methodology used compared bending moment and shear (over a range of span lengths) resulting from the current Canadian design truck model; with those resulting from the vehicles conforming to the N. B. Regulations. The actual vehicles consisted of various weights and spacings of single, tandem and tridem axle groups. This was done for simple span bridges with span lengths between 3 meters and 40 meters. The results are also be applicable to the central regions, or positive moment regions, of continuous span bridges with span lengths that are approximately 70 % longer than the simple span length.

An example of the comparison of the moment effects resulting from a short tridem-semi trailer truck with three different configurations is shown in Figure 4. A graphical representation of the comparison is shown in Figure 5. This same comparison would also be made for shear force effects in bridges.

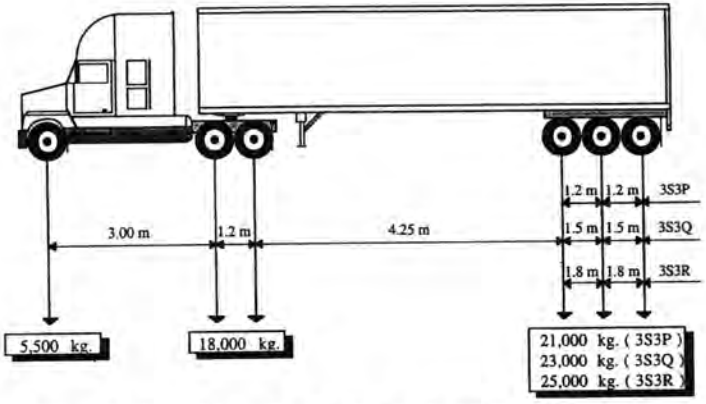


Figure 4 - Short Tridem Semi-Trailer Trucks (3S3P, 3S3Q, 3S3R)

When comparisons of various vehicles are made in this manner, it is evident that the allowable gross vehicle weights and individual axle weights depend upon interaxle spacings and overall vehicle base length. Generally, vehicles with smaller base lengths will be allowed smaller GVW and in some cases, smaller individual axle loads.

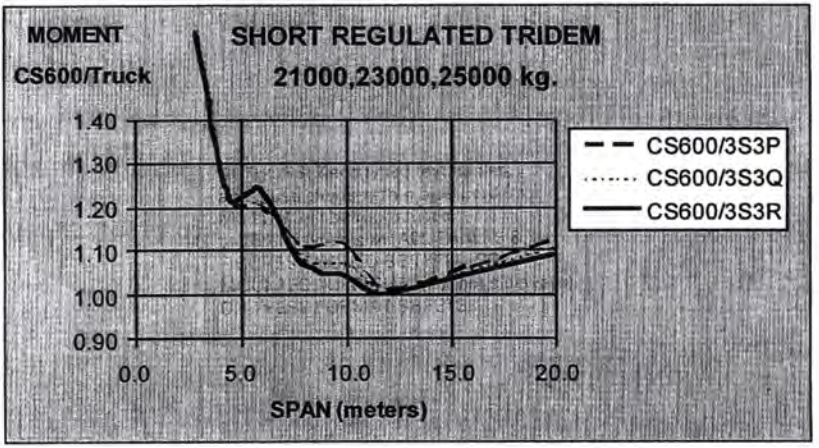


Figure 5 - Moment Comparison of Tridem Truck To CS600

It is important to note that this comparison is the best-case scenario because it applies only to bridges designed and constructed to recent bridge codes. Existing bridges may be evaluated for non-permit vehicular traffic using the CS600 model with Clause 12 of the CSA Standard S6 - 88 Design of Highway Bridges. The CS600 model must be treated as a real truck with the associated load factors when using Clause 12, if the model is to be used as a representation of non-permit traffic. The correlation of non-permit vehicular traffic to the design model must be established by the Regulatory Authority by comparing moment and shear effects, as shown in the preceding example.

Bridges must be assessed for strength on an individual basis in order to determine the live load capacity as compared to the model. When this model (CS600) is used to evaluate existing bridges for non-permit vehicular traffic, the ratio of the capacity of the bridge to the applied factored forces caused by the CS600 model is often less than 1.00. This implies that such a bridge may not be

capable of safely resisting loads that are in excess of the model, and may only be capable of safely carrying a fraction of the weights that are represented by the model. Since the inventory of bridges, may only be capable of safely resisting a portion of the CS600 model, it follows that force effects are greater than the design and evaluation model are not acceptable (from a standpoint of public safety and structural damage to existing bridge infrastructure).

The Structures Branch of the New Brunswick Department of Transportation analyzed the affects of the 26.5 m B-train and found it to be well within the bridge capacity guidelines for the intended route.

5. ROAD LIFE ASSESSMENT

The Equivalent Standard Axle Loads (ESAL's) are calculated to quantify the impact that a vehicle will have on the roads. The ESAL's per vehicle are based on standard axle loads of:

- 8200 kg single
- 14,500 kg tandem
- 19100 kg tridem

The relationship used is a s follows:

$$\text{ESAL} = \left(\frac{\text{Design Axle Loads}}{\text{Standard Axle Load}} \right)^{3.8}$$

The original request considered a 7 axle B-Train with anticipated axle loads of:

- steer axle 4,620 kg
- drive axles 12,857 kg
- center tandem 17,583 kg
- last tandem 10,559 kg

The total ESAL's for the 7-axle B-train are 3.13 ESAL's per vehicle or 0.31 ESAL's per roll of paper.

As an alternative, if an 8 axle B-train were used where a center tridem axle replaced the center tandem. The center tridem axle load reflects the increased tare weight attributed to the additional axle. The anticipated axle loads of the 8-axle B-train are as follows:

This situation requires that the rolls be offset from side to side, to create sufficient space for all 5 rolls to be loaded into each trailer as shown in figure 7.

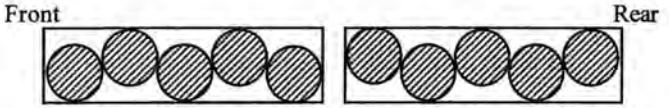


Figure 7. Offset loading pattern minimizes trailer length.

Because there are an odd number of rolls in each trailer, the lateral offset loading pattern will result in a shift of the mass center of the vehicle laterally from the centerline. The lateral shift will result in reduced vehicle roll stability. The problem can be made worse if the lead trailer is loaded identically to the rear trailer, as shown in figure 8, so that there are 4 rolls along one side of the vehicle and 6 rolls on the other side.

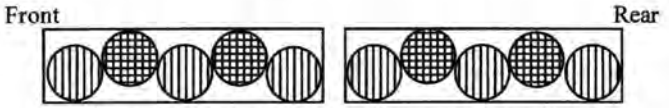


Figure 8. Worst scenario loading pattern resulting in lateral shift in the center of mass.

Such a loading situation will make the vehicle more susceptible to rollover on the side containing the extra rolls. To reduce this risk, a specific loading pattern was mandated eliminating any lateral imbalance as shown in Figure 9. The two rolls positioned in the center of the trailer required dunnage to prevent shifting. The consequence of improper loading is sufficiently serious to make it an ongoing condition for permit operation.

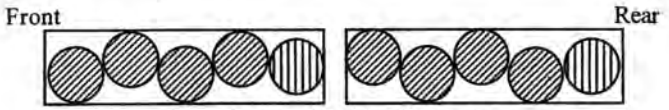


Figure 9. Mandated loading pattern to achieve maximum stability.

The rolls are protected on all surfaces by plastic wrap. Any snow or ice present on the floor of the vehicle can lead reduced friction resulting in shifting of the rolls. In addition, the wooden floors of truck vans result in poor frictional coefficients between the floor and the cargo. Rubber flooring mats

- steer axle 4,620 kg
- drive axles 12,857 kg
- center tridem 18,436 kg
- last tandem 10,559 kg

The ESAL's per vehicle for the 8 axle B-Train are reduced by 41% to 1.92 ESAL's or 0.19 ESAL's per roll compared with the 7 axle B-Train.

6. OPERATIONS

The transport company, Sunbury Transport was hauling the paper rolls in 16.15 m (53 foot) vans. A total of 8 rolls can be loaded into these trailers but because of variations in roll diameter and loading procedures, the majority of trips (70%) are made with only 7 rolls. In essence, the current 53 foot trailer moves 7.3 rolls per trip.

The proposed B-trains transport 10 rolls per trip. The company expects to transport 34,783 rolls per year between the two cities. If current equipment is to be used, 4765 trips will be required. By using the proposed B-train, 3,479 trips will be required thereby reducing the number of return trips per year by 1286, representing a 27% reduction in truck traffic for this particular haul, as mentioned earlier. This reduction of 1,286 truck movements is considered significant in the reduction of accident exposure and in truck traffic.

7. SPECIAL SAFETY ISSUES

7.1 LOAD SECURITY

The paper rolls are loaded into the vans standing on end. There is lateral free space between the rolls and the wall of the van amounting to 39 cm (15 inches). Each roll has a nominal diameter of 213 cm (84 inches). When loaded in a straight line, the rolls occupy 10.65 m which exceeds the interior space available within the trailers as shown in Figure 6.

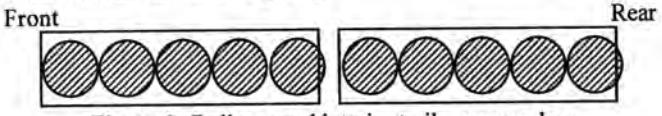


Figure 6. Rolls exceed interior trailer space when loaded in a straight line.

designed to improve the friction between the floor and the payload were mandated. Because the rolls are tall, strapping was mandated between the top of the rolls and security rails along the walls of the trailer. The compliance criteria imposed required that the rolls be restrained against longitudinal deceleration of 0.6 g for 5 seconds.

7.2 OTHER SAFETY PRECAUTIONS

The special pilot project to operate a longer (26.5) than normal (25m) B-Train, provided an opportunity to evaluate special safety precautions that would enhance the visibility and the overall safety of the vehicle as it traveled between the two cities. The following enhancements were put in place:

Two Additional rear top mounted red brake/tail/signal lights on the rear of the second trailer. Two 4" light emitting diode (LED) signal lights were mounted on each side at the top outer corners of the rear doors. An additional LED signal light was mounted on the rear of the trailer at frame height. These lights are in addition to the normal brake/tail/signals already on the stock trailer.

It is expected that the top mounted brake/tail/signal lights will be less likely to pack with snow. Because of the high location of the lights, it is also expected that in congested or urban traffic, several cars behind the B-train will have clear line of sight so as to be pre-warned of turns or braking. Conventional trailers signal lights are often "hidden" by the vehicles directly behind the trailer. Additional amber clearance lights have been installed on the top center of rear trailer and along the side of the trailers, to improve side visibility.

Reflective Tape Markings are also required as per the conspicuity specification now in place.

A Special Sign was required to be placed on the rear of the trailer indicating to the public that the length of this unit was 26.5 meters. It is expected that this sign will warn those who intend to pass, that this is a long vehicle. The sign shows a side-on image of the B-train with a 26.5 m dimension marker.

Splash-Spray suppressing devices have been required for this unit. Over all wheels except the steering axle, special fenders were fitted with plastic brushes to shed water.

Rear Under-ride protection is required as per New Brunswick's current regulation.

Side under-ride protection is being developed for trial during this pilot project with assistance of UNB's Truck Accident Investigation Group

Anti-lock brakes on the tractor drive axles have been shown to reduce jackknife accidents. The proposed tractor did not have a long wheelbase therefore it is more susceptible to jackknife induced accidents caused by over-braking. Anti-lock brakes contribute significantly to overall vehicle safety and were mandated.

Automatic brake slack adjusters are required on all wheels.

Tridem Axle Group: By replacing the center tandem axle group with a close spread tridem group, a 41% reduction in ESAL's for the vehicle will be realized. This represents a significant reduction in pavement wear attributed to the vehicle. Using a tridem axle group improves the roll stability of the vehicle by increasing the roll stiffness.

Tractor Wheelbase: The tractor specified in the design has a wheelbase of 4.5 meters (177 inches). Short wheelbase tractors are not recommended for B-Train operation (3). No tractor shorter than 4.5 meters can be used for this application.

Routing considerations: The condition of the special permit restricted the vehicle to a specific route. The roadway was studied for geometric and safety concerns and the final route was chosen after consultation with city traffic officials in Saint John and Moncton. Passing sight distances associated with the additional vehicle length were considered. No special hours of operation were deemed necessary at this time.

Load Security requirements were put in place to prevent both longitudinal and lateral movement of the 5 paper rolls being transported in each trailer. These rolls can weight up to 5600 lbs with diameters of up to 84" and 103" height. The use of cargo-belt strapping was used for lateral and longitudinal restraint such that the paper rolls could withstand a deceleration of 0.6g for at least 5 seconds. The rear roll was also restrained from falling out of the trailer should the doors open. Friction mats were placed on the floors of the trailers. A specific loading pattern was specified to eliminate any lateral imbalance of the vehicle as shown in Figure 10. The two rolls positioned in the center of the trailer will require dunnage to prevent them from shifting



Figure 10. Recommended loading pattern
to achieve maximum stability.

Driver training requirements are being mutually agreed upon.

Gross Vehicle Combination Mass is limited to 47,000 kg.

9. REFERENCES

(1) ERVIN, RD & GUY, Y (1986) The influence of weights and dimensions on the stability and control of heavy trucks in Canada. Roads and Transportation Association of Canada. Ottawa

(2) LAM, CP & BILLING, JR (1989) Stability and control characteristics of straight trucks and truck trailer combinations. Ontario Ministry of Transportation, Downsview.

(3) El-GINDY, M & WOODROOFFE J (1991) Influences of tractor wheelbase tandem-axle spread and fifth-wheel offset on commercial vehicle dynamics".

(4) TRB (1990) New trucks for greater productivity and less road wear. Transportation Research Board, National Research Council, Washington.

(5) GLENNON, JC (1988) A new and improved model of passing sight distance on two-lane highways. Transportation Research Record 1195, TRB, National Research Council, Washington.

(6) TROUTBECK, RJ (1981) Overtaking behaviour on Australian two-lane rural highways. Special report No. 20, Australian Road Research Board, Vermont South, Victoria.

BRIDGE COST ALLOCATION ISSUES

Edward Fekpe, PhD PEng

Transportation Division, Battelle Memorial Institute
505 King Avenue, Columbus OH 43229-2693, USA.

ABSTRACT

The complexities, assumptions and improvements associated with highway bridge cost allocation process, and the validity and sufficiency of intra-incremental cost allocators are presented. Fatigue damage factor weighted average daily traffic is an appealing intra-incremental allocator because it considers, the number and magnitude of expected stress cycles based, vehicle class and weight, bridge support type and functional highway type to determine the cost responsibility.

1.0 INTRODUCTION

The allocation of costs associated with highway bridges is usually based on the principle of cost recovery of short term total average costs. The allocation process attempts to attribute costs to various vehicles using the highway in an equitable manner on the basis of the expected level of damage caused by the individual vehicle categories and weight classes. Vehicle weights, live load bending moments and dynamic impacts, and some measure of usage of the facility are candidate parameters for allocating the costs.

Bridges are designed with consideration for the probability of extreme loading from heavy vehicles and fatigue damage. While the former criterion is common to all bridge types (span and material type), fatigue damage and hence the attribution of costs are considered minimal except for steel bridges. Within the context of cost allocation, certain questions respecting the validity and sufficiency of cost allocators remain, e.g. what measure of highway usage provides a better reflection of potential damage and therefore cost responsibility? Which parameter is a better discriminator for cost allocation purposes?. The paper describes the framework of allocating

bridge costs and attempts to address these questions by presenting concepts for doing so. An analysis exploring different intra-incremental cost allocators is presented. The expectation is to provide an exposition to the complexities, assumptions and new directions of bridge cost allocation process.

2.0 COMPLEXITIES

Before discussing the bridge cost allocation process it is worthy to appreciate the complexities associated with bridges within the context of cost allocation.

- Design standards or design loadings are different and many. The designs are also varied and it is difficult to specify a typical design process. Also, spans and materials, construction techniques are varied. Consequently, it is difficult to isolate the costs of the individual components from the cost of the overall structure and apply these costs to all bridge types.

- Bridge deterioration is complicated by the number of components, substructure, superstructure, deck, safety structures etc. Factors affecting deterioration and hence rehabilitation or replacement are many including, deicing (result of environment); weathering effects; traffic loading (axle weights, spacing and spreads, gross vehicle weight). The interactive effects of traffic, environment and constructional materials are complicated and not easily isolated.

- Bridges are site specific and the design may be independent of the number and types of vehicles expected to use the facility. Site specific factors are more likely to influence the costs of both substructures and superstructures considerably.

Under these complex circumstances, it is impractical to ignore the differences and treat all bridges as one or each individual bridge type with specific design details. Some level of disaggregation is necessary, that to a reasonable extent, can account for the differences and responses to the different vehicle configurations and loadings. The level of disaggregation is a function of the availability of data and the relative effects of the parameters e.g. the effect of bridge span on live load moments. Improvements on previous methods aimed at addressing some of these issues are discussed below.

3.0 BRIDGE COST ALLOCATION METHODOLOGY

Bridge cost allocation involves three main activities:

- classification of vehicles into meaningful groups;
- selection of allocating logic or methodology and cost allocators (i.e., parameters to determine cost responsibility);
- determination of cost categories and functions (i.e. relationship between vehicle and bridge characteristics and expenditure levels);

The current bridge cost allocation methodology used by the Federal Highway Administration of the US Department of Transportation is an improvement on the previous ones and is based on some basic principles (Saklas, 1995) :

- it embraces agency costs recovery as its theoretical basis;
- uses cost occasioning of short term average costs as its procedural framework;
- seeks equity as its basic goal in allocating costs.

A major improvement is disaggregation on the basis of bridge and vehicle characteristics, design increments and cost categories. Disaggregation allows the vehicle and bridge factors directly affecting performance hence expenditure to be easily identified and adequately accounted for in the analysis. The improved method considers: 10 bridge design load increments; 20 vehicle classes, 30 weight groups in increments of 2.3t (5000lbs); 12 functional classes of roads; 2 bridge support types (simple and continuous); and 4 cost categories - new construction, replacement, major rehabilitation and minor repairs.

Given identical vehicles on bridges of equal spans, the only distinguishing bridge characteristic that can affect the moment produced by the vehicles is support type, simple versus continuous. The span is an equally important bridge characteristic that determines live load moments. A simple and a continuous support bridge of the same span subjected to the same loading do not yield the same bending moments. The spans in turn depend on the functional class of the highway. Functional highway class is an indication of the level of service offered and therefore determines the composition and volume of vehicles using the facility and therefore the bridge design standards.

Another major improvement is the use of live moments as the basis of cost attribution instead of gross vehicle weight (GVW) as used in the previous methods. Bridge damage is caused by load effects rather the load itself. It is therefore important to evaluate bridges by considering the moments generated by traffic loads. Live load moments are calculated for each vehicle weight/class and for each mean span, functional class and support type. Previous methods assume GVW acting as point load on simple supported spans.

3.1 Assignment of Bridge Costs

The allocation process is a two stage incremental approach. First, vehicles are assigned to different weight cells for each bridge design increment on the basis of the vehicle's maximum live load moment and dynamic impact. The assignment is based on the potential bridge effects of traffic loads relative to the design vehicles. Second, all vehicles in each weight class are allocated costs based on some measure of facility usage. Such a parameter is termed the intra-incremental cost allocator. The premise being that vehicles in the same weight cell should be charged differently according to their usage of that facility. Saklas (1995) noted that the intra-incremental allocator is intended to account for costs to repair fatigue damage, the costs of new construction associated with fatigue based design fatigues factors; and some inherent common costs.

3.1.1 Live load moment allocator

The basic theory to assign vehicles to the various design increments was that all vehicle classes or weight groups, which produced a live load moment greater than or equal to that of the design vehicle that determine the increment, would be placed in that design vehicle's increment. For example, for the H-10 increment, the incremental costs over and above those associated with the H-5 and all lesser increments were allocated only to those vehicles who produced moments greater than that of an H-10 design vehicle.

3.1.2 Intra-incremental allocator

Bridges are designed to withstand overstress from heavy loaded vehicles, and fatigue from repeated usage. Fatigue damage is considered insignificant for most bridge types except steel bridges. The use of an intra-incremental allocator is to further discriminate between trucks in the same vehicle class/weight group but impose different number of load, therefore stress repetitions on the bridge structure. Furthermore, fatigue being a consequence of repeated loading is indirectly accounted for in the intra-incremental cost allocation process where some measure of usage of the facility is applied.

Annual vehicle miles traveled (VMT) is currently used as the intra-incremental cost allocator, being considered the best indicator for relative use. For bridges however, the loading event governing design is the occurrence two or more heavy trucks on the bridge simultaneously and not based on axle or vehicle repetitions. An argument against VMT as a suitable intra-allocator for bridges is that it does not represent differential load effects on the bridge structure. For example, for the same design increment and functional class, lighter vehicles with higher VMT values will be assigned higher cost responsibilities than heavy vehicles with higher moments and less VMT values.

The validity and sufficiency of VMT and other possible candidates such as, average traffic mix values or (average annual daily traffic, AADT); VMT weighted by a fatigue damage factor are examined. Some preliminary analysis to compare cost assignments based on these candidate intra-incremental allocators are presented in the following sections.

4.0 DAMAGE FACTOR CONCEPT

Hutchinson (1990) observed that studies of the relative damage created by different vehicle types tended to use the stress range concept, where the severity of damage depends on the number and magnitude of stress cycles. For example, Fisher (1977) suggested that the number of design fatigue cycles that a bridge might receive in its lifetime may be estimated from a relationship that is a function of $(GVW_i/GVW_s)^3$ i.e., the ratio of the gross weight of a given vehicle to that of the design vehicle. Heywood (1995) also noted that the relative total bridge damage induced by a vehicle (D_v) can be compared to that induced by a standard vehicle (D_{std}) based on the midspan deflections under action of load (Eqn. 1). Saklas et al. (1988) developed a similar relative damage function (Eqn. 2) for steel, prestressed and reinforced concrete bridges with the number of passes of truck as a variable. In all cases, the relation for the relative fatigue damage is a cubic function of the ratio of some measure of the total weight of the vehicles.

$$Relative\ total\ damage = \frac{D_v}{D_{std}} = \frac{\sum_{i=1}^{l} d^3_{v,i}}{\sum_{i=1}^{l} d^3_{std,i}} \qquad (1)$$

where: l - cycles that occur during passage of a vehicle;
$d_{v,i}$ - midspan deflections by vehicle;
D_v - damage by vehicle v.

$$d_j = \frac{n_j (F_{rj})^3}{\sum_{j=1}^{J} n_j (F_{rj})^3} \qquad (2)$$

d_j - relative damage factor;
n_j - number of passes of truck;
F_{rj} - stress range induced by truck j on a specific bridge detail;
j - number of truck cells (groups).

$$df = \left[\frac{GVW_i}{GVW_{des}}\right]^3 \quad (3)$$

df - damage factor;
GVW_i - gross weight of vehicle i;
GVW_{des} - gross weight of design vehicle;

Based on the above, a relative damage factor is assumed to be the cubic ratio gross vehicle weights (Eqn. 3) and is applied as a weighting factor to the intra-incremental allocator. This is based on the knowledge that fatigue life measured by the number of stress cycles is inversely related to the third power of the stress range. The stress range in turn is a linear function of the gross vehicle weight. Weighting accounts for the potential relative damage associated with usage. For example, vehicles with high gross weights but low usage will be assigned cost responsibilities based on both usage and potential relative damage. The proposed fatigue damage weighting concept is applied to the intra-increment allocators in the following sections.

5.0 ANALYSIS

This analysis explores the adequacy of four candidate intra-incremental allocators namely, (I) VMT; (ii) VMT weighted by fatigue damage factor; (iii) AADT; and (iv) AADT weighted by fatigue damage factor. The assignment of costs to 20 vehicle classes in 4 cost categories (new construction; replacement; major rehabilitation and minor repairs) are compared. This analysis considers only rural interstate highway functional class and HS-20 bridge design increment.

Sample 325 vehicles from a truck weight survey are used for the calculation of live load moments and VMT values by vehicle weight/class (Saklas, 1995). Except for cars and buses, AADT data for each vehicle weight/class cell was derived from the 1992 Truck Inventory and Usage Surgery (TIUS) database. The values for automobiles and buses were derived as percents of total AADT based on some previous estimates (Euritt et al., 1994). The AADT values are approximate and used for illustrative purposes only.

Live load moments and dynamic impacts are calculated for the mean bridge spans on rural interstate system i.e. 64ft for simple supported spans and 70 ft for continuous supported spans. These are determined for the 325 sample vehicles and HS-20 bridge design vehicle. The results are used to assign vehicles to the weight/class cells. This is achieved by comparing live load and impact values from

the design vehicles with those of the actual vehicles, as described earlier. The result is a 0 and 1 matrix where 1 indicates that vehicle weight/class exists on that functional class. A typical matrix is shown in Fig. 1.

The candidate intra-incremental allocators considered are each assigned to vehicle class/weight cells with "1". The proportions of the total in each cell of the matrix representing the cost share factors are then calculated. These factors are applied to the cost obligations of the various cost categories to arrive at the cost responsibilities for each vehicle class. The total allocations to each vehicle class is then determined.

Weight Group	Vehicle Classes																		
(kips)	UT	LT4	SU2	SU3	SU4	CS3	CS4	CS5	CS6	CT4	CT5	CT6	DS5	DS7	DS9	TRP	IBU	SBU	TBU
5	1	1	1	1						1									
10		1	1	1						1								1	
15		1	1	1	1	1	1			1	1						1	1	1
20			1	1	1	1	1	1	1	1	1		1				1	1	1
25			1	1	1	1	1	1	1	1	1	1	1				1	1	1
30			1	1	1	1	1	1	1	1	1	1	1	1			1		1
35			1	1	1	1	1	1	1	1	1	1	1	1	1	1	1		1
40			1	1	1	1	1	1	1	1	1	1	1	1	1	1	1		1
45			1	1	1	1	1	1	1	1	1	1	1	1	1	1	1		1
50			1	1	1	1	1	1	1	1	1	1	1	1	1	1	1		
55			1	1	1	1	1	1	1	1	1	1	1	1	1	1			
60			1	1	1	1	1	1	1	1	1	1	1	1	1	1			
65			1	1	1	1	1	1	1	1	1	1	1	1	1	1			
70				1	1	1	1	1	1	1	1	1	1	1	1	1			
75				1	1	1	1	1	1	1	1	1	1	1	1	1			
80				1	1	1	1	1	1	1	1	1	1	1	1	1			
85				1	1	1	1	1	1	1	1	1	1	1	1	1			
90				1	1	1	1	1	1	1	1	1	1	1	1	1			
95				1	1	1	1	1	1		1	1	1	1	1	1			
100					1			1	1		1	1	1	1	1	1			
105								1	1		1	1	1	1	1	1			
110								1	1		1	1	1	1	1	1			
115								1	1		1	1	1	1	1	1			
120								1	1		1	1	1	1	1	1			
125								1	1		1	1	1	1	1	1			
130								1	1		1	1	1	1	1	1			
135								1	1		1	1	1	1	1	1			
140								1	1			1	1	1	1	1			
145									1			1		1	1	1			
150									1					1	1	1			

NOTE: The number "1" indicates, vehicle exists, i.e., it operates on this funtional class

Source: Saklas (1995).

Fig. 1. Vehicles Assigned to HS-20 Design Increment for Simple Support Bridges on Rural Interstate.

6.0 DISCUSSION OF RESULTS

Table 1 shows the total cost responsibilities of groups of vehicle classes for rural interstate system and HS-20 bridge. The discussion is limited to the relative validity of the four intra-incremental cost allocators. The values obtained in the analysis are for illustrative purposes. The following observations are made:

- Based on VMT, automobiles are assigned a very large portion of the bridge costs. This directly reflects their share of the VMT. The results suggest that light vehicles (therefore low bridge load effects) with high VMT values are assigned higher costs responsibilities relative to heavier vehicles with low VMT values. For example, the results in Table 1 show that the cost responsibility of automobiles is about five times that of five and six axle tractor semitrailers, even though the potential fatigue damage of these trucks measured by the damage factor is more than 70 times that of automobiles. The table also shows that the cost responsibility of straight trucks, known to generate high moments particularly on short span bridges, is about a third of automobiles. Earlier findings, Saklas et al. (1995) indicate cost assignment based on VMT is directly proportional to the VMT shares.

- Cost assignments using the damage factor weighted VMT intra-incremental allocator provides a better reflection of both usage and relative damage. Compared to the unweighted VMT allocation, applying the damage factor results in significant changes in the percent costs assigned to certain vehicle classes particularly, automobiles and vehicle combinations (i.e. tractor semitrailers, doubles and triples). For example, five and six axle tractor semitrailers are responsible for over 80% of bridge costs (within the limitations of the analysis) while cars are charged less than 1% of total bridge costs.

  The VMT weighted allocation shows an improvement over the unweighted VMT allocator. Nonetheless, the question of whether VMT is a good indicator of relative use of bridge structures and cost discriminator remains.

- AADT is an indicator of the traffic mix and for the purposes of bridge cost allocation, it shows the number in each vehicle weight/class cell. It is traditionally used as an indication of load repetitions in pavement design. As an intra-incremental cost allocator, it can serve the same purpose, i.e., indicate relative usage of the highway facility. Like VMT, it does not indicate the relative damage from the vehicles with different weights.

Comparing assignments based on AADT to VMT clearly shows that vehicles with high VMT are not necessarily the most common (i.e. high AADT). For example, while straight trucks are assigned only 3% of total bridge costs based on VMT, they are assigned 32% based on AADT. On the other hand, five and six axle tractor semitrailers are charged about 13% based on VMT and less than 10% based on AADT. It can said that though AADT might give a good indication of the expected number of repetitions of loading on a bridge structure, by itself, does not give an indication of the magnitude of potential damage.

- Damage factor weighted AADT allocator assigns over 99% of total bridge costs to trucks with straight trucks accounting for about 16%. The five and six axle tractor semitrailers are responsible for most of the costs. This is attributed in part to the wider range of gross weights that these trucks are operated, and the relatively large population in the vehicle combination category.

Table 1. Summary of Allocations

Vehicle Type	Intra-incremental allocators				
	VMT (%)	VMT-DF (%)	AADT (%)	AADT-DF (%)	DF (%)
PASSENGER VEHICLES	**63.34**	**0.28**	**40.37**	**0.09**	**0.30**
Autos	63.09	0.19	40.12	0.07	0.00
Buses	0.25	0.09	0.25	0.02	0.30
TRUCKS	**36.65**	**99.49**	**59.63**	**99.91**	**99.70**
Single Units	20.83	1.82	44.99	16.40	5.13
Light trucks	17.80	0.01	12.92	0.20	0.00
Straight trucks	3.03	1.81	32.07	16.20	5.13
Combinations	15.82	97.67	14.64	83.51	94.57
Tractor semis (3 & 4 axles)	1.29	1.98	2.95	8.06	4.19
Tractor semis (5 & 6 axles)	12.56	82.86	9.71	66.99	22.15
Truck trailer (4+ axles)	1.11	4.56	1.32	3.29	21.00
Doubles	0.81	7.23	0.65	5.08	34.69
Triples	0.04	1.04	0.01	0.09	12.55
TOTAL	100.00	100.00	100.00	100.00	100.00

This allocator seems appealing because it combines the number and magnitude of expected stress cycles based on the vehicle class, vehicle weight, bridge support type and functional highway type to determine the cost responsibility. Unlike the three other allocators, three and four tractor semitrailers are assigned a fair portion of total costs.

7.0 CONCLUSIONS

The complexities and improvements to bridge cost allocation process are presented. VMT as an intra-incremental cost allocator for bridges does not assign costs fairly on the basis of the potential relative damage. AADT parameter reflects the percent distributions among the vehicle categories but like VMT, does not indicate the relative damage from the vehicles with different weights. Damage factor weighted AADT is an appealing intra-incremental cost allocator because it considers, the number and magnitude of expected stress cycles based on the vehicle class, vehicle weight, bridge support type and functional highway type to determine the cost responsibility. Nonetheless, the validity of this allocator needs to be further explored.

8.0 REFERENCES

Euritt M.A., Walton C.M., Goff Z.A. and Burke D. (1994) Texas Highway Cost Allocation Analysis and Estimates, 1933-1995. Research Report 1919-3F/1910-4F. Centre for Transportation Research, Univ. of Texas, Austin. Texas, USA.

Fisher J.W. (1977) Bridge Fatigue Design Guide. American Inst. Of Steel Constr. New York, NY.

Heywood R.J. (1995) Dynamic Bridge Testing: Short Span Bridges. Proc. OECD DIVINE Project Mid -Term Seminar. Sydney Australia. Paper #19.

Hutchinson B.G. (1990) Large Truck Properties and Highway Design Criteria. ASCE Transportation Eng. Journal. Vol. 116, No. 1 pp 1-22.

Saklas J.G. (1995) Highway Structures Cost Allocation. Draft Report. FHWA, Washington DC.

Saklas, J.G, Albrecht P. Sidani M and Schelling D.R. (1988) Bridge Fatigue Damage Cost Allocation: Executive Summary.

Performance-Based Truck Weight and Dimension Regulations

Fred P. Nix[1]
307653 Hockley Valley Rd
RR #1 Orangeville, Ont.
L9W 2Y8

John R. Billing[1]
Ontario Ministry of Transportation
1201 Wilson Ave
Downsview, Ont.
M3M 1J8

Michèle Delaquis[1]
17 Rosewarne Ave
Winnipeg, Manitoba
R2M 0V8

INTRODUCTION

Weight and dimension regulations are increasingly being devised using a concept that trucks should meet standards for performance, that is, they should be performance-based. The US Department of Transportation is reviewing domestic truck size and weight policies (for completion in December 1996), and is considering performance-based regulations. Since Canada has a reputation as a leader in this area and since there are significant proposals for change in Canadian and American regulations – the NAFTA negotiations are on-going – 'performance-based' requires careful examination. This paper examines the concept and describes developments in Canada. It also describes Saskatchewan policies as an important case study and it concludes with an overview of implementation approaches and policy issues.

There are three types of regulations: prescriptive, parametric and performance-based.[2] However, as considerations of performance lie behind all three, this last might better be labelled a 'pure' performance-based regulation. These can be illustrated using, as an example, offtracking and rear outswing performance measures that arise during low-speed turning. The prescriptive regulation states "no semitrailer with a length greater than . . . shall operate on the highway." This is rather clear and simple, but it obscures any original performance intention to ensure that vehicles can turn within the space provided by the road.

The parametric regulation attempts to control truck performance by setting values for and relationships between key parameters affecting performance measures. For example, the regulation might limit semitrailer swing radius, wheelbase, and effective rear overhang to ensure that offtracking and rear outswing are within the performance standard.

A 'pure' performance regulation addresses performance directly by setting standards for maximum offtracking and outswing during a particular test manoeuvre. This is equivalent to saying to the truck manufacturer or operator, "We don't care how long the vehicle is, or what any of its internal dimensions are, as long as it meets these performance standards."

Knowing where the boundaries are between weight and dimension and other regulations is a source of confusion in discussing performance-based truck regulations. Superficially, the boundaries appear clear: regulations having something to do with truck dimensions or weights. In practice, however, debates on the subject touch on other areas: safety (trucks must have brakes of a certain kind, trailers must have conspicuity markings); drivers (trucks over a certain size may require more highly-qualified drivers); traffic (large trucks may be limited to certain speed limits); and taxation (special fees for trucks of certain sizes).

In this paper, the focus is on performance-based regulations addressing one of four concerns: roadway geometry, pavements, bridges, or safety. Truck weights and dimensions have an impact on geometry (eg, underpass clearance or lane width and turning radius of an exit ramp), pavements (axle loads wear out pavements), bridges (bridges may fail under extreme loading, or lesser repeated loadings), and safety (the size and weight of a truck, and the way it is designed affect stability and/or controllability). There are overlaps and conflicts between these four concerns. For example, roadway geometric considerations lead to limits on maximum wheelbase, whereas consideration of dynamic stability sets limits on minimum wheelbase.

While the focus is on regulations dealing with the above four concerns, it is recognized that, performance-based or otherwise, these regulations have implications in other areas: 'pure' safety regulations, vehicle licensing, training and controlling drivers, traffic regulations, taxation; etc. In the following discussion it is important to remember that regulations in one area have broad ramifications for regulations in other areas.

BACKGROUND ON PERFORMANCE-BASED REGULATIONS

The concept of performance-based regulations is not new. As trucks became increasingly numerous in the early part of the century, it was necessary to put limits on gross weight and tire loads. Maine and Massachusetts were the first states to control truck weight in 1913; Ontario passed a "Load of Vehicles Act" in 1916. In 1918, American highway officials met in Chicago and agreed on a maximum truck weight of 28,000 lb with a tire load limit of 800 lb/in.[3] These were then used as design limits for roads. Canadian provinces probably either attended this meeting or were strongly influenced by the outcome. The early limits on tire loads (solid tires) appear to have been arbitrary because the understanding of truck performance relative to pavements was not well developed. However, even at this stage, there was a concept of truck performance (how total weight or tire load affected roads) incorporated in the regulations, albeit generally in prescriptive or parametric form.

There is an important point revealed in this history lesson. Regulations were performance-based to the extent they controlled tire loads or axle loads to protect road surfaces, even if the limits were arbitrary. Regulations became more performance-based when they were built on a formal model which related static axle loads to pavement performance developed from road tests in the 1950's. But, regulations were still not as performance-based as they could be, as they did not (for instance) relate characteristics

of different suspension systems to pavement performance, nor did they consider user fees as a means of controlling performance (higher fees for heavier axles). The concept of performance-based is relative.

A further move to performance-based regulation was the 1974 US federal adoption of a bridge formula (originally recommended in 1964). Prior to this, the recommended practice specified a maximum truck weight for a minimum distance between the first and last axle of the truck. Spreading the weight protected the bridges. The bridge formula did the same, but allowed more flexibility. Now a large number of axle spacings were possible, each resulting in different truck or axle group weights, so that operators could configure vehicles more specifically to their operational needs.[4]

Ontario also moved to a more performance-driven approach in 1970, four years before the US Congress, when it adopted a formula in its regulations for axle group and gross weight control. Surveys in the late 1960's found many overloaded trucks, but analysis also found significant bridge capacity not considered in bridge design and evaluation procedures. The Ontario Bridge Formula was developed to evaluate the effect of multiple axles on simple span structures. This 1970 regulation allowed any arrangement of axles, simply controlling their weight in accordance with the formula. Enforcement proved difficult, so in 1978 it was replaced with a series of five tables for axle group weights and 29 tables for gross vehicle weight. This regulation, either as a formula or a series of tables, was the first Canadian example of a 'pure' performance type of regulation.

As a result of this regulation, allowable weights increased to 63,500 kg in Ontario, putting pressure on other provinces to modify their regulations. A national study showed that, as in Ontario, bridges could accommodate heavier trucks as long as the distribution of the load was controlled. In the 1970's, the federal government made money available to the Prairie provinces and Atlantic Canada to upgrade pavements so that they too could increase axle and gross weights. The difficulty, at this juncture, was that regulations in each region were quite different from those in Ontario. Other provinces were not prepared to accept the axle arrangements allowed in Ontario, particularly the widespread use of liftable axles. As a result, the Canroad study was initiated in 1984 to address questions of axle configuration and load on pavement performance and truck stability and control.[5] Eight measures of stability and control were considered: rollover threshold; high-speed offtracking; handling (understeer coefficient); load transfer ratio; transient high-speed offtracking; low-speed offtracking; outswing; and friction demand in a tight turn. These all have some relationship to the potential for various accidents.

The Canroad study, and subsequent development of the 1988 interprovincial Memorandum of Understanding (the MoU), were milestones in the evolution of weight and dimension regulations. For the first time, anywhere in the world, bridge capacity, pavements, and stability/control factors were considered together in the development of regulations. Calls for a similar coordinated effort had been made in the United States since at least 1981[6]: the 1990 'Turner Truck' study accomplished this task.[7] Probably

the Turner Truck study's greatest contribution was its attempt to relate safety performance measures to accident rates. The impact performance standards have on accident rates is critical in judging the success of any policy.

The most significant development in the United States since the Turner Truck study has been the Department of Transportation's decision to conduct a Comprehensive Truck Size and Weight Study. Phase I of the study suggests as one option the possibility of controlling truck size and weight through performance regulations. While bridge/pavement performance is not ruled out, safety performance is the main concern underlying this suggestion.[8] In connection with this, a workshop on performance-based regulations was held in Ann Arbor on June 30, 1995.

Countries other than Canada and the United States have also been moving towards a more performance-based approach.[9] Some approaches differ from those in Canada. New Zealand adopted an 'envelope vehicle' approach similar to Canada's, but it took the performance-based idea one step further by allowing industry to develop its own designs. As long as trucks meet the performance standards, they may operate under permit. The European Union (EU), New Zealand and Australia all use a turning circle requirement that are closer to 'pure' performance standards than most Canadian regulations. The regulation in the EU is the most comprehensive, as it defines both an outer and an inner radius. The EU also attempts to take pavement-performance regulations one step further by giving a weight bonus for air or equivalent-to-air suspensions. This is done to encourage pavement-friendly vehicles.

While considering a truck's performance is not entirely new, there has been a noticeable movement towards a greater consideration of performance in developing weight and dimension regulations. This is particularly evident as regulations have advanced from the entirely prescriptive to the parametric and, more recently, to the first tentative attempts at 'pure' performance-based.

THE DEGREE TO WHICH CANADIAN REGULATIONS ARE PERFORMANCE-BASED

Canroad, the 1988 MoU and subsequent additions are considered to have advanced the concept of performance-based regulation significantly. Since many regulations are now under review and since there are misunderstandings about the development of Canadian regulations, it is important examine the Canadian case carefully.

In terms of roadway geometry, a shift to more performance-based regulations implies changing dimensional limits of trucks for different road classes. Some American states do this by classifying roads according to kingpin limits; other states classify roads according to offtracking characteristics.[10] Specifying dimensional limits by road class is not too common in Canada, although some provinces do allow, by permit, such things as overheight trucks on a regular basis as long as they have pre-checked the routes to be used. Other provinces use road classes to restrict longer combinations allowed under the MoU to certain roads.

In terms of pavements, it is difficult to know what 'performance-based' means until engineers can agree on a model of pavement performance. It can be noted, however, that whatever model is used, there is a relationship between axle loads and pavement wear. Canada's MoU set allowable axle group loads so that pavement impacts from a single passage by a single, tandem, or tridem axle at maximum load all would be about the same. The MoU also constrains allowable loads on the narrower tridem spreads based on bridge considerations, and contains an important load-sharing condition for axle groups, though this is rarely enforced. In most provinces, regulations also recognize different classes of roads: stronger or weaker pavements with heavier or lighter axle loads as the case may be. But, in general, Canadian regulations do not go much further. They do not encourage pavement-friendly suspensions or relate axle loads (pavement wear) to road taxes.

In terms of bridges, Canadian regulations are about as performance-based as they can be given the bridge designs used and the margins of safety built into the current truck regulations; this is true whether the regulations are of the bridge formula type in Ontario, the minimum interaxle spacing rules of the interprovincial agreement, or the more complex spacing requirements and weight penalties of Québec and several other provinces.

In terms of safety – and this is where Canadian regulations have earned their reputation as performance-based - there are several misunderstandings. First, the MoU is not 'the' Canadian regulation; it is simply an agreement describing trucks each province allows on a designated highway system. The four western provinces each immediately adopted the MoU into their regulations, so now, in practical terms, the trucks of the MoU are the trucks of these provinces. However, Ontario and provinces to the east initially merely accommodated the trucks of the agreement, while continuing with their existing regulations. Trucks under existing eastern regulations move an estimated 27 to 31 percent of total tonnes of truck freight on trucks with liftable axles.[11] Liftable axles are not allowed under the MoU. The point is that, in Eastern Canada, a significant amount of trucking occurs on trucks that would not pass the benchmark performance standards used for the MoU.

Second, trucks allowed under the MoU were never actually checked for their levels of performance. The Canroad study compared various generic vehicles for levels of performance. Once this was done, and a number of generic vehicles selected, a committee developed detailed specifications (inter-axle spacings, hitch offsets, load bonuses for certain trucks judged to be more stable, etc) on a prescriptive or parametric basis so that the trucks would be likely to meet the performance standards most of the time.

Third, there is a misunderstanding about the concept of an envelope vehicle. Regulations developed this way can check a vehicle for compliance with safety-performance levels in a general manner. What they cannot do is guarantee that, under all actual operational conditions, a particular truck will always meet these levels. For example, a configuration with specified dimensions and other characteristics can be

checked for its rollover threshold, with the critical case usually being a load that reaches full legal height at the same time as full legal weight. Unfortunately any operator can do something foolish – load the steel plate on top of the styrofoam packing – that has the effect of increasing the height of the centre of gravity of the payload and thereby lowering the rollover threshold to something less than the performance standard.

Fourth, there is a misunderstanding about the extent to which weight and dimension regulations can actually bring about a large-scale safety benefit. Accident rates for the truck fleet as a whole change only marginally under different weight and dimension regimes. The largest safety benefit actually occurs when a policy allows relatively large loads and thereby reduces the amount of truck travel.

Fifth, rigorously imposing the configuration restrictions that arise from safety performance standards does, perhaps fortuitously, seem to prevent the extremely bridge- and pavement-aggressive truck configurations that have evolved in places like Ontario and Michigan. The regulations in these jurisdictions are rather narrowly based on bridge and pavement performance criteria.

All of the preceding points about geometry, pavements, bridges, and safety make it sound as if Canada's regulations do not rate highly on a performance-based scale. In fact, this is misleading. While all the above points are true, (though not generally recognized when Canada's regulations are discussed in international forums), it is also true that the Canroad study and subsequent series of agreements immediately shifted the truck fleet more in the direction of performance-based than had previously been the case. Canroad has had a slow, incremental impact on changing regulations in various provinces: New Brunswick's new regulations are a 'vigorous' move towards the principles contained in the MoU; Québec has recently made changes encouraging more desirable axle types and discouraging less desirable double-trailer configurations; and Ontario adopted the exact terms of the MoU when adding a regulation allowing 25 metre doubles and 16.2 metre (53 ft) semitrailers. Canadian trucks tend now to be more stable and controllable then they would have been without the performance-based (albeit, parametric) regulations emanating from the Canroad study. Finally, and despite a range of inconsistencies in implementation of the agreement by the provinces, there are now standard vehicles that can run coast-to-coast across Canada in a range of gross weight classes up to 62,500 kg.

SASKATCHEWAN AS A CASE STUDY

Policies in Saskatchewan provide an interesting case study when analyzing options for implementing performance-based truck weight and dimension regulations. The province has pushed the concept of performance-based regulations further than others. First, after Saskatchewan signed the 1988 MoU, it fully adopted the envelope-vehicle approach. There were few exceptions for other vehicles, no long phase-in periods, and few more liberal allowances beyond the MoU. In other words, Saskatchewan bought into a concept of envelope-vehicle regulations where the envelope is defined in part by a

truck's performance. Admittedly, it was easy for Saskatchewan to do this as it had fewer changes to make than others.

Second, Saskatchewan increases axle weights and some gross weights during the three winter months. There are qualifications and exceptions to this, but these are not important here. Considering that the object is to control trucks and pavement performance, the rationale is that frozen pavements are stronger than unfrozen ones. So, for almost the entire Saskatchewan road system, things like the tandem axle load are increased from 17,000 to 18,000 kg during the winter.[12]

Third, Saskatchewan's development of a performance-based approach becomes most apparent in the treatment of large trucks carrying divisible loads. Many jurisdictions will not even issue permits for divisible loads which, by definition, can be hauled by legal trucks. Saskatchewan does issue such permits, and in relative terms is probably the most active jurisdiction in North America. There are three divisible-load permit policies, for timber, bulk commodities, and long trucks.[13]

The development of the Timber Transportation policy is important as it explains how Saskatchewan became interested in performance-based regulations. Plans were made in the mid 1960's for a pulp mill in Prince Albert. One way to attract development was to allow log hauling in permit trucks at high weights. Permits were issued without any consideration of truck stability. However, an increasing problem with rollovers and further requests for expanded permit operations forced the province to rethink its approach. The result is that, today, there are about 250 carriers, representing 400 to 500 power units, hauling 3.5 million cubic metres of logs under permit on configurations that have been checked for safety performance. Each carrier must obtain an annual permit for each truck, and the province ensures the equipment used will be suitable by, for example, checking that the manufacturer's rating on the axles is actually up to the job of hauling the loads allowed.

The most common logging truck is a 6-axle tractor-semitrailer with a maximum gross weight of 62,500 kg. The MoU allows this truck at 46,500 kg and this is the maximum weight in non-permit operations on Saskatchewan primary highways. Other logging trucks include 7- and 8-axle B-trains, and a 7-axle tractor-semitrailer, operating at weights up to 75,000 kg. There are even plans for a 26.5 metre long 10-axle B-train at 99,000 kg – payload will exceed 70,000 kg. Even heavier trucks are being considered.

A unique feature is that these logging trucks do not operate just under a permit from the province. Their operators are also subject to a legal agreement between the province and the forest-products company controlling the operation. This puts the onus for enforcement on the forest-products company. The company can tell who is cheating by watching the weight of the incoming loads, and the same weights can be audited by the provincial staff. The forest-products company is anxious to maintain the agreement, as transportation costs for incoming logs are a large part of total costs. Big loads on permit trucks keep these costs down.

The performance aspect of the regulation is in the fine print of the "Truck Haul Operating Plan." Typically, this document covers every aspect of the operation. Weights and dimensions are specified in minute detail: what axle loads, on what roads, at what time of the year, and on and on. There are even details on how and when Central Tire Inflation devices are to be used, to ensure a more pavement-friendly operation. Another part of the Plan covers track width and load height. If axles with a track width of 2.44 metre are used, load height is limited to 3.4 metres, but if track width is 2.59 metres, load height may rise to 3.7 metres, both less than the general height limit of 4.15 metres. However, with a track width of 2.9 metres, load height may reach 4.6 metres. The non-permit width limit is 2.6 metres. These measures control the rollover threshold. As load height increases, so does the centre of gravity of the payload, but with wider track widths, the rollover threshold is held to acceptable levels. This height/width relationship is only used for old equipment. New equipment must be checked by a specialist who ensures that the proposed truck meets the stability performance standards on which the MoU is based.

Other parts of the Truck Haul Operating Plan cover: operating procedures (everything from how accidents are reported, to the qualifications of the driver, to the requirement that the tractor have a speed-recording device); vehicle specifications and standards (everything from the ordinary – truck must have fire extinguisher, to the more esoteric – a weight-to-horsepower requirement and a brake performance test). There is even a part of the Plan whereby the operator agrees to work with the province on research and development projects!

There is one final part of the timber policy: the forest-products company agrees to pay for increased road costs – say higher maintenance because of the use of heavier trucks. The agreement also requires that the province and the company calculate the total transportation cost saving resulting from the use of big trucks. They then split this savings. That is, the company 'keeps' a certain portion – in effect, the benefit from lower transportation costs. The balance goes into a special account, and when the province and company have identified a needed road project in the area, say a strengthened bridge or a new intersection, the forest-products company pays. The significance of this last point may not be obvious. In Canada, provincial governments operate under consolidated revenue, which does not allow specific taxes to be dedicated to particular expenditures. What Saskatchewan has done, by making the forest products company open a special account that can only be used for road projects, is to find a way around this constraint.

The Bulk Commodity policy[14], formally adopted by Saskatchewan in January 1980, is similar to the timber policy in its use of a legal agreement between the province and the shipper (rather than just a truck permit), careful control of all aspects of the operation, and infrastructure cost-recovery. Its origins, however, go back to the late 1960's and early 1970's when the International Minerals and Chemical Corporation (IMC) asked the province for help in getting potash to market. The company wanted to haul potash from Esterhazy to a Burlington Northern Railway loading facility at Northgate just across the North Dakota border. The threat of potash bleeding off the Canadian

railways system could be used as a bargaining chip in discussing rates with the Canadian railways. IMC was willing to pay any extra pavement costs if it was allowed to use big trucks. The province agreed.

When originally formulated, the policy allowed any bulk shipper moving 30,000 tons or more annually between two points in Saskatchewan to enter into an agreement that, in return for allowing the shipper to use extra large trucks, gave the province a series of controls over the operation. It also required the shipper to pay incremental road costs, calculated as the difference in ESALs per tonne on non-permit trucks and permit trucks multiplied by the pavement costs per ESAL, for each road section. Sometimes there are also bridge costs. Further, if the bulk haul was rail competitive, the shipper had to pay for the total pavement costs (total ESALs per tonne times pavement cost per ESAL). (*'ESAL' or 'equivalent standard axle load' is a measure of the impact an axle has on pavements.*)

Currently there are six bulk haul agreements. Potash is still moving from Esterhazy to Northgate, now in 10-axle stabilized A-trains operating at a gross weight up to 89,813 kg. There are several chemical bulk hauls to the uranium mines in the north, moving in 9-axle B-trains at 72,500 kg. In the south, fly ash, petroleum products, and crude oil are also moving in heavy B-trains.

The Long Truck[15] policy started in 1979, when two carriers requested permission to operate triples on four-lane divided highways. On January 30, 1982, after reviewing the operation of triples in Alberta, and conducting trials of its own, Saskatchewan introduced a policy allowing carriers to operate long trucks on designated roads.

The two safety performance measures of most concern were offtracking and trailer sway. Because of this concern over trailer sway, Saskatchewan became the first jurisdiction to encourage the C-dolly – a double draw bar converter dolly that eliminates the articulation point of the more common A-dolly. A great debate rages about the relative safety and other merits of A- and C-dollies. Some carriers loath them ("they're expensive to maintain and drivers have a hard time hooking up"), while other carriers and some drivers swear by them. The small numbers of doubles make it unlikely that there will ever be sufficient data to demonstrate a statistically significant difference in accident rates between trucks using A- and C-dollies.

There are now 25 companies in Saskatchewan with long truck permits, operating about 300 power units under these permits. There is nothing unusual about long trucks. In fact, five Canadian jurisdictions, 21 US jurisdictions and Mexico allow them. But again, Saskatchewan is different. Most jurisdictions recognize just three kinds of long trucks: Rocky Mountain doubles, turnpike doubles, and triples. In Saskatchewan, there are 48 designs falling into four broad classes: Rockies, turnpikes and triples as in other places, plus a Queen city triple, which has one long trailer and two short ones. The important point is that each of the designs is actually a carefully controlled combination of trailer length, hitching mechanism and gross weight, set to ensure they meet most of

the same standards used for the MoU. (*They are not required to meet the same offtracking standards.*)

Saskatchewan's regulations differ from others in the way the allowed weight of the long truck depends on both the hitching mechanism – B-trains are allowed more weight than C-trains and C-trains are allowed more than A-trains – and the lead trailer length. Lead trailer length is a factor because research has shown that rearward amplification decreases as trailer wheelbase increases. A Rocky Mountain double with a long lead trailer is more stable than one with a shorter lead trailer and, hence, can be allowed more weight. The rough rule of thumb is 1,000 kg in weight for every extra foot in lead trailer length.

The Queen city triple is operated by SLH Transport between Regina and Saskatoon and Regina and Swift Current.[16] The company conducted trials in 1993, and has been operating them ever since. At close to 38 metres, these are among the longest legal trucks in North America. The average weight is only 42,000 to 44,000 kg and the payload is only 22,000 to 23,000 kg – not a heavy truck or a heavy load in Saskatchewan. With a volume just under 194 cubic metres, this suggests a freight density of 100 kg/m^3 (*6.6 lbs/ft^3*).

IMPLEMENTATION APPROACHES AND POLICY ISSUES

There are three basic approaches to performance-based truck weight and dimension regulations: the envelope-vehicle, the permit, and the pure performance standard. The choices between these approaches have to be made on the basis of a number of policy issues.

The first approach is the envelope-vehicle, used by Canada in the 1988 MoU. It works by having an agency consider a truck design, check it for compliance with pre-determined performance levels, and, upon satisfaction that these are met, enshrine the design into regulation. This is a prescriptive and parametric approach, not a pure performance one. However, since performance is considered, many people label this as a performance-based approach. The advantage of this approach, from a regulatory perspective, is that everyone knows exactly what a legal truck is – the equipment manufacturer, the motor carrier and the enforcement officer. This is particularly helpful when vehicles are to be common to a number of jurisdictions. Canada, both in the MoU and several provincial regulations, actually uses diagrams of the trucks, showing all the approved loads and internal and overall dimensions that make the truck legal. This approach has disadvantages. The amount of regulatory detail can be quite large. Further, an envelope vehicle may not be the best vehicle for all situations, and may not even be suitable for some.

A permit approach is another model, not mutually exclusive of the previous. Trucks over a threshold size can be required to operate pursuant to a permit. This is the practice in the 27 North American jurisdictions allowing long trucks (LCVs as they are also known). The permit allows the regulator to control the routes, to require higher

standards (eg, more qualified drivers), to exact fees (to pay for infrastructure), and to achieve better performance levels. It is also easier to ensure the matching of actual loads with manufacturer's ratings of components of the truck under a permit system. One key feature is that the threat of revoking permits ensures a better level of compliance.

A third option is the pure performance standard – that is, a regulation solely on performance. This could be in the form of a motor vehicle standard – in Canada and the United States, federal governments regulate manufacturers with respect to certain new vehicle standards (eg, brakes, lighting). These motor vehicle standards could be extended to require that vehicles meet objective performance standards under specified test conditions. Alternatively, this pure performance approach could be implemented with new on-highway safety regulations making motor carriers responsible for ensuring compliance with the performance levels. Either way, the advantage of this approach is that regulations are aimed at the primary object of interest: rollover threshold, rearward amplification, offtracking, etc. Further, the regulation tackles an issue directly ("no vehicle shall operate with a rollover threshold of less than . . .").

This third approach also allows for flexibility on the part of vehicle manufacturers and motor carriers. They can experiment with different designs as long as they achieve the performance standards. However, as has been argued[17], there are drawbacks to a pure performance-based approach. For one thing, measures of performance relating to safety (the eight measures used in the Canroad study, plus, conceivably, another dozen) are extremely complex. Some can only be evaluated by computer simulation. Testing methods have been agreed upon in only a few cases.[18] Even the simpler concepts, such as low-speed offtracking, are probably impractical for on-road enforcement. Another drawback is that if the performance-based system of regulation is not complete, unforseen things can happen.[19] This has been the experience in Ontario where, in effect, a form of pavement performance and a fairly new approach to bridge performance were used to develop regulations. Unfortunately, the stability/control aspects of the trucks were not considered and there is evidence that some trucks in Ontario are not as stable as might be desired.[20] Furthermore, even the pavement/bridge performance aspects of the new trucks is not what had been envisaged because the regulations did not control liftable axles.[21]

There are a range of other possible approaches; perhaps even some combination of the previous three (envelope, permit, pure performance standard). For example, 'go anywhere' trucks could be allowed under the envelope-vehicle approach. Other local or regional trucks could be controlled by permit with respect to a whole series of things including safety performance. Still other trucks, developed by motor carriers or equipment manufacturers, could be allowed either under permit or a certification program, as long as they met a series of performance standards. Whatever the approach – any of the above three or some combination – there are a number of broad policy considerations.

First, there is the issue of the complexity and conflict that can arise when jurisdictions agree on principles but fail to deliver harmonious programs when they implement these principles into regulation. This happened in the US when states implemented the provisions of the 1982 Surface Transportation Assistance Act, and in Canada when provinces and territories implemented details of the MoU. There are still provinces trying to amend regulations to reflect what was agreed upon eight years ago. In both cases, implementation of measures intended to harmonize regulations arguably ended up with more differences at a greater level of detail. An alternative to this is to have jurisdictions regulate by reference to technical standards residing with a non-governmental body, a practice followed in other areas where common technical standards must be adopted.

Second, there is the issue of the impact any of these approaches have on enforcement mechanisms. Some argue that enforcement considerations favour the permit approach; others argue that enforcement considerations favour the envelope-vehicle approach. It is difficult to resolve this debate with the current information available. However, enforcement considerations are extremely important and may be the main reason a pure performance-based approach would be difficult to implement.

The third issue is that there is an inherent conflict in developing 'good' weight and dimension policies. On the one hand, long haul trucking, NAFTA, globalization and simplicity cry out for harmonization among the (approximate) 62 NAFTA jurisdictions, and even countries beyond NAFTA (ie, consider ISO containers). On the other hand, roads and a great deal of trucking are relatively local affairs. The Timber and Bulk Haul policies in Saskatchewan provide evidence that local truck designs work well and that local agencies, who are close to the scene, know best how to manage their affairs, particularly when they also are responsible for the roads.

A fourth policy issue, taxation, is controversial because powerful interests are ranged on either side of the question as to whether trucks "pay their fair share" or not. Without attempting to resolve this, there is an aspect of taxation peculiar to Canada which impinges on truck weight and dimension regulations. The problem is that the taxation of road users is divorced from the building, maintenance and regulation of roads. To see how this impacts on weight and dimension regulatory policy, one has only to recall the development of toll roads in the United States and the somewhat more imaginative size and weight policies developed by toll roads in comparison to other road agencies. There is an obvious link between weight and dimension policies and taxation when the same agency administers both. Again, the Saskatchewan examples are instructive as the only instances in Canada where there is an absolute link between truck sizes and road-user fees; there is also some assurance that the revenues so raised are actually spent on roads.

Fifth, an important policy issue is whether the levels of safety performance known today should be cast into something as rigid as a regulation – that is, a regulation defining the acceptable performance level in an envelope vehicle, a permit operation, or an actual performance standard regulation. An argument could be made that, given the

measurement difficulties and the uncertainties surrounding the technical issues, it may be preferable to leave performance measures as administrative decisions. Motor vehicle administrators then have more latitude to exercise judgement and make adjustments. As the Saskatchewan experience suggests, it would be difficult to negotiate trade-offs between various aspects of performance if performance levels were set too rigidly.

Finally, how important is it to modify weight and dimension regulations in a more performance-based (particularly safety performance-based) direction? There is no easy answer. On the one hand, since available statistics make it difficult to relate aspects of truck design to the accident rate, it is not possible to know with certainty how much safety can be achieved with the imposition of any particular performance standard. The Turner Truck study made the first good attempt at this. However, current accident causation models are not sufficiently robust to allow an adequate assessment of risk as it relates to many changes in safety performance levels – particularly where tradeoffs have to be made among performance attributes.[22] This even makes it difficult to know how much effort to put into weight and dimension programs in comparison, say, to improvements in driver training/qualification programs or motor carrier safety rating efforts.

On the other hand, political realities make it necessary for governments to respond to public perceptions. The public wants safer roads; the public perceives trucks to be part of the problem; and, clearly, weight and dimension regulations have something to do with large trucks. Given this, governments must have a safety component to their weight and dimension regulations (and all road policies). Performance-based weight and dimension regulations will not, in themselves, deliver startling improvements in safety. They are, however, a tool for the creation of more rational weight and dimension regulations.

NOTES & REFERENCES

1 Fred Nix is a consultant who, among other things, is currently working for the US DoT on the subject of performance-based size and weight regulations. He was also the 'scribe' for the Ann Arbor workshop. John Billing has been involved with the development of weight and dimension regulations since the days of the Canroad study. Among other things, his work on the safety performance of trucks and truck-trailers was used to broaden the agreement in 1991 to include these vehicles. He was also on the TRB committee overseeing the Turner study. Michèle Delaquis is a consultant who assisted in the work for the US DoT. She has also undertaken research on weight and dimension regulations as they relate to long truck permit operations and has just finished work on weight and dimension limits associated with the NAFTA negotiations.

2 This threefold classification scheme was provided by C.G.B. Mitchell, Transport Research Laboratory (retired), at the Ann Arbor workshop.

3 Transportation Research Board. 1990. *Truck Weight Limits, Issues and Options*. Special Report 225. Washington, DC. p 35.

4 Information on the US bridge formula is from TRB 225, pp 38-39. The recommended practice prior to the bridge formula was a 1946 AASHO policy. (TRB 225, p 36).

5 Canroad Transportation Research Corporation. 1986. *Vehicle Weights and Dimension Study: Technical Steering Committee Report*. RTAC. Ottawa.

6 United States Department of Transportation, Office of the Secretary of Transportation. 1981. *An Investigation of Truck Size and Weight Limits*. Report No. DOT-30-82-07. Washington, DC.

7 Transportation Research Board. 1990. *New Trucks for Greater Productivity and Less Road Wear*. Special Report 227. Washington, DC.

8 United States Department of Transportation, Federal Highway Administration. 1995. *Comprehensive Truck Size and Weight Study, Summary Report for Phase I--Synthesis of Truck Size and Weight (TS&W) Studies and Issues*. Washington, DC. p 18.

9 Maze, T.H., J.S. York & M.T. Scott. 1995. "Policy Issues for the Application of Performance-Based Standards to U.S. Truck Size and Weight Regulations," Iowa Transportation Center, Iowa State University. Ames, Iowa. A full report, with a review of the practices in 30 countries, is expected sometime in the spring of 1996.

10 California, for example, uses a series of kingpin-to-rear axle measurements to classify roads; Idaho, as another example, classify roads according to the amount of offtracking.

11 Nix, F.P., M. Boucher. 1990. *Economics of Liftable Axles, Final Report*. Vehicle Technology Office, Transportation Technology and Energy Branch, Ontario Ministry of Transportation. Toronto.

12 Several other jurisdictions have a winter bonus on axle weights, but these programs are more limited than Saskatchewan's. A second point to note is this: if pavements can take

higher loads during the winter, why not allow any weight on an axle? The reason for not doing this is that higher axle loads have an impact on aspects of safety performance and on bridges. A third point is that in other jurisdictions, the freeze/thaw cycle is such that a simple-minded carte blanche increase in axle loads is not possible.

13 Policies are changing as this is written, making it difficult to describe them. Nevertheless, current policies are still so different than anywhere else that they deserve mention. People interviewed at Saskatchewan Highways and Transportation are: Bernie Churko, Norm Burns, Greg Gilks, Peter Hurst, Ken Jarvis (manager of the Timber policy), Finlay MacLeod (manager of the Bulk Haul policy) and Don Gatzke (manager of the Long Truck policy).

14 Aspects of this description are based on Churko, B., P. Hurst. 1985. "Improving Productivity of Freight Transportation in Saskatchewan", *Proceedings*. CTRF, 20th annual meeting. Toronto. pp 314 - 333.

15 Aspects of this description are based on (i) Burns, N. 1983. "The Operation of Overlength Vehicles -- The Saskatchewan Experience," *SAE Technical Paper Series*, 831164. Warrendale, Pennslyvania; (ii) Nix, F.P. 1995. *Long Truck Activity in Canada*, Canadian Trucking Research Institute, Ottawa.

16 Aspects of this description are based on an interview with Dave Greschuk, safety supervisor at SLH in Regina. In addition, a visit was made to the SLH yard on January 30, 1996 to observe the operation of these C-train Queen City triples. In the three years the company has operated them, 600,000 to 700,000 kilometres has been driven with no reportable accidents and no significant reaction from other motorists.

17 Billing, J.R. 1995. "Implementation of Vehicle Weights and Dimension Regulations," Fourth International Symposium on Heavy Vehicle Weights and Dimensions. Ann Arbor, Michigan, June 25-29.

18 Only two test procedures, for roll thresholds and rearward amplification, are recognized by the Society of Automotive Engineers (SAE J2180 and SAE J2179).

19 Billing, J.R. 1992. "A Discussion of Performance Standards for Configurations of Heavy Trucks," *Heavy Vehicles and Roads: Technology, Safety and Policy*, Third International Symposium on Heavy Vehicle Weights and Dimensions. Cambridge UK. D. Cebon & C.G.B Mitchell (eds), Thomas Telford Services Ltd, London. pp 278-283.

20 Billing, J.R. 1993. "Configuration of Heavy Trucks with Liftable Axles," paper 932941, SAE Truck and Bus Meeting. Detroit. November 1993.

21 Billing, J.R. 1993. "Highway Considerations from the Use of Liftable Axles by Heavy Trucks," paper 932989, SAE Truck and Bus Meeting. Detroit. November 1993.

22 Aspects of this argument have benefited from a discussion with Joe Morris, TRB, National Research Council, Washington, DC.

The Role of Transportation in Canada–ASEAN Trade: Is it a Barrier to Trade?

David A. Saucy

Introduction

Canada and International Trade

Canada is the seventh largest trading nation in the world. Exports contribute more than $130 billion per year in revenue, or 25% of gross domestic product (GDP).[1] As a result of its successful trade record, Canada enjoys one of the highest standards of living in the world. However, the country now faces many challenges; it is plagued with a sizable national debt, high unemployment and decreasing consumer spending, all of which threaten to erode the foundation of prosperity enjoyed by Canadians for so long. This dilemma is, however, not limited to Canada; it affects most of the industrialized nations of the world. To seek solutions to these problems, national governments are now turning to each other for support, and through cooperative efforts hope to foster economic renewal for their respective countries by the end of this century. The globalization of the world's economies is well under way, as witnessed by two general trends: the emergence of economic free trade zones in North America (NAFTA), Europe (EC) and Southeast Asia (AFTA) and the growing emphasis on searching for compromises and resolutions to global trade conflicts through GATT.

Although Canada's trade success depends on favourable international relations, it has always followed a distinctive pattern whereby a majority of exports are destined for the U.S. market. Currently over $110 billion, or 75% of all manufactured goods designated for export, are shipped to the United States, the single largest consumer market in the world and Canada's most important trading partner. Domestic exporters benefit from this close relationship, confirmed by Canada's $24 billion trade surplus with the U.S.[2] However, it can be argued that a certain degree of complacency has set in, as Canadian exporters are often hesitant to venture beyond the safe confines of the North American market. This dependency will only intensify in the coming years as NAFTA welcomes Mexico into the North American free trade family, thereby creating the largest single trading bloc in the world and further reinforcing the tendency to stay close to home.

Canada–ASEAN Trade

The Association of Southeast Asian Nations (ASEAN) was established in 1967 and now comprises six member states: Brunei, Indonesia, Malaysia, the Philippines, Singapore and Thailand. In 1992, the ASEAN governments created the ASEAN Free Trade Area (AFTA) in order to reduce internal trade barriers and promote further economic integration; if these plans prove successful the ASEAN region will encompass a population of over 325 million people and a combined GNP of US $300 billion.[3] The economies of ASEAN have performed remarkably in the past decade with sustained annual growth rates of over 8%. Economists have estimated that by the end of the decade, the Asia-Pacific region will represent 60% of the world's population, 50% of the global

production and 40% of total consumption.[4] Although these figures might be slightly optimistic, they are none the less impressive.

Exhibit 1: Canada and ASEAN Key Economic Indicators (US$)

Country	Pop. (millions)	GNP per capita	GDP growth	Inflation CPI	Annual Exports (in billions)
Canada	27.8	$19,936	0.9%	1.8%	$ 130.3
Indonesia	188.2	$ 645	5.8%	9.5%	$ 33.9
Singapore	3.1	$ 15,200	7.1%	2.4%	$ 63.4
Malaysia	18.8	$ 2,065	8.1%	4.3%	$ 41.0
Thailand	58.5	$ 1,660	7.4%	4.5%	$ 33.2
Philippines	64.1	$ 835	-.02%	6.7%	$9.7
Brunei	0.3	$ 17,500	3.0%	2.5%	$ 2.6

Source: ***Asiaweek*****, July 7, 1993**

Even with ASEAN's impressive record of sustained economic growth, rapidly increasing consumer spending and favourable economic prospects, Canada's trade with Southeast Asia has been fairly anemic. In 1992, only 1% of total exports were destined for ASEAN (see Exhibit 2). During the same period, Canada accumulated a $948 million trade deficit with these countries.[5] Currently, Canada's largest export market in the region is Singapore, although it is ranked only twenty-first among nations in terms of total exports. While Canadian foreign investment in ASEAN stands at $2 billion, ASEAN investment in Canada exceeds $6 billion![6] Needless to say, Canada has lagged far behind its industrialized allies in creating inroads into the dynamic economies of Southeast Asia. While Mexico is being heralded as the next frontier, Southeast Asia continues to be overlooked. Kim Campbell, Canada's new prime minister, recently stated her desire to explore the possibility of extending certain NAFTA provisions to the ASEAN region. In addition, a new study is being commissioned by the members of ASEAN, with the cooperation of the Canadian government, on the possibility of promoting further integration between business and government in North America and Southeast Asia.[7] Whether these discussions will materialize into increased trade between Canada and the ASEAN countries remains to be seen.

Exhibit 2: Canada–ASEAN Trade Data (Thousands of CDN$, 1992)

Country	Rank	Exports*	Imports	Balance	% of CDN Exports
Indonesia	24	437,366	394,449	42,917	0.24
Singapore	21	330,893	644,727	(313,834)	0.26
Malaysia	26	234,501	598,108	(363,607)	0.20
Thailand	22	331,207	577,123	(245,916)	0.25
Philippines	31	206,492	276,273	(69,781)	0.13
Brunei	195	2,358	207	2,151	—
TOTAL		**$1,542,817**	**$2,490,887**	**($948,070)**	**1.08 %**

*** Includes re-exports**

Source: Statistics Canada, ***Domestic Exports by Country*****, 1992**

The primary exports of manufactured products to ASEAN is centred on Canada's industrial heartland. Exports from the prairie provinces consist primarily of cereals and fertilizers while wood products and wood derivatives remain British Columbia's chief export. Trade from the Atlantic provinces is fairly diversified yet very weak compared to the rest of the country. The two Canadian territories, the Yukon and Northwest Territories, do not currently engage in trade with Southeast Asia (see Exhibit 3).[8] Trade figures compiled by Statistics Canada are not always consistent with trade figures from other organizations; these discrepancies can sometimes significantly distort Canada's bilateral trade picture. For example, in 1992 Statistics Canada recorded exports to Indonesia at CDN $431,567,000 while PT Surveyor Indonesia indicates CDN $657,004,190 for the same time period. This difference alters the balance of trade by CDN $225,437,190 in Canada's favour. One of the possible explanations for these inconsistencies is that over 15% of Canadian exports to Indonesia are transshipped through Singapore or Batam Island and then re-exported to Indonesia, unrecorded by Statistics Canada.[9]

There are several valid explanations for Canada's delayed response in developing inroads into the growing consumer markets in the Far East. Canada cannot aspire to compete with countries such as Japan, South Korea, Taiwan, Hong Kong and China in the production of mass consumer goods. Domestic exporters are still searching for profitable ventures closer to home in the U.S. and Europe, relying on the extensive networks established over the years exporting to these markets. Southeast Asia, on the other hand, is further down on Canada's learning curve. For these reasons, the prospect of exporting across the Pacific cannot be justified by many firms considering the extra time and financial investment required to develop new markets in the region. Canada's export potential to ASEAN is also limited by the manufacturing capacity of its domestic producers. Currently, 40% of manufacturing output is exported, of which 75% is sold directly to the United States. Canadian exports enter the U.S. with little difficulty, benefiting from relatively low tariffs and few non-tariff trade barriers. In addition, the Canadian manufacturing belt, strategically situated in Ontario and Quebec, benefits from excellent transportation links to major U.S. markets in the northeastern, midwestern and western States. The remaining 25% of exports is traded to closer and more familiar markets in Europe as well as to Japan, leaving Southeast Asia with a very small piece of the export pie.

Although Canadian manufacturers cannot compete in the area of consumer products, certain value-added sectors promise rewarding futures for Canadian exporters. The Canadian government has identified key opportunity sectors: advanced technology industries (telecommunications, aerospace, agri-food, defence equipment), energy (oil and gas) and the environment.[10] These are all industries for which Canada has the expertise to remain competitive; these sectors offer the best hope for increasing export revenue to ASEAN. Traditional areas of export such as pulp and paper, fertilizers and cereals continue to make up a significant portion of Canadian trade, although these industries are declining in importance as other sources of raw materials are being discovered within Asia. Canada is also responding to the worldwide growth in specialized service industries such as finance, environmental management, oil and gas exploration and business and engineering consulting.

Exhibit 3: Canadian Exports by Region (thousands of CDN$, 1992)

Country	Maritimes	Quebec	Ontario	Prairies	B.C.
Indonesia	4,550	52,140	65,657	234,330	74,886
Singapore	3,711	110,026	130,531	29,222	45,754
Malaysia	8,851	38,712	40,063	118,769	24,428
Thailand	3,509	84,846	78,808	67,977	90,909
Philippines	3,740	30,864	29,155	60,886	80,980
Brunei	—	379	138	60	726
TOTAL	**24,361**	**316,967**	**344,352**	**511,244**	**317,683**
% Total Trade	**1%**	**21%**	**23%**	**34%**	**21%**

Source: Statistics Canada, *Domestic Exports by Country*, 1992

This discussion inevitably leads to the question of whether some other factors are influencing the flow of merchandised goods from Canada to ASEAN. Since transportation linkages are the ultimate facilitators of trade, they must be one of the first sources examined for potential trade impediments. The question then arises whether Canada's poor trade performance with Southeast Asia is a result of the perception of inadequate transportation linkages to these markets. How Canadian exporters and freight forwarders view the role of transportation in the export process is critical in determining whether these linkages pose a barrier to trade. If, on the other hand, transportation is not perceived as an obstacle to increased trade, then perhaps some other factors are causing Canadian exporters to shy away from Southeast Asia. The perception of cultural barriers and different business practices as well as Canada's unfamiliarity with Asian issues must also be explored in order to determine their overall significance in developing a comprehensive export strategy. Once these perceptions have been identified and examined in greater detail, recommendations can be made as to how Canadian exports can become more competitive in the dynamic markets of Southeast Asia.

Exhibit 4: Major Canadian Exports by Country in 1992

Indonesia
- Cereals
- Nuclear reactors, boilers, machinery
- Plastics and articles thereof
- Pulp of wood
- Salt; sulphur; earth & stone; plastering mat; lime and cement
- Machinery, boilers, mechanical appliances, engines, parts

Singapore
- Aircraft, spacecraft and parts
- Machinery, boilers, mechanical appliances, engines, parts
- Paper & paperboard
- Electrical machinery equipment parts
- Natural/cultured pearls, precious stones & metals, coin .

Malaysia
- Fertilizers
- Cereals
- Paper & paperboard
- Salt; sulphur; earth & stone; plastering mat; lime and cement
- Plastics and articles thereof

Thailand
- Machinery, boilers, mechanical appliances, engines, parts
- Salt; sulphur; earth & stone; plastering mat; lime and cement
- Pulp of wood/of other fibrous cellulosic material; waste etc.
- Iron and steel
- Organic chemicals

Philippines
- Ores, slag and ash
- Cereals
- Electrical machinery equipment parts
- Mineral fuels, oils and products of their distillation
- Iron and steel

Brunei
- Special classification (designed for re-export)
- Electrical machinery equipment parts
- Iron and steel
- Machinery, boilers, mechanical appliances, engines, parts
- Arms and ammunition

Source: Statistics Canada, *Domestic Exports by Country*, 1992

Objectives

The primary objective of this paper is to conduct exploratory research into the current state of the shipping links between Canada and the ASEAN countries of Indonesia, Singapore, Malaysia and Thailand in an attempt to uncover whether these linkages restrict the flow of Canadian exports to Southeast Asia.

Scope

Indonesia, Singapore, Malaysia and Thailand were selected for this project based on their performance as Canada's four principal trading partners in ASEAN. In order to achieve the objectives mentioned above, personal and telephone interviews with all prominent players in the international transportation chain, in Canada and ASEAN, were conducted in order to identify critical problems areas. A majority of the interviews in Canada were conducted in the Halifax/Dartmouth area and, as expected, the answers were more focussed, although not exclusively, on issues central to the Atlantic provinces. The companies interviewed were limited to the industries of major Canadian exports selected from the list in Exhibit 4.

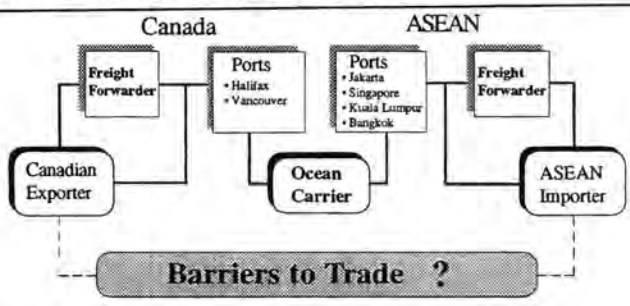
As the major mode of transportation to Indonesia, Singapore, Malaysia and Thailand, this paper emphasizes the use of containerized shipping outbound from Canada. Although air carriers may be used, ocean shipping is the primary focus of this paper.

Methodology

Research Design

This research was administered in three distinct steps. First, Canadian exporters, freight forwarders and port authorities were interviewed and their perspectives on the transportation links to Asian export markets were analyzed. Second, field research was conducted in Indonesia, Singapore, Malaysia and Thailand where Asian importers, freight forwarders, shipping lines and port authorities were interviewed in order to gather information on their assessment and perceptions of past and present dealings with Canadian companies and products. In this paper, the third step is to bring together the Canadian and Asian perspective in order to determine the weak links, if any, in the transportation process outbound from Canada (see Exhibit 5). The paper concludes whether the objectives of each player in the transportation process are compatible and, if not, what can be done from a Canadian perspective to reduce the potential problems. This in turn leads to an identification of practical solutions which might increase the flow of exports to the Southeast Asia thereby easing Canada's burgeoning trade deficit with the ASEAN members.

Exhibit 5: International Export Goods Flow



The Canadian Exporter Sample

Canadian exporters to either Indonesia, Singapore, Malaysia or Thailand were selected and interviewed. A list of Canadian exporters was generated from the *Business Opportunities Sourcing System (BOSS)* maintained by Industry and Science Canada. In order to simplify the search process for Canadian exporters to Southeast Asia, a list of exports was obtained from Statistics Canada (*Domestic Exports by Country, January to December 1992)* in order to identify the most common exports to each target country. The latter was then cross-referenced with the *BOSS* search to obtain the names of Canadian companies exporting to Southeast Asia. This search reveals that the key Canadian exports common to Indonesia, Singapore, Malaysia and Thailand are in the following industries:

- paper, paperboard and paper pulp;
- machinery, boilers, mechanical appliances and engine parts;
- electrical machinery equipment parts;
- aircraft, spacecraft, and parts;
- plastics and articles thereof.

Following this selection process, seven companies were contacted and an interview was scheduled. As this is an exploratory study, no attempt was made to identify the full set of companies exporting these commodities. Instead, a sample of firms in a variety of industries was selected. A majority of the exporters surveyed are located in the Halifax/Dartmouth area. They were selected for convenience as they could be easily reached for personal interviews and follow-ups. Exporters situated outside the Halifax/Dartmouth area were interviewed by telephone.

The firms surveyed were questioned on their exporting history to Southeast Asia and what types of problems they encountered, in Canada or Asia, which had a negative impact on the flow of trade. How much authority the exporter wishes to exercise in the selection of carrier services and how much control is relinquished to a freight forwarder was also examined. For those exporters which surrender the decision-making process to an intermediary, it was important to establish what level of service was expected and whether the current level of service offered by freight forwarders and ocean carriers was satisfactory.

Exporters were also questioned on the shipping terms quoted to potential customers (i.e. FOB, CIF). ASEAN importers were then questioned on their purchasing behaviour to determine whether the selling terms of Canadian products are a potential barrier to trade.

In addition, three non-ASEAN exporters were interviewed to establish the reasons why their products are not currently marketed in Southeast Asia. This last step was crucial in determining the perceptions held by Canadian companies with regard to Southeast Asia as a potential export market and to establish whether these perceptions were linked to the transportation network as a potential barrier to trade. Interviews with non-ASEAN exporters uncovered the existence of any preconceived notions Canadians hold on doing business with ASEAN importers. This led to the conclusion of whether perceived cultural differences produced a negative effect on Canada's exporting potential to Southeast Asia.

The Canadian Freight Forwarder Sample

Two international freight forwarders with operations in Canada were selected and interviewed at their Halifax premises. Freight forwarders were chosen because of the prominent role they play in determining the shipping requirements for Canadian exporters. Freight forwarders were asked to comment on the level of service offered to Canadian exporters by international shipping lines. Of particular importance was the availability of service to major ASEAN ports of call, shipping rates, and any other factors which would greatly influence the flow of trade outbound from Canada. Freight forwarders were also asked to describe in greater detail the relationship they hold with exporters and various shipping lines.

The Canadian Port Sample

Two Canadian container ports were selected for this research paper—Vancouver and Halifax. The Port of Vancouver is significant due to its proximity to the "booming" economies of the Pacific rim. As a result, the volume of trade passing through Vancouver has increased substantially in recent years. The Port of Halifax is the largest port of call in the Atlantic provinces and provides the Maritimes with a strategic link to major U.S. and European markets. In addition, a niche market is developing for carrier service from Halifax to ASEAN ports via the Suez Canal. The purpose of these visits was to witness first-hand the port infrastructure and container terminals, and to gather information on the services that are provided to international shipping lines. Although the Port of Montreal is the largest container port in Canada it was not included in this study due to the low volume of containerized shipments bound for Asia. The Port of Montreal serves northern

Europe and the Mediterranean as its primary market; 96% of its containerized shipments are destined for these regions.[11]

Informal discussion with representatives of Canada Ports Corporation were conducted at the Atlanta Intermodal Expo in April 1993, prior to the commencement of this project.

The ASEAN Port Sample

Interviews with port authorities and personal visits were pursued at four major ports of call in Southeast Asia: Tanjung Priok (Indonesia), Tanjong Pagar (Singapore), Port Klang (Malaysia) and Klong Toey (Thailand). The purpose of these visits was to identify the range of services offered by the respective ports and to determine whether the current infrastructure is able to accommodate the demands of international ocean carriers. In addition, personal observations were made as to how the flow of goods is handled once in the Asian port of call and whether the materials handling process is efficient. The assistance of freight forwarders with operations in ASEAN was imperative in determining their level of satisfaction with the overall operations of the various ASEAN port authorities in meeting the needs of importers in Southeast Asia. Promotional material was also collected from the respective port authorities in order to obtain basic statistical information on the infrastructure (e.g., berthing facilities, handling capacity, handling equipment), available services and planned future development. This process enables a brief comparative examination of the ASEAN ports.

The ASEAN Freight Forwarder Sample

Freight forwarders operating in Indonesia, Singapore, Malaysia and Thailand were interviewed. Commercial officers at the Canadian High Commission or Embassy were asked to assist in selecting appropriate companies to interview with preference given to freight forwarders with experience in clearing Canadian cargo through customs in Asian ports of call. The results of this search produced a total of ten freight forwarders for possible interviews: three in Indonesia and Thailand, and two in Singapore and Malaysia. Apart from general discussions on trade between Canada and Southeast Asia, the forwarding agents were asked to comment, based on their experience, on the level of efficiency, regulations, custom procedures and any other concerns they might have with regards to the overall functioning of the various ASEAN ports.

The ASEAN Importer Sample

Meetings with eight Asian importers of Canadian products were organized with the assistance from the Canadian High Commission in Singapore and Malaysia and from the Canadian Embassy and CIDA officials in Indonesia and Thailand. Commercial officers proved to be a valuable source of contacts as they are very often requested to link Canadian exporters with prospective Asian importers; these officers are also the main source of information on Canada for the Asian business community. The purpose of these interviews was to establish what the business practices of Asian importers are with respect to foreign goods. Asian importers were asked to comment on any past or present experiences they had with Canadian companies and what perceptions they hold about Canadian products in general. Differences in culture were also examined to determine how they may affect business transactions with Canadian companies and whether they represent a potential barrier to trade.

The Shipping Line Sample

Various international shipping lines operating between Asia and North America were interviewed to determine, from a shipping perspective, the level of services available

between Canada and major Southeast Asian ports. Five shipping companies were selected based on referrals from ASEAN freight forwarders and importers. The ocean carriers selected were OOCL, Sea-Land, N.Y.K./T.S.K. Lines, American President Lines (APL) and Mitsui/O.S.K. Lines. Representatives of these shipping lines were asked to comment on the level of proficiency of various Asian and Canadian ports. In addition, these representatives were in a position to comment on the competitive nature of the global shipping industry and what effect it has on the shipment of cargo originating in Canada. A review of the ocean carrier links between Canada and Southeast Asia is examined in order to determine Canada's competitive situation in comparison with other industrialized countries, namely the United States, Western Europe, Japan and Australia.

Other Sources of Information

Finally, a variety of discussions were held with experienced business people and consultants familiar with Asian markets, members of Canadian business associations, university professors in Asia and Canada, Canada–ASEAN country coordinators, representatives of the Vancouver Port Corporation in Southeast Asia and Canadian commercial officers in Asia.

Study Findings

Canadian Freight Forwarders

Three interrelated issues concerning containerized transportation to Asia emerged from the discussions with the two Canadian freight forwarders: distance, transit time and cost. The vast distance carriers must travel between Canada and Southeast Asia entails transit times anywhere from 20 to 40 days, depending on the port of origin and number of stops along the way. As a direct result, shipping costs tend to be very high for Canadian exports to Asia. These three elements create an environment which might pose a barrier to trade for certain products shipped by ocean freight. The range of potential Canadian exports to ASEAN is therefore greatly limited by the combined effect of lengthy transit time and high freight cost.

Shipping service from Canadian ports to Southeast Asia is perceived to be more than adequate, considering the low volume of exports to ASEAN. The availability of certain routes can sometimes lead to complications when shipping from the Port of Halifax. However, this is a relatively minor issue since Halifax benefits from good rail links to major ports in the northeastern United States, central and western Canada. The majority of the shipments bound for Asia are usually consolidated in larger Canadian or American ports, either Vancouver, Montreal, Chicago, Seattle or Tacoma. This highlights the most significant hindrance to exporting to Southeast Asia; due to the low volume of trade, freight forwarders sometimes encounter difficulty in consolidating a shipment. This inconvenience often leads to delays in shipping and added costs for the exporter.

Freight forwarders expressed concern over the relative lack of attention firms in the Maritime Provinces place on establishing a comprehensive transportation strategy. All too often these firms are unfamiliar with important shipping terminology and requirements such as Incoterms, bills of lading and letters of credit. Surprisingly, some do not even know of the existence of freight forwarders and the services they offer! In addition, many exporters based in the Maritimes are oblivious to education programs and information packages provided by the federal government to assist in the development of an export strategy. In the forwarders' view, some of this responsibility is reluctantly being shifted onto them.

The inexperience of certain companies in the Maritime Provinces creates tension in the relationship between exporter and freight forwarder. The latter feels burdened with the responsibility of educating the exporter who is often only interested in obtaining the

lowest shipping quote possible, not taking into account the intricacies involved in developing a comprehensive transportation strategy. This oversensitivity to price is also manifested in the unwillingness of many exporters to quote CIF (Cost, Insurance and Freight) prices to potential customers. Many exporters do not possess the skills required to calculate a CIF price, and as a result are forced to quote in FOB (Free On Board) terms or rely on the assistance of a freight forwarder. This is identified as a potential problem area when contemplating an export plan for Southeast Asia. Quoting CIF entails providing better service to the importer; Canadian products would be better received in a global marketplace if sold on these terms.

Exporting to Asia requires a different approach than dealing with more familiar markets in the U.S. and Europe. Freight forwarders emphasize the importance of "starting off right" when planning an export strategy for Asia. Canadian firms are advised to do their "homework"; this requires gathering information on the potential market as well as becoming more familiar with shipping terms and conditions and using proper financial instruments such as letters of credit. Freight forwarders believe that many exporters in the Maritimes feel intimidated. There is a clear perception that exporting to Asia entails sustaining additional risks, and therefore many companies are simply reluctant to look at ASEAN as a potential market. Freight forwarders are seen as able to reduce this risk as much as possible. However, freight forwarders stress that although they can assist in demystifying some preconceived notions about shipping to Asia, they do not have the ability to eliminate the perception of risk altogether.

The perception of adequate port facilities in ASEAN is generally positive with a few exceptions. Although no specific port was mentioned, some are perceived to possess less than efficient infrastructures, namely limited crane facilities. Poor inland road networks in some countries is another negative perception held by freight forwarders. The Port of Singapore on the other hand is held in high regard, consistent with its stature as one of the world's premier container ports.

Potential cultural barriers between Canada and ASEAN were also discussed. Canadian exporters have limited experience dealing with Asian companies; this translates into preconceived ideas about Asian business practices and culture. The issue of bribes in Indonesia and Malaysia is mentioned as an example of a local business custom foreign to many Canadian exporters. Although not considered a barrier to trade, it is just another illustration of the importance of learning more about the export market prior to committing to any business venture. In the freight forwarders' view, cultural barriers are more prevalent for Canadian firms which have little or no experience exporting to Asia.

In many ways, a freight forwarder is akin to a travel agent for exporters; this is the perception that freight forwarders have of the services they offer. Their duties entail obtaining the most favourable shipping rate and terms for the customer, as well as handling all necessary export documentation such as bills of lading and letters of credit. In addition, freight forwarders must follow up on the delivery of the cargo in order to provide a complete service to the client. In other words their role is seen as a transportation supervisor working on behalf of the local exporter. The role of educator is not a responsibility that freight forwarders are particularly content with, however some of these services must be offered in order to attract new customers and maintain a client base. Word of mouth is one of the most important sources of advertising for a freight forwarder. An exporter who is satisfied with the performance of a particular freight forwarder is more likely to use its service again or refer a business associate or friend.

Canadian Exporters

The response obtained from the seven Canadian companies followed a distinct pattern depending on the level of exporting experience and industry classification of the firm. Inexperienced companies had more preconceived notions about doing business with Asia.

Companies operating in high technology industries experienced no difficulty with the transportation linkages, however they faced a host of other restrictions, namely government regulations at home and abroad.

The companies interviewed did not experience or foresee any difficulty with the transportation linkages between Canada and Southeast Asia. However, two concerns were voiced by one export manager working for a small high-tech firm with little export experience who perceives that slow customs clearance in ASEAN ports is problematic; the issuance of bribes is identified as a way of accelerating the clearance process, thereby eliminating some of the "red tape." The other concern involves the perception of frequent delays when shipping containers from Canada to Asia.

Foreign governments are the largest purchasers of defence-related equipment. As a result, companies in the defence sector are more likely to ship FOB as the customer prefers to exercise control over the transportation of the products. Companies with a quality product in high demand are also in a position where they can afford to quote FOB prices. All the exporters interviewed were in one of the two positions mentioned above; consequently the prices quoted were FOB.

In the world of business, communication is very important. A time difference of up to 12 hours between Canada and Asia makes it inconvenient to arrange for phone conversations during business hours. The prominence of fax machines has alleviated this problem somewhat, although Indonesia was singled out as having poor communication facilities, as fax machines were perceived not to be widely available there.

There is a perception among smaller, inexperienced companies that dealing with Asia will expose them to added risks. The greatest risk mentioned is the potential for not receiving payment from the Asian importer. More experienced companies did not have these perceptions as they were more familiar with ways to insure themselves through the use of letters of credit and export insurance programs offered by the federal government.

Many firms suggested that they did not have the internal resources required in order to make well informed decisions regarding the selection of carriers and the preparation of all relevant export documentation. Apart from not having the internal resources, small and medium-size companies did not have the volume of trade, and therefore lacked the bargaining power to negotiate more favourable shipping rates. Freight forwarders were viewed as being able to provide a complete range of professional services, thereby saving the exporter a great deal of paperwork and time.

Firms were asked what their level of expectation was regarding carrier service performance. The answers obtained were unanimous: timing is crucial. The goods to be delivered have to be at the port of destination when promised. Companies operating in today's economic environment can ill afford to have idle resources tied up in inventory; the old adage "time is money" was often used to convey this message. Another important factor mentioned by many firms was the desire to have up-to-date information on the status and whereabouts of a particular container. With modern technology this is feasible; however it is limited to the few ports which employ such sophisticated information systems. Although the move towards greater port integration and automation is accelerating at a rapid pace in Canada and other western countries, the process is much slower in many developing ASEAN countries.

Although air freight has not been the focus of this paper, several exporters expressed their preference for using air transportation when feasible. Although shipping cargo by air is more expensive, the cost is partially offset by cheaper insurance premiums, lower inventory carrying costs and faster delivery times. For these reasons, many exporters are turning to air freight. The air link is also becoming more popular within the intermodal network as the use of specialized sea-air containers is becoming more prevalent in ASEAN.

Canadian Ports

The two container ports featured in this study, Halifax and Vancouver, account for over 51% of the containerized cargo tonnage handled in Canada.[12] Of the two ports, Vancouver is clearly the leader in terms of handling containerized shipments to growing markets in the Orient. As a result, the Port of Vancouver has experienced an average container growth rate of 11.4% during the period from 1987 to 1991, with over two-thirds of cargo traffic destined for the Asia Pacific region.[13] However, it faces stiff competition for Pacific-bound cargo from the ports of Tacoma and Seattle. Extensive promotion of the Port of Vancouver in North America and Asia is seen as one solution to retaining its competitive edge. The Port has a strategic alliance with Jardine Pacific Shipping Agencies, with representatives throughout Asia. The Port of Vancouver is also planning for future growth by constructing a new container terminal which will ensure that the port achieves its goal of handling up to 1.1 million TEUs by the year 2010.[14]

In the late 1980s, The Port of Halifax witnessed significant growth in the handling of containerized cargo, averaging 13.4% growth per annum in container tonnage throughput. However, by 1991, containerized tonnage had plummeted 21.2%. Several factors account for the Port of Halifax's declining market share, namely route rationalisation due to a flat Transatlantic market.[15]

The Port of Halifax is struggling to recoup from its dramatic fall in 1991. With funding provided by the Nova Scotia government, double-stack rail service is now being offered from Halifax to Montreal and Toronto. This new development, it is hoped, will partially offset the losses from the container business.[16]

The perception of adequate shipping service from major Canadian ports is confirmed by publications such as *Canadian Sailings* and the *Directory of Ocean Shipping Services* published by the Atlantic Provinces Transportation Commission. Weekly service is offered from Halifax and Vancouver to all major ports in ASEAN. The majority of the routes are offered by conference carriers sailing to Singapore or Hong Kong with feeder service to Indonesia, Malaysia and Thailand.

Exhibit 6: Canadian Container Ports – Key Figures (1991)

	Halifax	Vancouver
Containerized Tonnage Handled ('000 Tons)	2,900	3,400
TEUs Handled ('000 TEUs)	307	348
% Empty	12.4	12.2
% Containers Destined for Asia	3	78

Sources: ***Shipping in Canada*, Statistics Canada, 1991.**
Canada Ports Corporation, 1991.

Shipping Lines

The five international shipping lines interviewed acknowledge that the level of service from Canadian ports is very efficient at the present time, especially from the Port of

Vancouver. Many ships calling the northwestern ports of Seattle and Tacoma make a final stopover at the Port of Vancouver before heading across the Pacific. This ensures that Vancouver receives adequate service to major ASEAN ports and continues to be Canada's gateway to the orient. Canadian exporters benefit from the existence of an integrated transportation network within North America, and as a result are able to use American port facilities with few complications.

The most efficient service to ASEAN is provided from Vancouver to Singapore. Apart from being the largest container port in ASEAN, the port of Singapore acts as the major transshipment hub for all of Asia. Feeder vessels then provide service to other secondary ports such as Jakarta, Kuala Lumpur (Port Klang) and Bangkok. As a result, the transit times to all major ASEAN ports are not equal; service to the ports mentioned above results in an additional three to four days transit time by feeder vessel from Singapore.

Shipping lines showed some concern over the current trade imbalance between Canada and ASEAN. Ships most often carry full loads when sailing to North America, but remain half empty on the return trip home. This trend is also supported by a recent study which states that the trans-Pacific trade is over-tonnaged and that "in 1991 there is too much eastbound cargo, with ships light westbound".[17] This has an noticeable affect on the competitive position of the shipping lines and could lead to the eventual disruption in the current level of service offered to Canadian ports. This also creates equipment problems for the carriers in terms of containers shortages. One ocean carrier mentioned the difficulty in locating empty boxes in eastern Canada, emphasizing that the pool of containers is located in the United States. Boxes located in areas of low trade volume, such as the Maritime Provinces, must be trucked empty to the United States or, in the worst case scenario, shipped empty back to Asia.

The topic of shipping costs from Canada to Southeast Asia proved to be the most difficult issue on which to achieve a consensus among the shipping lines and freight forwarders interviewed. Shipping costs obviously vary greatly depending on the type of cargo being shipped and the equipment handling procedures requested. The price also depends on who the shipper is and its ability to negotiate favourable rates. The majority of the quotes obtained for the shipment of general containerized cargo indicated that the freight costs from Canadian or U.S. ports were very similar, usually within US$100 for a 20-foot container. In several cases freight rates from Canada were slightly cheaper. The reason cited was the intense regulatory environment which conference carriers are forced to operate under in the United States. Therefore, in terms of shipping costs to ASEAN, Canadian exporters are not placed in a disadvantageous position compared to exporters in the United States. However when compared to other industrialized countries, Canada faces greater competition. Freight rates from Europe to ASEAN are on average 15% to 20% cheaper than what Canadian exporters would pay, while shipments from Japan and Australia are one-third the price. Three representatives from the shipping lines suggested that inland transportation costs in Canada are very high, especially when shipping westward from the eastern or central part of the country.

Shipping lines did not express any concerns in regards to the port facilities in Canada. Traffic congestion in major ASEAN ports, however, caused problems for the shipping lines. The Port of Bangkok was singled out as having very slow service and frequent delays, largely due to excessive traffic in and out of the port. The recent completion of Laem Chabang, a deep-sea port 130 kilometers southeast of Bangkok is anticipated to alleviated some of this congestion and service is expected to improve in the near future. The Port of Singapore was unanimously appointed as the most efficient among all the ASEAN ports. This should not come as a surprise considering the modern facilities available at the port and its tremendous record for loading and unloading a ship, thereby decreasing the time ships must remain in port.

ASEAN Importers

The overall awareness of Canada and Canadian products in ASEAN is very weak. A majority of the ASEAN importers suggest that Canadians are not aggressive enough and are much too conservative in seeking new business ventures in ASEAN. This is in sharp contrast to other industrialized nations which have a strong presence in Southeast Asia such as the United States, Japan, Germany, the Netherlands, Denmark, Australia and Great Britain. The countries with aggressive marketing campaigns in Asia also benefit from extensive networks established over the years.

Canada is given very little hope of competing successfully in Southeast Asia in the area of consumer goods. The consumer products market is dominated by major Asian countries such as Japan, South Korea, Hong Kong, Taiwan and China. Canada's best chance therefore lies in the production of high value added capital goods where the quality of manufacturing usually takes precedence over price.

ASEAN importers expressed their preference for dealing with importers who quote in CIF terms. This saves the customer the burden of finding appropriate transportation and enables them to forecast more effectively the quantity of sales required to break even, therefore reducing the risk for the importer. Any exporters who do not quote under these terms place themselves at a competitive disadvantage within the ASEAN region.

One importer of Canadian products made an interesting comment—Canadian firms do not trust Asian importers, even after a relationship has been established for quite some time. This is manifested by the desire of exporters to deal only with terms of payment involving a letter of credit. Opening a letter of credit entails excessive paperwork and bank charges for the importer which would not exist if the terms were payment upon receipt of documents. No other importer interviewed, however, saw this as being a major inconvenience.

On the subject of cultural barriers, ASEAN importers emphasize that doing business in Southeast Asia is not all that different than doing business anywhere else in the world, but they stress the importance of establishing closer relationship with customers in ASEAN. This entails making an investment in travelling to the region and meeting with potential customers.

ASEAN Freight Forwarders

The eight freight forwarders interviewed all play a prominent role in Canada–ASEAN trade once the cargo arrives at the port of destination. Clearing customs can sometimes be an arduous task; it is therefore important to deal with a reputable freight forwarder with very close connections to the port authority. Poor facilities at some ASEAN ports can cause substantial delays in getting the product to the market. Unloading a vessel and clearing the cargo through customs should take between two and three days under normal circumstances. In some ports, namely Bangkok, Kuala Lumpur and Jakarta, this process can take up to seven days. The possibility of having goods tied up in bureaucratic "red tape" is unappealing to the customer and creates an atmosphere conducive to the request for bribes. One freight forwarder candidly stated that one cannot do business in Southeast Asia, and hope to be successful, without giving or receiving bribes. This common business practice is seriously frowned upon among the business community in Western countries, although by no means are they excluded from this activity. The difference is the degree of openness and level at which this form of corruption takes place in Asia. Workers in a position of even moderate authority use bribes as a means to supplement their low wages. Anyone who is in the position of delaying the flow of goods through the port system is subject to bribery. In the short run, the payment of bribes seems to have a facilitative effect on the flow of business, as it cuts down on excessive paperwork and

"red tape" which are notorious for slowing down the pace of business. In other words, freight forwarders, and ultimately the importer, look at bribes as just another cost of doing business, to be factored into the price of bringing a product to the market. Singapore is the notable exception where open forms of corruption are not as prevalent as in the other ASEAN countries.

ASEAN freight forwarders, according to their experience, did not see the transportation linkages between Canada and Southeast Asia as an explanation for the current low volume of trade between the two regions. It is pointed out that ASEAN importers do not look at Canada as being a large supplier of consumer products. ASEAN importers have a virtual unlimited supply of consumer goods produced within Asia at prices Canada could never compete with due to its socioeconomic structure.

ASEAN Ports

> In the age of multimodalism, a port must do more than just match cargo-handling facilities and tariffs with competitors.... A port is judged not just by cost and productivity, but also speed, reliability, multimodal links, information systems and distribution capabilities which have become important competition differentiators.[18]

The main purpose of conducting visits to the major ASEAN ports was to witness the overall efficiency of the port and to determine whether the infrastructure was able to accommodate international ocean carriers. The major problems encountered by freight forwarders at ports in Thailand, Indonesia and Malaysia were delays in unloading and then clearing the cargo through customs. Unnecessary delays in the movement of cargo will increase the inventory costs and ultimately increase the price of the goods. A recent study has attempted to place a monetary figure on the added cost of delays to containerized cargo (see Exhibit 7).

Exhibit 7: Added Inventory Cost of Delayed Cargo*

Vessel (TEU)	Cost per TEU ($US)	Cost per Vessel Day ($US)
600	13.7	8,220
1,800	13.7	24,660
3,000	13.7	41,100

*** based on an average load per container of 10 tons, average cargo value of $5,000 per ton and an average cost of capital of 10%**

Source: ***ASEAN Cargo Transport Study*****, 1992.**

Although the handling of cargo in some ASEAN ports can be characterized as slow, handling charges are not excessive. *The ASEAN Cargo Transport Study* concludes that "ASEAN port charges are at the low end of the international scale. The current charge structure cannot be regarded as an inhibiting factor in the movement of containers."[19]

Singapore is one of the world's busiest ports with over 70,000 vessel arrivals in 1991. In the same year, the port of Singapore handled over 6.1 million TEUs (Twenty-foot Equivalent Units) making it the world's leading container port (see Exhibit 8).[20]

Singapore continues to look at expansion as a means to retain its status as the primary transshipment hub for the Pacific; by 1995 the three current container terminals (Tanjong Pagar, Kepel and Brani) will reach a maximum capacity of 10 million TEUs, requiring the eventual addition of a fourth terminal.[21]

The container terminals at the port of Singapore are a model of efficiency. A vessel can be fully unloaded, the cargo cleared through customs and at the customers' doors within 24 hours. This high level of productivity is achieved in part by sophisticated management information systems. PORTNET enables shipping documentation to be handled more efficiently through on-line processing. Other systems include MAINS (Maritime Information System), CITOS (Computer Integrated Operations System) and CIMOS (Computer Integrated Marine Operations System) all of which are designed to integrate and streamline the flow of information and to direct and plan all operational moves in the container terminals.[22] Tremendous gains in efficiency and productivity have been made through EDI links with Bremen, Le Havre, Bangkok, Hong Kong and Seattle.[23]

Exhibit 8: ASEAN Container Ports at a Glance

Port	# Terminals	# Berths	Number of Gantry Cranes	Productivity (TEUs /Hour)	TEUs Handled (1991)
Singapore	3	10	22	28	6,190,000
Port Klang (Kuala Lumpur)	1	3	8	25	603,257
Klong Toey (Bangkok)	1	6	9	14	1,084,000*
Laem Chabang (Bangkok)	4	4	6	N/A	N/A
Tanjung Priok (Jakarta)	1	6	7	18	693,613*

* Estimate

Sources: *Malaysian Maritime SourceBook 1992*; *Port of Tanjung Priok Today*, 1990; Port Authority of Thailand, *Annual Report 1992*; *The Port Authority of Thailand: Growing with Thailand's Economy*, 1991; Port of Singapore Authority, *Annual Report* 1991; *Containerisation International Yearbook 1992*; *ASEAN Cargo Transport Study*, 1992

The Port of Singapore plays a pivotal role in the flow of trade within the ASEAN region. Thailand, Malaysia and Indonesia rely on the efficiency of Singapore as a transshipment point. A majority of the cargo destined for these three countries is siphoned through Singapore, arrives on megacarriers and is then loaded onto smaller feeder vessels.

Although the ports in Thailand, Malaysia and Indonesia have not achieved the levels of productivity associated with the port of Singapore, they are responding to the challenge. These countries realize that one of the keys to becoming a member of the NIC is the establishment of an efficient and modern port facility.

The Port of Bangkok is currently facing the greatest challenge of any ASEAN port; many infrastructural problems are linked to the tremendous growth of the Thai economy in the last decade. While the Port of Singapore is the embodiment of efficiency and order, Bangkok is ruled by intense traffic jams leading to what looks like mass confusion. The port infrastructure is currently stretched to its maximum capacity of 1.3 million TEUs. The port will be limited to this handling capacity for the next few years until new solutions can be found.

Even if the port can remedy its infrastructural problems, Klong Toey is still at the mercy of the City of Bangkok. Currently traffic congestion is the largest problem facing Bangkok. As a result, drastic measures had to be taken; all trucks are banned within the city limits during business hours. Therefore, even if the port can clean up its act, very little traffic will be able to circulate in and out of the port during the day, placing serious limitations on the future growth opportunities of the port. In order to defuse some of these problems, the Port Authority of Thailand constructed Laem Chabang, a deep-sea port, which officially opened in early 1991. Laem Chabang operates a feeder service to the heart of Bangkok via intermodal train or truck. It is still too early to tell whether Thailand's newest port can be considered a success, however it is definitely a step in the right direction. OOCL and American President Lines are betting on the success of Laem Chabang, launching a shared feeder service in order to provide improved connections to the major trade lanes of North America and intra-Asia.[24]

After Singapore, Port Klang is considered the most efficient port, although significant delays are still caused by the customs inspection process and bribery is prevalent. Port Klang benefits immensely from its proximity to Singapore, and in fact some Malaysian importers use Singapore as an unloading point and then use other intermodal carriers, either rail or truck, to enter the country. Malaysia also benefits from very good rail and road facilities providing access to markets throughout Thailand and Singapore.

Tanjung Priok in Indonesia faces many of the same problems encountered at Klong Toey and Port Klang, namely traffic congestion and slow customs clearance. Bribery is also prevalent at Indonesian ports, but it has been on the decline since S.G.S. (Société Générale de Surveillance) was implemented in 1985. S.G.S. has been very successful in Indonesia: "the result was an increase in efficiency, transparency and predictability, savings in time, and a reduction in questionable payments."[25] Under these guidelines, pre-inspection service is performed at the country of origin, prior to shipment arrival in Indonesia. Although S.G.S. has been relatively successful, the Indonesian government is planning to replace it in 1994 with a new system; Indonesia could see a return to slow inspection and unscrupulous behaviour of custom officials witnessed prior to 1985.

Other Findings

The most common complaint about Canadian exporters is that they are simply not aware of the potential of the Asian markets. This problem threatens to worsen if Canadians do not alter their focus away from the stagnant markets of the Western world towards more dynamic regions such as Southeast Asia. Canadians are seriously disadvantaged as a result of not having established extensive networks in ASEAN. Canadian exporters are perceived to be too conservative in seeking out new opportunities. Our image as a peaceful nation is conducive to good international relations, but it is of little advantage when making business deals.

Another common grievance expressed about Canadians, and Westerners in general, is that they are blinded by short-term thinking. Being successful in Southeast Asia requires a long-term perspective. However, Canada can still hope to play a

prominent role in the region in the years to come. Countries such as Vietnam, Laos and Burma hold promising futures; the economic climate in these countries suggests that now is the time to invest.

All too often Westerners' perspective of the Orient is centred on Japan and China, providing for a very distorted and single-minded view of Asia. The six countries of ASEAN offer a world of diverse cultural experiences. There is a tendency for exporters to view all the ASEAN countries as a single homogeneous market, not realizing that each country has its own unique blend of cultural and political values, cuisine, language, religion and business practices. As a result, products marketed in ASEAN must reflect this cultural diversity. This requires a considerable amount of investment on the part of the Canadian exporter and the establishment of proper contacts is essential.

The first source of information for firms wishing to export to ASEAN is the Canadian Embassy or High Commission. The primary function of commercial officers in these institutions is to assist in establishing contacts between the exporter and prospective clients. Some commercial officers, however, voiced their disappointment about the attitude of many Canadian businessmen and women who do not make the necessary effort to build a long-term relationship with clients. Once a trade mission has been completed, Canadian firms do not engage in any follow-up activities; this is a considered a serious mistake and shows disrespect towards the Asian hosts.

The topic of piracy in the South China Sea and Malacca Straits is attracting extensive media attention in Asia. The mention of the word "piracy" will send many chills down the spine of any ship captain venturing into the seas surrounding Southeast Asia. The issue of piracy, however, is of little concern to freight forwarders and major shipping lines. Pirate attacks are shrugged off as mere media publicity. In part this is due to the pattern of attacks—pirates do not plunder and pillage large container ships. The majority of aggressions are towards smaller feeder vessels operating between Southeast Asian ports. The number of piracy attacks does seem to be on the decline, although little is said about the seriousness of the crimes committed. In 1992, 83 attacks were reported worldwide as opposed to 107 such incidences in 1991.[26] However, it is believed that many attacks go unreported by the shipping lines. In addition the area of greatest concentration is shifting from the Malacca Straits to a triangle between China, the Philippines and Hong Kong, often referred to as the Hong Kong triangle.

Talks between government officials of Indonesia and Malaysia have resulted in an agreement to provide added patrolling in the affected areas. Of greater concern is the increasing traffic in the Malacca Straits, already the most congested water passage in the world. The threat of a collision poses a far greater security problem than an act of piracy. In any case, neither piracy nor high traffic congestion in the Malacca Straits is seriously hindering the flow of trade within Southeast Asia and shipping lines are not planning to alter the current route patterns.

Discussion

This study examined the transportation linkages for Canada to ASEAN trade with the intent of exploring whether transportation acts as a barrier to trade, or is perceived as a barrier by Canadian exporters and freight forwarders. The findings of this research indicate that the ocean carrier linkages between Canada and Southeast Asia represent a potential barrier to trade for certain Canadian exports, i.e., goods which are time sensitive with high value inventory costs. Consumer products which fall in this category may be uncompetitive for other reasons including high domestic wage rates. High inland transportation costs and distance to market make it virtually impossible for Canadian firms to compete successfully in Southeast Asia under these circumstances.

This study also indicates that Canadian exporters are not disadvantaged compared to their American counterparts in terms of ocean shipping costs. The majority of Canada's manufacturers reside in the industrial heartland, centred in Quebec and Ontario. The location of these producers results in high inland transportation costs which places them at a serious disadvantage. Therefore in order to minimize the impact of high transportation costs, Canadian firms must look at the production of high-technology, high-value added capital goods as the future growth sector for exports. Although transportation costs and transit time remain relatively high, their combined effect is not as significant on these range of goods, where quality of manufacturing is usually the key determining factor for remaining competitive in a global economy.

The Canadian companies surveyed in this study did not perceive the transportation linkages to be a significant hindrance to trade. In fact, when companies are planning an export strategy, the logistics of transportation are not a high priority. This is evident in a publication issued by Industry and Science Canada entitled *Exporting for Competitiveness – Ten Steps For Small Business*. According to this guide, the evaluation of transportation options is step number eight. Step three suggests that exporters familiarize themselves with shipping terminology while step five recommends that an export price be determined. Exporters are faced with a host of other potential obstacles including marketing strategy, export pricing, financing and government regulation. Once the issue of transportation arises, research has concluded that the responsibility is being passed on to freight forwarding agencies with greater frequency. It is therefore important that these intermediaries did not foresee any significant weak links in the transportation chain outbound from Canada although they had volume concerns; it is critical to expand the trade so that there is sufficient volume to consolidate.

The level of service offered by shipping lines to destinations in ASEAN is more than adequate. Combined with excellent access to all major U.S. ports through an integrated intermodal network, Canadian exporters and freight forwarders are provided with several viable alternatives when shipping containerized cargo to Southeast Asia. The port infrastructures in Canada do not pose any difficulties for Canadian exporters, freight forwarders and international shipping lines.

The Port of Singapore is one of the most efficient in the world and as a result it represents a tremendous asset to all of Southeast Asia, acting as a gateway to the other developing countries in the region. The improvements currently taking place in most ASEAN ports is mirroring the tremendous economic growth of the entire region. Ports in Thailand, Malaysia and Indonesia still face many challenges; however, the respective port authorities have already enacted plans for the further development of port facilities and are committed to achieving the levels of productivity associated with the Port of Singapore and other leading international ports by the year 2000. Although inadequate port infrastructure may create a barrier to trade, an adequate port infrastructure does not necessarily promote greater trade. Therefore, Canadian exporters should be more concerned with trying to reduce delays in clearing goods through customs at ASEAN ports, as these could result in payment delays from the importer. There seems to be a concerted effort on the part of the ASEAN governments to improve the delays in customs clearance, but it may be several years before any improvements are apparent. Although delays in the ports are an important consideration for any exporter, these problems are not unique to Canada. Any country wishing to export to ASEAN faces these same obstacles; Canadian exporters are, therefore, not placed at a competitive disadvantage.

Although this study's primary focus is the transportation linkages, it is only one piece of an intricate puzzle. Many countries in the world are very successful in Southeast Asia because of the extensive networks established over the years. Countries such as the Netherlands, Denmark, Germany, Japan and the United States are making great strides in the Far East. Their success is partly attributable to well-established links and the overall

attitude of exporters. Communication is a very important element in business and establishing proper networks is the foundation of any successful venture in Asia. There is a great deal of evidence which suggests that Canadians do not have this extensive web of networks, placing them at a serious disadvantage compared to exporters of other nationalities. The time difference between central Canada and Southeast Asia, along with poor communication facilities (e.g. faxing capabilities) in the lesser developed ASEAN nations, are problematic for exporters. These inconveniences cannot be considered a barrier to trade, though, as they do not place Canadian exporters in a disadvantageous position compared to exporters in other industrialized countries.

Since the transportation linkages are not perceived to be a barrier to trade by either the shippers or the freight forwarder, attitudinal factors must be examined. The negative perceptions associated with exporting to Asia represent a substantial obstacle for Canadian exporters. This study indicates that Canadian awareness of Asia is very poor, which probably acts as a barrier to trade. Negative perceptions may constitute a significant obstacle for Canadian exporters, seriously impeding their ability to make inroads into the dynamic markets of Southeast Asia.

As a result of the real barriers discussed above (e.g. market, tariff, communication, distance and time) and perceived barriers (e.g. increased risk), Canadian companies are continuing to explore opportunities closer to home in North America and Europe by taking advantage of the extensive networks and experiences built up over the years. Although Canadian companies are successful in these markets, economic indicators call into question whether these regions will continue to prosper. The question then arises whether Canada is selling itself short by not paying more attention to Southeast Asia as a potentially profitable market.

Recommendations

Exporters cannot realistically change some of the barriers that exist between Asia and Canada, namely problems associated with time, distance and inland transportation costs, so the main barriers which must be addressed are attitudinal. The negative perceptions held about exporting to Southeast Asia can only be overcome through education. The question is who should bear the brunt of this responsibility. The first cry for help is usually directed towards the federal government, and there is evidence that the federal government is making a concerted effort to educate domestic exporters about the tremendous potential in Southeast Asia. The Canadian government already spends more money on export promotion programs than any other government in the world.[27] Either these programs are not being properly communicated to exporters, or the latter are simply not interested in exploring new market opportunities.

The programs currently offered will only be successful with the full cooperation of management in Canadian firms. The market conditions in the Western world will eventually dictate whether Canadian exporters will continue to concentrate all their energy and resources in North American and European markets, or whether they need to seek out new frontiers. This latter proposition is, however, very risky as Canada's ability to compete in a global economy is intimately linked to the prosperity of a future generation of Canadians.

For the Federal and Provincial Governments

Educational programs targeted at exporters should try to break down some of the misconceptions about dealing with Asian importers. Although certain barriers are legitimate, others can be overcome with the proper education and experience. Promotional campaigns (e.g. Canada Export Month) and publications endorsed by External Affairs Canada are very informative and these efforts should continue. However, there is

evidence that there are still too many companies which do not benefit from these programs due to lack of awareness. One complaint voiced in the Maritimes and provinces west of Ontario is that too much of the governments' resources are concentrated on central Canada. This study has emphasized that although the shipping links do not pose a barrier to trade for exporters, inland transportation costs can be a serious hindrance, so many companies situated in central Canada experience great difficulty competing in Asia. Exporters situated in British Columbia and the Maritimes benefit from excellent shipping links to Asia. As a result, the federal and provincial governments should encourage firms in high-tech and capital goods industries to locate on the east or west coast of the country.

A great deal of research remains to be conducted in order to determine the full impact of transportation on the range of products for which Canada wishes to remain competitive. Possible areas for future studies include air cargo connections, specifically for high tech products, as well as bulk shipping for raw materials and project cargo.

For the Exporter

Ultimately the bulk of the responsibility lies with the Canadian exporter. The future prosperity of Canada rests in the hands of small and medium-size companies as they are becoming the largest source of employment and revenue for the country. Exporters should invest some time and money every year into looking for new markets. These investments need not be in the millions of dollars. Canadian exporters should make use of some of the contacts established by the federal government in ASEAN and already paid for by their tax dollars; a simple fax to a commercial officer at the Canadian Embassy or High Commission requesting information about a potential opportunity is a start in the right direction. Canadian businessmen and woman should educate themselves about the various customs of Asia where relationship building is very important. Canadians must also break away from the habit of always thinking with a short-term perspective. A trade mission to Asia may not yield profitable results immediately; however, those who have been successful in the past have also been the most persistent. The entire ASEAN region is currently thriving with profitable business opportunities not witnessed since the industrial revolution. The only difference is that with the pace of modern technology these opportunities disappear as fast as they arrive. The time for Canadians to invest in ASEAN has arrived.

> When written in Chinese, the word 'crisis' is composed of two characters—one represents danger, and the other represents opportunity.
>
> *John F. Kennedy*

Notes

1 *Canada and Trade*, January 1993
2 *Statistics Canada*, 1992
3 *Business Guide to ASEAN 1992*, External Affairs and International Trade Canada
4 *Business Guide to ASEAN 1992*, External Affairs and International Trade Canada
5 *Statistics Canada*, 1991
6 *Business Guide to ASEAN 1992*, External Affairs and International Trade Canada
7 *Globe and Mail*, July 29, 1993.
8 An important detail left uncovered in traditional trade figures is the growing importance of the service trade industry, a dynamic field for which Canada is very competitive on a global scale. Trade in services is impossible to measure with any degree of accuracy.
9 *Canadian-Indonesian Trade*, April 1993, p.2.
10 *Canada and Trade*, January 1993
11 *Canada Ports Corporation*, 1991, p.32.
12 *Statistics Canada*, Shipping in Canada, 1991, p.82.
13 *Statistics Canada*, Shipping in Canada, 1991, p.80, 82.
14 *Port Development International*, March 1993, p. C15.
15 *Statistics Canada*, Shipping in Canada, 1991, p.82.
16 *Statistics Canada*, Shipping in Canada, 1991, p.82.
17 *ASEAN Cargo Transport Study*, p. 24.
18 *Business Times*, August 10, 1992, Malaysian Transport Minister Datuk Seri
19 *ASEAN Cargo Transport Study*, p. 59.
20 *Port of Singapore Authority*, Annual Report 1991, p. 11.
21 *Port of Singapore Authority*, Annual Report 1991, p. 20.
22 *Port of Singapore Authority*, Annual Report 1991, p. 39-40.
23 *Port of Singapore Authority*, Annual Report 1991, p. 43.
24 *Cargonews Asia*, April 12, 1993, p. 29.
25 *Amcham Trade Expansion Committee Newsletter* (AMTEC), May 1993
26 *The Jakarta Post*, May 24, 1993, p.5.
27 U.S. Department of Commerce, as cited by Czinkota, M.R. and I.A, Ronkainen (1993), p.50.

References

Brooks, M.R. (1984), "An Alternative Theoretical Approach to the Evaluation of Liner Shipping, Part I. Situational Factors," *Maritime Policy and Management*, 11, 1, 35-43.

Brooks, M.R. (1985), "An Alternative Theoretical Approach to the Evaluation of Liner Shipping, Part II. Choice Criteria," *Maritime Policy and Management*, 12, 2, 145-155.

Brooks, M.R. (1990), "Ocean Carrier Selection Criteria in a New Environment," *Logistics and Transportation Review*, 26, 4, 339-56.

Brooks, M.R. (1991), "Assessment of the Ocean Carrier Decision Environment: A Longitudinal Study," *Journal of the Transportation Research Forum*, 31, 2, 219-29.

Brooks, M.R. (1993), "The Ocean Container Carrier Market—Is It Segmentable?" Presentation to the Transportation Research Forum, New York, October 12-14,1993.

Collison, F.M. (1984), "Market Segments for Marine Liner Service," *Transportation Journal*, Winter, 40-54.

Czinkota, M.R., I.A. Ronkainen (1993), *International Marketing*, Fort Worth, Texas: Dryden Press.

Davies, G.J. and C.E. Gunton (1983), "The Buying of Freight Services: The Implications for Marketers," *Quarterly Review of Marketing*, 8(3),1-10.

D'Este, G.M. and S. Meyrick (1992), "Carrier Selection in a RO/RO Ferry Trade: Part 1: Decision Factors and Attitudes," *Maritime Policy and Management*, 19, 2, 115-26.

D'Este, G.M. (1992), "Carrier Selection in a RO/RO Ferry Trade: Part 2: Conceptual Framework for the Decision Process," *Maritime Policy and Management*, 19, 2, 127-38.

Jerman, R.E., R.D. Anderson and J.A. Constantin (1978), "Shipper Versus Carrier Perceptions of Carrier Selection Variables," *International Journal of Physical Distribution and Materials Management*, 9, 1,29-38.

LaLonde, B.J. and M.C. Cooper (1989), *Partnerships in Providing Customer Service: A Third-Party Perspective* (Council of Logistics Management).

McGinnis, M.A. (1978), "Segmenting Freight Markets," *Transportation Journal*, Vol. 18, 58-68.

Other Sources

Amcham Trade Expansion Committee *Newsletter*, May 1993.

Asia Week

Atlantic Provinces Transportation Commission, *Directory of Ocean Shipping Services*, April 1993.

Business Times

Canada Ports Corporation, *Towards a Canadian Intermodal System: Recommendations for Change*, 1991.

Canadian Sailings

Cargonews Asia

CPCS Ltd., *ASEAN Cargo Transport Study*, Montreal, Canada, June 1992.

External Affairs and International Trade Canada, *Safe Stowage*, Ottawa: Supply and Services.

External Affairs and International Trade Canada, *The Exporter's Guide to Transportation*, Ottawa: Supply and Services.

External Affairs and International Trade Canada, *Business Guide to ASEAN.*

Globe and Mail

Jakarta Post

Industry and Science Canada, *Exporting for Competitiveness- Ten Steps For Small Business* , Ottawa: Supply and Services.

International Chamber of Commerce, *Incoterms 1990,* Paris.

Port Development International

Port of Singapore Authority, Annual Report 1991.

Port of Tanjung Priok Today, 1990.

Resource Development Consultants (REDECON), *Canadian-Indonesian Trade Statistics Study*, April 1993.

Société Générale de Surveillance, *Preshipment Inspection of Imports: Guidelines for Exporters*, 1988.

Statistics Canada, *Domestic Exports by Country*, 1992.

Statistics Canada, *Shipping in Canada*, 1991.

Straits Shipper

Straits Shipper, Malaysian Maritime SourceBook 1992.

Straits Shipper, Malaysian Transport Review 1993.

The Port Authority of Thailand, Annual Report 1992.

The Port Authority of Thailand: Growing with Thailand's Economy, 1991.

World Trade

Final Note

This research paper is made possible by an Academic Travel Grant funded by the **Canada–ASEAN Centre**. Other noteworthy sponsors of this research paper are: **The Centre for International Business Studies (CIBS)** at Dalhousie University and the **Environmental Management Development in Indonesia (EMDI)** Project, Dalhousie University. I would like to acknowledge the contribution of **Dr. Mary Brooks**, Professor and Director of the Centre for International Business Studies at the School of Business Administration, Dalhousie University, in appreciation for her guidance and time commitment to this project.

UNCERTAINTIES IN TRUCK REGULATORY POLICY AND HIGHWAY INFRASTRUCTURE MANAGEMENT

Edward S. K. Fekpe[1]
Centre for Surface Transportation Technology
National Research Council
Ottawa, Ontario K1A 0R6

ABSTRACT

Policy decisions regarding the regulations governing truck size and weight and management of highway infrastructure are complex and clouded by uncertainties surrounding the relevant input parameters. Addressing the uncertainties is required to improve objectivity and instill confidence in the decision making process respecting regulatory and highway infrastructure management policies.

1. INTRODUCTION

The physical and operating characteristics of the trucks are primarily determined by the regulations governing their sizes and weights, the extent of their enforcement, the characteristics of freight and technology in the motor manufacturing industry. The performance of the highway infrastructure on which these trucks are operated is influenced largely by the weights imposed. In order to arrive at objective infrastructure management decisions therefore, it is important to understand the interaction between the heavy freight vehicles and the infrastructure.

In recent years emphasis has shifted from construction of new infrastructure to cost-effective management of existing infrastructure, where truck movements and the implications of their operations become a major concern to highway agencies all over the world. It was also observed (OECD, 1988) that better technical understanding of the impact of heavy vehicles on highway infrastructure is required to permit more cost-effective resource allocation for road construction and maintenance.

---

[1] Formerly, Dept. of Civil Engrg, Univ. of Manitoba, Winnipeg, Canada R3T 5V6

A number of trade agreements intended to facilitate international trade have been initiated in recent times. For example, the North American Free Trade Agreement (NAFTA); the European Economic common Transport policy). In implementing such agreements consideration is given, among other things, to regulatory policies governing the physical and operating characteristics of the freight vehicles. This is because the trucking industry plays a key role in the economic development and has implications for the infrastructure.

The objectives of this paper are to: (i) identify the policy and technical aspects of the regulations, and the linkage between the regulations and highway infrastructure management; (ii) highlight the uncertainties associated with input variables required for policy decisions regarding weight regulations; and (iii) present the results of current research that address some of the uncertainties with examples.

2. POLICY AND TECHNICAL CONSIDERATIONS

In discussing the regulations governing the size and weights of vehicles and highway infrastructure, it is important to understand the interaction between the policy and technical aspects from which they ultimately evolve. Highway infrastructure are provided and maintained with the primary objectives to: improve economic and social well being; improve accessibility; promote efficiency in the transportation of goods and services; promote commerce and economic activity. These objectives are however subject to constraints including funding levels, competing investments in other sectors of the economy, safety, energy and environmental concerns. The management and engineering of the infrastructure can be visualized as a complex web of technical and policy relationships, processes and issues requiring addressing trade-off questions relating to resource allocation.

The policy aspects are concerned with directives that are translated into regulations and standards relating to (i) the types and operating conditions of equipment using the infrastructure and (ii) design, construction, maintenance and management of the infrastructure. The standards and regulations are expressions of the quality and values as desired by the road users or the general public. These in turn give rise to technical and technological innovation creating new vehicles, operating conditions, new design practices, construction and maintenance and management practices and processes. Policy intervention is also required in controlling the effects of the regulations and standards. The

technical aspect is primarily concerned with identifying and executing technically feasible procedures to ensure efficient and safe operating conditions that meet the objectives of the policies and within the constraints imposed. The technical analyses also provide the necessary input for policy decisions.

3. REGULATORY REFORM

It is difficult to achieve uniformity in the regulations within any geographical region because of differences in transportation needs, infrastructure provision, regulatory and operating environments. For example, in the USA and Canada, individual states and provinces have regulations that are different from the national regulations. The national regulations are the least common denominator types. Very often policy makers are confronted with decisions that ultimately require a revision of the regulations governing vehicle weights and dimensions. The need to revise the regulatory limits may be the result of one or more of the following conditions:

- pressure from the trucking industry to improve economic efficiency (e.g. based on the premise that larger trucks are more economical because of the associated larger payloads) may necessitate relaxing the limits;

- in pursuance of technical harmony and promoting international trade in a geographic region (e.g. North American Free Trade Agreement (NAFTA) involving Canada, USA and Mexico; European Economic Community (EEC), consideration must be given to the regulations governing the size and dimensions of heavy freight vehicles. This is because freight movement by trucks has important economic implications in terms of transport costs and the highway infrastructure;

- the need to improve competition between alternative modes of freight transportation;

- the need to conserve energy.

The decision to revise the regulations to meet any of the above objectives is constrained by a number of competing and conflicting issues. Some of these issues are: (i) trucking productivity; (ii) implications for the maintenance and management of existing infrastructure: and (iii) safety

and environmental aspects of truck operations. The interactions and interrelationships among these issues are obviously complex. Policy decisions therefore require trade-offs between these issues. A recent study (TRB, 1990a) concluded that the issue of truck sizes and weights is complex and controversial involving sensitive trade-off between costs to build and maintain highways and the costs to transport goods by trucks and they have implications for highway safety, traffic flow and highway finance. These decisions should be based on sound analytical procedures backed by meaningful assumptions that take into consideration the interaction and complexity of the variables in question. Such analytical procedures allow objective assessment of the relative impacts of any regulatory changes in order to arrive at policies and strategies that are feasible both physically and economically. Currently, these procedures are either not available or have notable weaknesses and characterized by uncertainties in the input variables.

4. HIGHWAY INFRASTRUCTURE ISSUES

The provision, maintenance and management of highway infrastructure requires sufficient knowledge about the factors affecting their performance. McNeil et al. (1992) observed that maintenance, rehabilitation and reconstruction of transport infrastructure is important to the economic well being of any nation but due to constrained resources, it is necessary to make cost-effective decisions with respect to which facilities, which strategy, when to apply it, and to what extent. Fundamental to infrastructure management is the availability of relevant data and prediction models on the condition of the facility and the factors that determine the performance, structural and functional, of that facility. The factors dictating the performance of infrastructure are traffic, environment and operating conditions. Among these, traffic is considered the major contributing factor. Traffic characteristics of interest include: the gross weight; type and spacing of axles; weight distribution on the axles; tire per axle; suspension system; tire size and pressure; volume of traffic and composition by axle or weight class. These characteristics are principally determined by the governing regulations.

Hutchinson (1990) identified and discussed the linkages between the properties of trucks and highway transportation system characteristics. Among the various classes of impacts, infrastructure impacts (bridges and pavements) are affected by all truck properties with pavements being the most affected. It is observed that axle loads and configurations along with suspension properties and tire pressures have important influences

on pavement design, maintenance costs and bridge design. The overall dimension and geometry of vehicles influence turning movement requirements, highway capacity and safety and are generally related to the geometric features of the highway. The gross vehicle weight has important linkages with bridge design as well as a number of geometric design problems. For example, the weight/power ratios of trucks influence their performance on grades and because of this the traffic capacities and safety on upgrade sections.

Pavement damage is assessed in terms fatigue failure or permanent deformation. The condition of the pavement is reflected in operating speeds, vehicle operating costs, transportation costs and fuel economy. Bridge responses to heavy vehicle loads are evaluated in terms of overstress and fatigue. Overstress concerns the possibility of severe damage and possible collapse caused by a single extreme loading. Fatigue concerns the cumulative damage caused by repeated load passages. Fatigue for both pavements and bridges are due to heavy vehicle loading.

Managing the infrastructure involves making decisions that are based on current data and models capable of predicting the performance (or condition) of the infrastructure. This is facilitated by deterioration models which are used to estimate life-cycle impacts of investments in transportation facilities in terms of costs and benefits and condition of the facility. These models are developed with sufficient knowledge of the probable weight distributions of all truck types under the prevailing weight limit and enforcement strategy.

5. REGULATORY AND INFRASTRUCTURE UNCERTAINTIES

As earlier noted, the physical and operating characteristics of trucks and hence the impacts on highway infrastructure are primarily determined by the regulations and the intensity of enforcement of these regulations. Invariably, the regulations form the core of transport policies relating to the productivity, infrastructure provision and management. It is therefore considered appropriate to address the issue of highway infrastructure management from the stand point of the regulations governing the size, weights, safety characteristics and operating conditions of freight vehicles and their enforcement.

In evaluating alternative regulatory and infrastructure management policies and strategies, estimates of the following variables are required:

- probable weight distributions (axle and GVW) under the alternative weight limits. This includes estimates of trucks operating in overloaded condition and the amount by which the weight limits are exceeded. This information is required to estimate the infrastructure loading and payloads under alternative weight regimes;

- changes in truck fleet mix given a change in the regulatory limits. The impacts on infrastructure and productivity depend on the composition of truck fleet;

- vehicle utilization measured by the average annual total vehicle distance travelled under the alternative limits is required in comparing the productivity and infrastructure impacts of various truck types, alternative limits and operating strategies;

- definition of acceptable level of enforcement for given regulatory limit to achieve tolerable violation rates. This information is important in establishing penalties or fine structure that can be effective in discouraging overweight trucking and has implications for resource allocation.

Recent studies have observed that reasonable estimates for these items are difficult to obtain because the mechanism of change associated with them is not fully known. This makes the estimates uncertain. These uncertainties are major constraints to objective evaluation of alternative weight limits and control policies. For example, recent US studies on truck weight limits observed that impact estimates of alternative weight limits are highly uncertain. This is because the inputs cannot be predicted with any degree of certainty (e.g. TRB, 1990a, 1990b). Difficulties encountered in estimating these parameters include:

- the responsive change in any variable is a function of the magnitude of change in the governing limits. This relationship is different for each variable and in some cases compounded;

- given a change in the regulations, it is difficult to predict with any degree of certainty changes in fleet mix. The time lag before the system fully adjusts to the revised limits to attain a reasonably stable condition is dependent on a number of factors including the magnitude of change, equipment replacement cycle, and availability and access to funds.

- innovations and new technologies in the truck manufacturing

industry results in the introduction of new truck types which may distort understanding of the fleet mix and their physical and operating characteristics. Such innovation is even encouraged when new weight limits are introduced. Interviews with trucking firms have been the main source of information on probable changes in fleet mix. These interviews generally provide only subjective assessments;

- until recently, the relationship between enforcement effort and the level of compliance with the regulations and subsequent effect on truck weights was not understood. In fact, the enforcement factor has not been adequately accounted for in truck weight analysis in some studies e.g TRB, 1990a, 1990b principally because of the uncertainty;

- there are uncertainties respecting the safety aspects of new vehicle designs that may result from revision of the regulations. Subsequently, estimates of usage of new vehicle types and changes in the fleet mix are uncertain.

With regards to infrastructure management, information on current condition of the facility, provided by facility inspection, and future condition, forecast by performance models are required. An important variable in performance models is the traffic. Infrastructure performance models are used in selecting activities for the current period and for planning future maintenance and rehabilitation activities. Madanat (1993) observed that both forms of information are characterized by a large degree of uncertainty. Model forecasts are characterized by a high degree of uncertainty due to measurement errors and inherent randomness in the variables used in the model.

These uncertainties from both the regulatory and infrastructure standpoints have important implications for the management process. Available literature indicates that the focus of research in uncertainty has been in pavement management systems decision making. There is little appreciation for uncertainty from the regulatory standpoint let alone adequately addressed in the context of infrastructure management. In the following section, results of recent research aimed at resolving some major uncertainties from the regulatory framework and how it relates to infrastructure management are reported.

6. ADDRESSING REGULATORY UNCERTAINTIES

Some results of recent research to address some of these uncertainties are presented. The focus has been on two closely related aspects (i) understanding the enforcement-violation relationship and (ii) predicting the weight distributions for various truck types.

An enforcement intensity-violation rate relationship has been developed (Fekpe et al. 1994a) that allows, among other things, the effect of enforcement on weight distributions to be accounted for. It also allows assessment of the consequences and implications of adopting alternative weight control strategies. In terms of policy issues, it allows the estimation of maximum possible violation rates for given intensities of enforcement and definition of acceptable levels of enforcement effort required to achieve tolerable violation rates.

A new procedure is developed to predict gross vehicle weight (GVW) distributions as a function of the weight limit and the intensity of enforcement (Fekpe et al. 1994b). The GVW predictive models are developed for the "steady state" condition for laden trucks, which represents the situation that would exist if any change in the limits had been in effect long enough for the trucking industry to have fully adjusted the fleet to optimize operation under the new limits.

In this procedure, two distinctive GVW distribution families were identified - each represented by the most common or dominant configuration that reflects the characteristics of that family. Predictive models are developed for the defined reference truck types in the two identifiable weight distribution families (the "all-commodity" family represented by the five-axle tractor-semitrailer (3-S2) and the "weight-out" family by 8-axle B-train). Predicted GVW distributions of reference trucks are converted into those of other trucks in the respective GVW families based on the second hypothesis using truck substitution ratios (Fekpe et al. 1993).

The procedure is demonstrated to have a strong predictive capability and stable (transferable) in time and space. The procedure includes an objective method of translating these weights into axle weight distributions and equivalent pavement loading. These models also allow average payloads to be estimated to assist in evaluating trucking productivity. The parametric form of the model is shown in Eqn. 1, a full description given in Fekpe et al. (1994b). Fig. 1. shows the model. For a given GVW limit and intensity of enforcement measured by the violation rate (i.e. number of trucks in violation as a percent of all trucks inspected), the GVW

distribution can be predicted.

$$F(.) = \frac{1}{1 + f(z)}\{Pr(x)\} \quad (1)$$

where
F(.) = cumulative density function - general GVW prediction model;
f(z) = violation rate (%) - function of enforcement intensity
Pr(x) = probability of operating at less than or equal to given GVW x;
x = GVW as percent of GVW limit.

Fig. 2 shows the steps involved in the evaluating alternative weight limits and enforcement levels for different truck types. The process involves predicting the cumulative GVW distribution for a specified GVW weight limit and intensity of enforcement measured by the violation rate. The GVW distribution is then converted into axle load distributions which are translated into pavement loading by calculating the equivalent standard axle loads (ESAL). The average payloads can also be determined. Parameters for assessing infrastructure impacts and trucking productivity can be subsequently determined and the effects of different weight limits and enforcement levels as well as the cost implications compared.

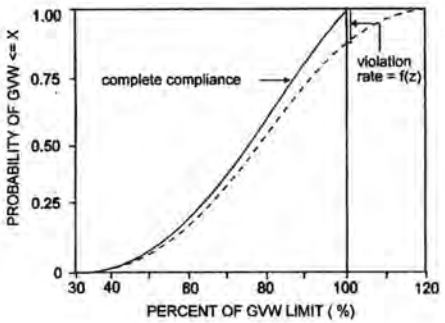


Fig. 1. GVW Prediction Model

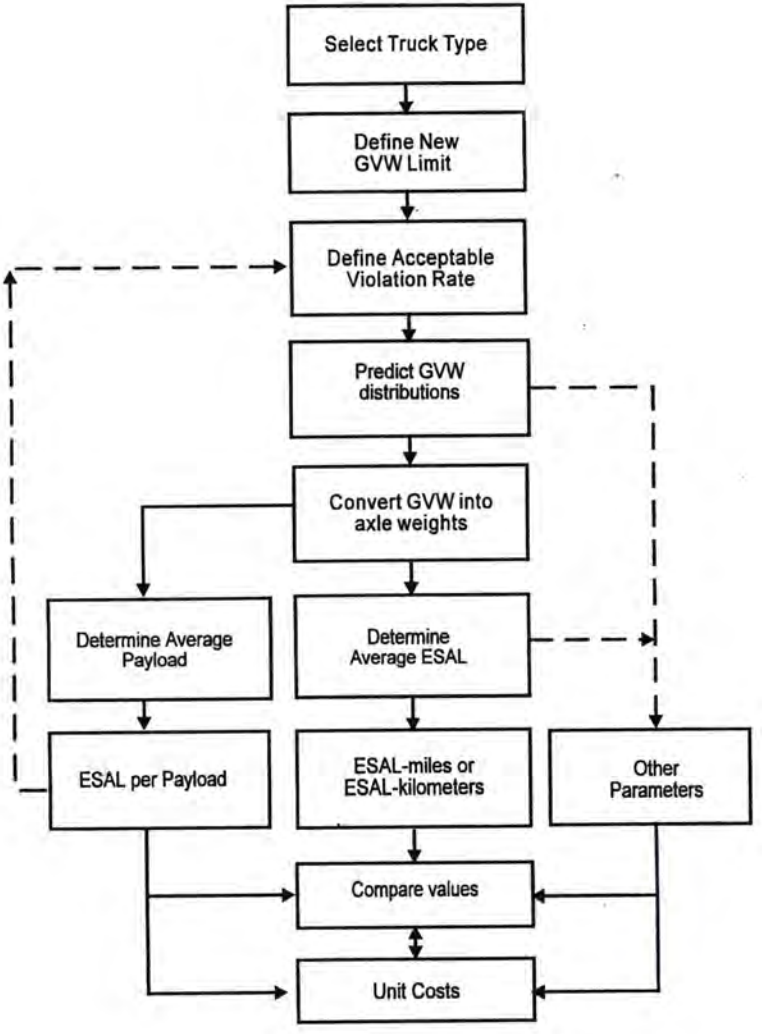


Fig. 2. Flow Chart of Evaluation Process

The application of the procedure is illustrated by two examples: (i) exploring the variation of pavement loading measured by the average ESAL per truck or truck load factor (TLF) with GVW limit for different truck types; and (ii) comparing the pavement effects and trucking productivity (using the average payloads as proxy) of the 5-axle tractor semitrailer truck in the NAFTA countries.

6.1 Variation of Truck Load Factor with GVW Limit

The predictive models are used to predict the GVW distributions for different weight limits for different truck types. These weights are distributed between the axles of each truck based on the weight split characteristics (Fekpe et al. 1993). The axle loads are then translated into TLF using the "fourth power" rule with an exponent of 3.8. Fig 3. shows the variation between TLF and GVW limit for five common truck types operating on the highways under complete compliance condition i.e violation rates are negligible.

It is observed that the pavement loading resulting from these trucks increases with the GVW limit in a curvilinear fashion with no apparent optimum GVW limit. It also observed that under complete compliance condition, 3-S3 truck operating under 48.5 tonne limit and 7-axle A-train truck under 52.5 tonne limit impose the same amount of pavement loading measured by the TLF. Similarly, 7-axle B-train at 58.5 tonne GVW limit will impact the same pavement loading as 8-axle B-train at 62.5 tonnes. The average payloads can be similarly compared.

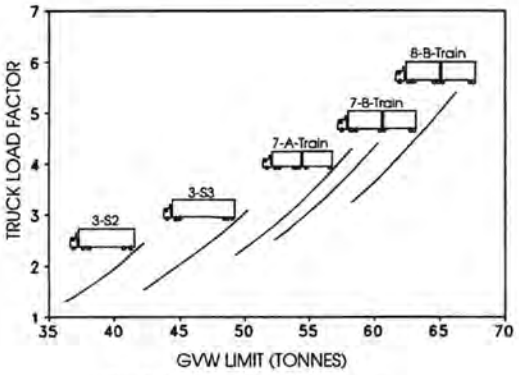


Fig. 3. TLF - GVW Limit Relationship

6.2 Enforcement Effect on Weight Distributions

The effect of enforcement on the weight distributions and hence pavement loading are shown for the 3-S2 truck in Fig. 4. It is noted that generally, higher infrastructure impacts result from low enforcement intensities as reflected in the violation rates. The relationship provides the means of assessing the incremental pavement costs measured by the extra ESAL associated with not enforcing the regulations. It is also noted (Fekpe et al. 1994c) that for all truck types, (i) the effect of enforcement on the TLF becomes more important at higher GVW limits and (ii) for a given violation rate, changes in TLF are greater at higher GVW limits.

The models therefore allow the infrastructure impacts of different truck types and weight limits and enforcement levels to be explored in order to arrive at regulatory and control polices that are compatible with the existing infrastructure capabilities.

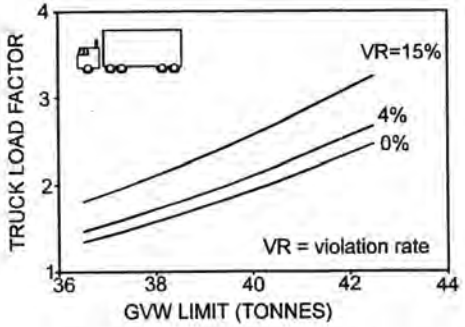


Fig. 4. Effect of Enforcement on TLF

6.2 Comparing 3-S2 Truck in NAFTA Countries

Within the context of the regulatory reforms under NAFTA a number of possibilities exist. For example Pearson (1993) identified three options namely: (i) common regulations in the three countries (this is considered highly unlikely); (ii) common acceptable limits for specific truck types; and (iii) acceptable standards for transportation of selected commodities. Whatever the case, it is imperative for each party to assess the implications for the options from the trucking industry's (productivity gains or losses) and agency's (infrastructure impacts and implications) points

of view. This example summarizes the differences in the pavement loading and average payloads for the 3-S2 truck operating under current conditions in each country (Table 1). The analysis is based on the premise that each jurisdiction maintains and enforces the existing limits.

It is observed that the GVW limit for this truck in Canada is about 9% higher than the US counterpart but the TLF is about 37% higher and carries an average payload that is about 8% higher. In terms of transborder trucking, Canadian trucks operating in the US at US limits will be losing 1.1 tonnes per truck per trip. Stated otherwise, on average 11 trips will be required to move the same amount of freight instead of 10 trips if Canadian trucks operate at US limits. On the other hand, if Canadian limits are allowed in the US the pavements will be subjected to at least 37% more loading than under the US limits. Similarly, the GVW limit for this truck in Mexico is about 15% higher than the US limit and about 5% greater than the Canadian limit. The TLF is 65% and 17% higher than the US and Canadian counterparts respectively. Obviously, Mexican trucks have higher payload capacity but impose much higher pavement damage measured by the TLF. For a given amount of freight that can be moved in 10 trips by trucks operated at Mexico limits, it will require 12 trips when operated at the US limit and 11 trips at the Canadian limits for this truck type.

Table 1. Summary of Regulations and Impact Indicators - NAFTA

Country	GVW Limit (tonnes)	Truck Load Factor †	Ave. Payload (tonnes)† [a]	ESAL per Payload
Canada	39.5	1.831	15.27	0.120
USA	36.3	1.337	14.17	0.094
Mexico	41.6	2.210	16.86	0.131

† - complete compliance condition.
[a] - average tare weight is assumed to be 15.5 tonnes for all regimes

The models enable prediction of weight distributions, pavement loading and average payloads for different truck types under different weight regulatory and enforcement strategies. This information can be used in modelling deterioration that would assist in facility condition prediction to facilitate infrastructure management. Note that these examples demonstrate only the relative infrastructure impacts in terms of pavement

loading and productivity in terms of the average payload. Other parameters relating to the physical characteristics of the trucks (e.g. suspension types, tire types, tire pressures etc) and condition of existing infrastructure are required in the overall impact analysis. For example, tire and suspension types have significant effects on load equivalency factors and these are becoming critical as more special tires and different suspension types are being used on the highways. Literature on the subject suggests that uncertainties regarding other input data are yet to be resolved. In order to improve the technical basis for the decision making process therefore, it is imperative that research efforts be directed to resolve the outstanding uncertainties.

7. CONCLUSIONS

The paper identified and discussed the linkage between the regulations controlling vehicle size and weights and highway infrastructure management focusing on uncertainties associated with regulatory policies. It is shown that policy decisions respecting the physical and operating characteristics of trucks have important implications for the highway infrastructure performance. Policy decisions regarding the regulations is complex and clouded by uncertainties surrounding the relevant input parameters. Although an objective and reasonably accurate procedure for predicting the GVW distributions is developed, there still exist other uncertainties that require attention to improve objectivity and instill confidence in the decision making process.

8. REFERENCES

1. Fekpe, E.S.K. and Clayton A. (1994a) "Predicting Heavy Vehicle Weight Distributions". ASCE Transportation Engineering Journal Vol. 121, No. 2 pp 158-168.

2. Fekpe, E.S.K. and Clayton A. (1994b) "Quantitative Assessment of the Effect of Enforcement Intensity on Violations of Vehicle Weight and Dimension Regulations". Journal of Transportation Planning and Technology. 1994, Vol. 18 No. 2, pp. 143-153.

3. Fekpe, E.S.K. and Clayton A. (1994c). " Estimating Pavement Loading in terms of Weight Limit and Enforcement Intensity". Proc. Annual CSCE Conf. Winnipeg, Manitoba. Vol. 3. pp 639-648.

4. Fekpe, E.S.K. and Clayton A. (1993). "A Rational Approach to Predicting Pavement Loading". Proc. CSCE Conf. Fredericton, New Brunswick. Vol 3. pp. 597-606.

5. Hutchinson B.G. "Large Truck Properties and Highway Design Criteria". ASCE Journal of Transportation Eng. 1990 Vol. 116 No. 1 pp 1-22.

6. McNeil S., Markow M., Neumann L., Ordaway J. and Uzarski D. "Emerging Issues in Transportation Facilities Management". ASCE Transportation Engineering Journal. 1992 Vol. 118, No. 4, pp 477-493.

7. Organization for Economic Co-operation and Development (1988) Heavy Trucks, Climate and Pavement Damage. OECD Road Research, Paris, France.

8. Pearson, J.R. (1993). " NAFTA Size and Weight Issues: A Canadian Perspective". Presented at 72nd Transportation Research Board: Annual Meeting. Washington D.C. USA

9. Transportation Research Board (1990a) Truck Weight Limits: Issues and Options. Special Report 225. Transportation Research Board: Washington D.C. USA.

10. Transportation Research Board (1990b) New trucks for Greater Productivity and Less Road Wear: An Evaluation of the Turner Proposal. Special Report 227. Transportation Research Board: Washington D.C. USA.

A Term Project
For
Decision Making Under Uncertainty
Model and Choices Course (CE 862)

Sensitivity Analysis
For
AASHTO Flexible Pavement Design

Submitted to
professor G. A. Sparks

By:
Hamidreza Soleymani (Ph. D. Student)
S.I.N. 909 157 240

Transportation Centre
Engineering Department
University of Saskatchewan
Saskatoon S7N 5A9

Submitted to Canadian Transportation Research Forum
Student Research Paper Competition
1993-1994

Sensitivity Analysis for AASHTO Flexible Pavement Design

1.0 Introduction

Sensitivity analysis is an important tool for evaluating the behavior of the performance models over a range of considered conditions that are practically expected. Sensitivity analysis could be applied for pavement design and analysis to understand the importance of the pavement design factors. For practical purpose, sensitivity analysis may require more laboratory characterizations and/or controls during construction and service life of the pavements.

In this paper, the 1986 AASHTO flexible pavement design equation for flexible pavement design was selected as a performance model and sensitivity analysis conducted for that model. There was two specific objectives: first, to evaluate the sensitivity of the 1986 AASHTO flexible pavement design to various pavement design factors (i.e., asphalt concrete modulus, layer thicknesses and traffic loads) and, second, to evaluate the effects of pavement design factors on performance of the pavement based on difference between Initial Serviceability Index and Terminal Serviceability Index (ΔPSI).

2.0 Performance Model for the Sensitivity Analysis

The experimental model of the AASHTO Guide for design of pavement structures was selected as a performance model for this sensitivity analysis (1). The main differences between the 1986 version of the AASHTO Guide and previous version are as follows:

1. Consideration of reliability concept,
2. Use of elastic (resilient) modulus in material characterization,
3. Consideration of drainage condition,
4. Consideration of the effect of frost heave, swelling soils, and thaw-weakening on pavement evaluation,
5. use of Non-Destructive Test (NDT) in pavement evaluation,

6. Use of life-cycle cost analysis in determining the most cost-effective construction/rehabilitation strategy.

The flexible pavement design equation based on 1986 AASHTO guide is as follows:

$$\log W_{18} = Z_R . S_0 + 9.36 \times \log(SN+1) - 0.2 + \frac{\log[\frac{\Delta PSI}{4.2-1.5}]}{0.40 + \frac{1094}{(SN+1)^{5.19}}} + 2.32 \times \log M_R - 8.07$$

Where:

W_{18}= predicted number of 18-kips Equivalent Single Axle Loads (ESALs),

Z_R= standard normal deviate

S_o= combined standard error of the traffic prediction and performance prediction,

ΔPSI=difference between the initial design serviceability index P_0, and design terminal serviceability index P_t,

M_R= resilient modulus of subgrade(psi)

SN = structural number

$SN=a_1D_1+a_2D_2m_2+a_3D_3m_3$

a_i= i th layer coefficient

D_i= i th layer thickness (inches)

m_i= i th layer drainage coefficient.

On the basis of Serviceability Index (ΔPSI) this model will be as follows:

$$\log[\frac{\Delta PSI}{4.2-1.5}] = [0.4 + \frac{1094}{(SN+1)^{5.19}}] * [\log W_{18} - 9.36 * \log(SN+1) - 2.32 * \log M_R + 8.07]$$

In a short term this equation could be summarized as bellow:

$$\Delta PSI = f(W_{18}, a_1, D_1, a_2, D_2, m_2, a_3, D_3, m_3, M_R)$$

In this sensitivity analysis, the difference between initial serviceability index and terminal serviceability index (ΔPSI) were calculated based on expected variations of other variables. The Z_R for the design values (or mean values) of variation assumed zero and this means that the reliability for design values are 50%.

3.0 Variation in Pavement Design Factors

All factors that are considered in pavement design have degree of uncertainty. On the basis of the available information or policy of designer, the most likelihood amount of these factors are selected for design of the pavements. In many cases, such design considered values change and consequently the performance of the pavement will be different from expected performance. The amount of change for each factor depends upon many other variables that could not be easily predicted, therefore the available deterministic design procedures are inadequate and some probabilistic models are used to have a better prediction of what may happen for a special pavement condition.

Thicknesses of pavement layers, layer material properties, traffic loads and repetition, subgrade strength and drainage coefficients, in the AASHTO flexible pavement design model, are factors that degree of uncertainty is expected on them. A brief discussion of the degree of uncertainty and variation on each of these factors is presented bellow.

3.1 Variation of Layers Thickness

Change in pavement layers thickness could lead to significant change in the pavement performance.

The variations of the individual layers thickness have been studied in numerous quality control studies. Sherman studied the thickness variation of the pavement layers and concluded that there was an increasing degree of accuracy in the layer thicknesses from sub-base through to surface. Table 1 shows the approximate standard deviations obtained by pooling the mean squares for the various pavements layers thickness (2). The type of distribution associated with the thickness variations would be expected to be approximately normally distributed due to the many contributing sources of error.

Pavement Layers Thickness	Standard Deviation
Asphalt Concrete	0.41 (in)
Aggregate Base	0.79 (in)
Aggregate Sub-Base	1.25 (in)

Table 1: Standard Deviation for Thickness of Pavement Layers

Lytton and Zollinger suggested a coefficient of variation for pavement layer thickness as given in Table 2 (3).

Pavement Layers Thickness	Coefficient of Variation %
Surface	5-12
Base	10-15
Sub-Base	10-15

Table 2: Coefficient of Variation of pavement layer thicknesses

For this study, the standard deviation for thickness of asphalt concrete, base and sub-base layers selected 0.41, 0.79, 1.25 (inches) and with assuming a normal distribution, the coefficients of variations were 5.5%, 12% and 11% for these layers respectively. With 90% and 95% reliability, the maximum and minimum thicknesses of these layers were calculated.

3.2 Variation of Pavement Material Properties

The variation of the layers strength is a function of type of material, layers location, quality control and uniformity in construction.

The typical values of coefficient of variation for the sub-base and base layers range from 15 to 40 percent. Lytton and Zollinger suggested coefficient of variation for the elastic modulus of pavement layers as in Table 3 (3).

Pavement Layers Modulus	Coefficient of Variation
Surface	10-20%
Base	8-20%
Sub-Base	10-20%

Table 3: The Coefficient of variation for the modulus of pavement layers

For this study, a coefficient variation of 15% selected for asphalt concrete, base and sub-base layers and pavement layer coefficient a_1, a_2, and a_3 changed with a reliability of 90% and 95%. It has been assumed that the distribution of variations for coefficient of the pavement layers is normal.

3.3 Variation in Traffic

Perhaps the most uncertain factor in the AASHTO flexible pavement design model is the traffic loading and traffic repetition over the life of the pavement. Many factors give rise the uncertainty in traffic prediction, which include the socio-economical factors.

Such uncertainties become more evident for high volume highways. For example, the pavement design of New Jersey Turnpike was based on 20 $\times 10^6$ vehicle applications, but estimates indicated that the 20-year count was over 90 $\times 10^6$ vehicle applications. This shows for some cases it is possible to have more than 4 times traffic loads than predicted traffic loads (4).

Deacon and Lynch estimated Equivalent Weight Load (EWL)'s using a new method with data from 1950 to 1968. EWL's were estimated using the method and compared to actual EWL's for loadmeter stations at which both vehicle classification and weight data had been obtained. Mean square residual of the log of the total EWL's was determined to be 0.0453 (5).

In this study, the design ESAL's were assumed 27 $\times 10^6$. With considering the Deacon and Lynch results the maximum and minimum expected traffic would be 44 $\times 10^6$ and 16.5 $\times 10^6$ ESALs. Distribution of traffic variation is skew to the right and it means that for most cases more traffic is expected than the design traffic.

3.4 Variation of Sub-Grade Strength

Lytton and Zollinger suggested the coefficient of variation between 10 to 30% for elastic modulus of sub-grade (3). Darter et al. studied the variation of pavement subgrade stiffness from the deflection based on the surface curvature index (SCI). The normal distribution was recommended for consideration of the variation of sub-grade stiffness. In another study, involving 153 projects, the average coefficient of sub-grade stiffness was 10 percent (6).

The design modulus of sub-grade was considered 5700 psi. With considering 20% of coefficient variation for sub-grade modulus, the maximum and minimum modulus of sub-grade with 90% reliability were 7580 and 3820 and with 95% reliability were considered 7930 and 3460 psi.

3.5 Variation of Drainage Coefficient

AASHTO 1986 has introduced a new concept for considering in the flexible pavement design equation called drainage coefficient. This coefficient rates the pavements as good, fair, and poor. In this study, the drainage coefficient (m_1 and m_2) selected for these three conditions 1.25, 1.15 and 1.05 from the AASHTO Guide respectively.

3.6 Summary of Expected Variation of the Pavement Variables

Table 4 Summarized the expected variation of different variables for a flexible pavement.

Factors	Standard Deviation	Coefficient Variation	Distribution
Asphalt concrete thickness(D_1)	0.4(in)	5.5	Normal
Aggregate base thickness(D_2)	0.79(in)	12	Normal
Aggregate sub-base thickness(D_3)	1.25(in)	11	Normal
Asphalt concrete modulus(a_1)	0.063	15	Normal
Aggregate base(a_2)	0.021	15	Normal
Aggregate sub-base(a_3)	0.012	15	Normal
Sub-grade modulus(M_R)	1140(psi)	20	Normal
Drainage coefficient(m_1, m_2)*			
Traffic(W_{18})**			

* Three selected level for m_1, m_2 were good, fair and poor.

**The mean square residual of the log of EWL's was selected 0.0453.

Table 4: Expected Variation for Flexible Pavement Variables

Three levels of expected variation for different variables for 95% and 90% reliability are listed in the Tables 5 and 6.

Factors	Low Level	Design Level	High Level
D_1(95%)	6.7964(in)	7.6(in)	8.40369in)
D_2(95%)	5.25(in)	6.8(in)	8.85(in)
D_3(95%)	8.55(in)	11(in)	13.45(in)
M_R(95%)	7930(psi)	5700(psi)	3460(psi0
a_1(95%)	0.5434	0.42	0.2965
a_2(95%)	0.181	0.14	0.098
a_3(95%)	0.1305	0.08	0.05648
W_{18}	1,650,000	2,700,000	4,400,000
m_1 and m_2	1.25(good)	1.15(fair)	1.05(poor)

Table 5: Three Levels for Pavement Variables with 95% Reliability

4.0 Results

With considering the 1986 AASHTO flexible pavement design model and three levels for pavement variables, a sensitivity analysis has been conducted and the results for change in serviceability (ΔPSI) are presented in Tables 7 and 8.

Factors	Low Level	Design Level	High Level
D_1(90%)	6.9235(in)	7.6(in)	8.2765(in)
D_2(90%)	5.5(in)	6.8(in)	8.1(in)
D_3(90%)	9.94(in)	11(in)	13.06(in)
M_R(90%)	7580(psi)	5700(psi)	3820(psi)
a_1(90%)	0.52389	0.42	0.316
a_2(90%)	0.174	0.14	0.105
a_3(90%)	0.0.0998	0.08	0.0602
W_{18}	1,650,000	2,700,000	4,400,000
m_1 and m_2	1.25(good)	1.15(fair)	1.05(poor)

Table 5: Three Levels for Pavement Variables with 90% Reliability

Factors	ΔPSI	ΔPSI	ΔPSI
D_1(95%)	1.97557	1.56915	1.2803
D_2(95%)	1.8544	1.59915	1.28579
D_3(95%)	1.8232	1.56915	1.36658
M_R(95%)	1.0882	1.56915	2.72883
a_1(95%)	0.92967	1.56915	3.35447
a_2(95%)	1.29277	1.56915	1.96247
a_3(95%)	1.08448	1.56915	1.91892
W_{18}	1.24016	1.56915	1.98152
m_1 and m_2	1.40133	1.56915	1.77086

Factors	ΔPSI	ΔPSI	ΔPSI
D_1(90%)	1.90054	1.56915	1.32003
D_2(90%)	1.80298	1.59915	1.37968
D_3(90%)	1.67167	1.56915	1.39604
M_R(90%)	1.14405	1.56915	2.44527
a_1(90%)	1.00226	1.56915	2.88488
a_2(90%)	1.33445	1.56915	1.88642
a_3(90%)	1.34654	1.56915	1.85561
W_{18}	1.24016	1.56915	1.98152
m_1 and m_2	1.40133	1.56915	1.77086

Tables 7 and 8: Different Serviceability Level (ΔPSI) for Different Pavement Variables With 95 and 90% Reliability

Figures 1 and 2 show change in serviceability (ΔPSI) of a flexible pavement based on expected change in variables with 90% and 95% reliability.

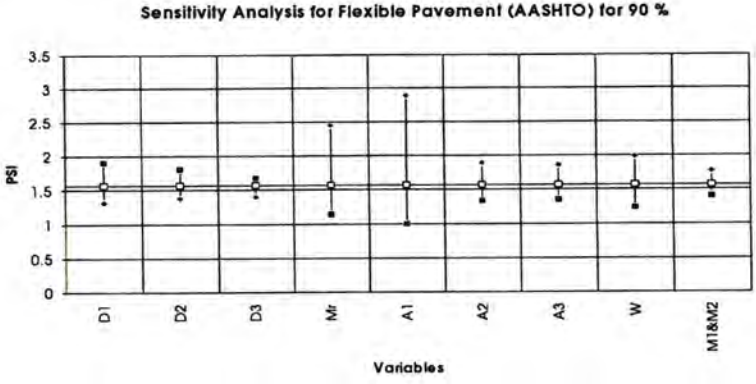


Figure 1: Change in Serviceability of Pavement With 90% Reliability

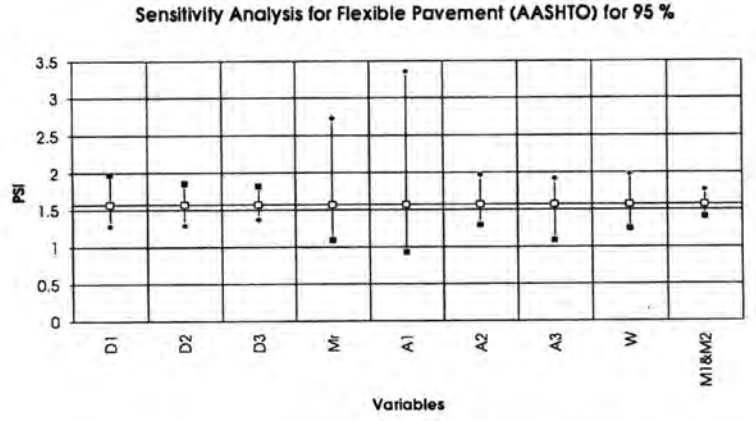


Figure 2: Change in Serviceability of Pavement With 95% Reliability

5.0 Conclusions

1-Thickness of asphalt concrete layer and modulus of subgrade are the most important variables in pavement serviceability based on AASHTO 1986 flexible pavement model.

2-With change in sub-grade strength and construction factors such as thickness of pavement layers, the performance of pavement could be different in different locations, as some areas of pavements have cracks and other distress and some other are as with enough strength.

3- The sensitivity analysis for change in serviceability could be used in Pavement Management System (PMS).

4-More quality control should be applied for accuracy of the asphalt concrete thickness during construction.

5-In the design of a flexible pavement, more attention should be considered in specifying the modulus of subgrade.

References

1- AASHTO, "Guide for Design of Pavement Structure", American Association of States Highway and Transportation Officials (AASHTO), 1986.

2-Sherman B. G., "In Situ Materials Variability", Special report 126, Highway Research Board, 1971.

3-Lytton R. L. and Zollinger D., "Modeling Reliabilities in Pavement", Transportation Research Board 1133, 1993.

4-Adnan A. B. and Adli H. B., "Probabilistic Design of Flexible and Rigid Pavements Using AASHTO Equation", TRB 1227, 1989.

5- Deacon J. A. and Lynch R. L., "Determination of Traffic Parameters for the Prediction, Projection, and Computation of EWL's", Kentucky Highway Department, 1968.

6-Darter M. I., Hudson W. R. and brown J. L., "Statistical Variations of Flexible Pavements and Their Consideration in Design", American Asphalt Pavement Technologist, 1972.

EVALUATION OF TRAFFIC SIGN RETROREFLECTIVITY

Susan S. Dableh, University of New Brunswick

Abstract

One cause of highway accidents is the inability to see traffic signs because of their decreased retroreflectivity which is necessary for night driving. Traffic signs must be clearly visible and easily understood in time for a driver to respond. The retroreflectivity of a sign depends on the sheeting material used and subsequently its ability to redirect a vehicles' headlights back to the driver.

Currently the only Canadian guidelines for nighttime visibility of signs are in the Uniform Traffic Control Devices for Canada (UTCD). The manual simply states that "warning and regulatory signs be reflectorized or illuminated to show the same colour and shape during day and night". No minimum retroreflectivity standards have been established. This research involved an evaluation of two-hundred-and-eighty-one in-service traffic signs for different districts and road classes in New Brunswick. The measurements were obtained in the autumn of 1994 during daylight using a retroreflectometer. A comparison is made between the data obtained and the recently developed United States Federal Highway Administration (FHWA) standards. Furthermore, a sign management system is discussed based on the survey results.

1.0 Introduction

Many people have driven past a STOP sign or missed a direction sign while driving at night, but do not realize why this happens. Perhaps the driver was not paying attention or there was a tree in the way or maybe the sign was not retroreflective enough. Retroreflectivity is the redirection of a vehicles' light from the headlight back to the driver as shown in Figure 1. The proper retroreflectivity of a sign is particularly important at night so that it is clearly visible, allowing enough time for the driver to react.

Accident statistics show that driving at night is more dangerous than day driving. The fatalities/vehicle miles travelled is more than three times higher at night than during the day (Paniati and Schwab, 1991). Fatigue, intoxication, and weather are

some contributing factors. Drivers depend on traffic control devices for warning, regulation, and guidance, particularly at night when visibility decreases. Traffic control devices must be easily visible in time for the driver to respond and make quick, safe decisions. Highway officials have determined that if the three elements of a sign - shape, colour, and legend - are uniformly established, traffic signs would be recognized more often. The ideal sign would have 24-hour uniformity, therefore the background should be reflectorized.

The luminance of a traffic sign is the degree to which a surface emits light while its brightness is the amount of light reflected from a sign that reaches the driver's eye. The conspicuity of a sign is the probability that a sign located within a visual periphery will be seen at a given distance. Sign legibility is the readability of a sign as a function of its contrast and brightness, as well as the size of the letters and/or symbols on the sign.

This research involves a traffic sign evaluation for different districts and road classes of New Brunswick. Canada presently only has the Uniform Traffic Control Devices for Canada (UTCD). The United States has the Manual for Uniform Control Devices (MUTCD) and the recently developed Federal Highway Administration (FHWA) standards which were used as a guide for this evaluation. The aging profile of sign retroreflectivity in New Brunswick is determined in this study. The determination of adequate service lives of traffic signs will enable the feasibility of a sign management system for the replacement of signs to be discussed.

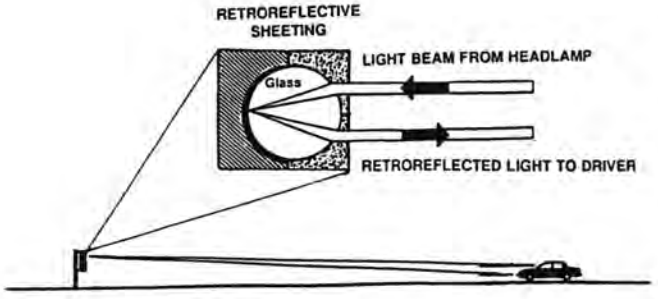


Figure 1: Retroreflection (Paniati, 1989)

2.0 Literature Review and Current Approach

2.1 Existing Retroreflectivity Standards

As automobile use increased and drivers travelled longer distances, an improved system of traffic control devices was needed. At first there were two national manuals, one for rural signing and another for urban traffic control devices. In 1927, the Association of State Highway Officials (AASHO) put out the urban manual, while the National Conference on Street and Highway Safety (NCSHS) put out the rural manual in 1930. These two manuals led to the publication of the MUTCD which sets the basic principles governing design and use of traffic control devices. The MUTCD was first published in 1935 and has been the "bible" of the profession (Hawkins, 1992). Since 1935, a new edition has been published every ten years.

2.2 Federal Highway Administration (FHWA)

The FHWA has been concerned about highway signs, markings, and other traffic control devices not being replaced when they should be. Studies of in-service traffic signs have shown that signs are acceptable during the day but defficient at night. An important property of traffic control devices that requires more study is retroreflection. The FHWA established the High Priority National Program Area (HPNPA). The goal of the HPNPA is not only to obtain information needed to determine the minimum retroreflectivity requirements, but also develop management programs and measurement devices necessary to effectively implement these requirements.

2.3 New Brunswick Department of Transportation

The New Brunswick Department of Transportation (NBDOT) Sign Shop is responsible for all traffic signs installed in the Province. Approximately fifty thousand signs are made per year by NBDOT. The warranted service life of engineering grade sheeting is normally seven years while that of high intensity grade is warranted for approximately ten years. Signs are only kept in storage for approximately one year at NBDOT before being placed in-service. The sheeting on traffic signs is sensitive to pressure therefore signs are stored in a slanted vertical position. If stored vertically the pressure would drive down the beads, that make up the sheeting, into the adhesives resulting in less resistivity or retroreflectivity. Pressure should not be applied to the face of a traffic sign for this reason.

Brilliance is only good to a certain extent. If the sheeting "glows" then the legend on the sign cannot be distinguished. The sheeting presently being used by

NBDOT sign shop is visible from a 1/2 km distance, if light is directly shining on it. A new engineering grade sheeting has been developed that is visible from a 1 km distance.

2.4 Manual on Uniform Traffic Control Devices (MUTCD)

2.4.1 Function of Signs

The Manual on Uniform Traffic Control Devices (MUTCD) identifies general sign specifications. Signs are needed where special regulations apply or where hazards are not obvious. They give information about highway routes, directions, destinations, and points of interest.

2.4.2 Sign Standardization

The "legend" or "symbol" on a traffic sign includes all word messages and symbols that have specific meanings. The basic requirement of a highway sign is that it be legible and understood in time for response. This means high visibility, lettering or symbols of adequate size, and a short legend for quick comprehension by a driver approaching a sign at high speed. Standardized colours and shapes are specified for quick recognition. Simplicity and uniformity in design, position, and application are important.

2.4.3 Design of Traffic Signs

Uniformity in design includes shape, colour, dimensions, legends, and illumination or reflectorization. Colour and shape determine the type of sign as shown in Table 1.

Table 1: Traffic sign designation

Colour	Shape	Type of Sign
Red	Octagonal	STOP or Prohibition
Green	Rectangular	Directional Guidance
Blue		Motorist Services
Yellow	Diamond	General Warning
Black on	Rectangular	Regulatory (ie Speed Limit)
Orange		Construction & Maintenance
	Inverted Triangle	Yield
	Pennant	No Passing
	Pentagon	School
	Circular	Railroad Crossing

2.4.4 Illumination and Reflectorization

Regulatory and warning signs shall be reflectorized or illuminated to show the same shape and colour by day and night.

2.4.5 Maintenance

Traffic signs must be kept in their proper positions, clean, and legible at all times. Damaged signs should be immediately replaced. To ensure adequate maintenance, a schedule for inspection, cleaning, and replacement of signs should be established. Employees of highway departments, police, and other government employees whose duties require them to travel on the highways should be encouraged to report any damaged or obscured signs as soon as possible. Special attention is needed to remove weeds, trees, shrubs, and construction materials. Regular schedule of replacement of lighting elements for illuminated signs should be maintained.

3.0 Problem Details

3.1 Reflectorized Traffic Signs

Retroreflective materials are used to enable drivers to easily see and recognize signs, delineation elements, and other reflectorized traffic control elements at night under headlamp illumination. Reflectorized traffic signs warn a driver of danger ahead or road regulations. It is obviously important for a driver to be able to detect signs in order to avoid hazards.

The detection and recognition distances of traffic signs and devices at night are determined by a variety of factors. These factors, which are found in the MUTCD, include size, colour, shape, background, weather conditions, and luminance of a sign. The luminance of a sign is determined by retroreflectance of the material; relative headlamp position, the sign, and driver's eyes; the intensity distribution of the headlamps; and the headlamp misaim (Zwahlen et. al, 1991).

3.2 Retroreflective Coefficient

The three main types of sheeting materials include the enclosed lens (glass beads), encapsulated lens (glass beads), and prismatic sheeting. Only the first two types of sheeting materials were encountered in this study. Retroreflectance of these is measured in units of candela per lux per square metre, (cd/lux/m^2), usually defined as R' which is the coefficient of retroreflection (Lumia, 1990).

$$R'=I/EnA$$

where I = reflected intensity
En = normal illumination incident on the retroreflection
A = area of the surface to be measured

Retroreflectance can be defined as the absolute brightness of the retroreflective target for a given amount of incident illumination on the target. This can be expressed as R'=L/En.

Difficult parameters to measure include the illumination of signs by vehicle headlamps, dirt on headlamps and signs, windscreens, and weather conditions. Influence of other factors on the retroreflectivity of traffic signs include (Colomb and Michaut, 1986):

a. Age of film, dirt, and dew deposit on the film,
b. Level of headlight intensity or dirt on headlights,
c. Additional illumination from the environment,
d. Geometry of lighting and observation,
e. Absorption of light by windshield, and
f. Visual acuity of drivers as a function of contrast between the letters and background for certain luminous ambient situations.

Luminance of retroreflective films depends on retroreflective characteristics of the film and on the degree of illuminance that strikes the sign. These two parameters are functions of geometric conditions of lighting and observation.

alpha (α) = angle of observation (ie angle between lighting and observation directions)
beta (β) = angle of incidence (ie angle between direction and a line perpendicular to the sign plane)

These angles change with the distance between the vehicle and sign and with the type of vehicle (ie truck or light vehicle) as the height of the observer's eyes varies with these two conditions. Calculations are made assuming that headlamps are at a height of 0.7 m for a light vehicle and 0.9 m for a truck and that the observer's eyes are at a height of 1.2 m in a light vehicle and 2.5 m in a truck (Colomb and Michaut, 1986).

Retroreflective coefficient (RC) defines the intensity of light reflected per unit of film surface for a certain degree of illumination received. This coefficient varies in terms of beta (2° and 40°). Degree of illuminance from vehicle headlights varies from vehicle to vehicle. The luminance of the film can be calculated by multiplying the illuminance on the sign by the RC of the film, taking into account the geometric conditions of lighting and observation that change with distance.

3.3 Retroreflective Sheeting Deterioration

The most effective way to attract a motorist's attention in time for him to react is to have signs with reflective colour backgrounds (AASHO, 1958). It must be a colour that is sharply contrasted against night blackness and stand out among other lights.

Retroreflective sheetings are complex multilayer composite products which deteriorate as a result of various factors. Typical failure modes include (Ketola, 1989):

a. Destruction of the metallic reflector coat,
b. Disruption or distortion of the optical elements within the sheeting, making retroreflection on incoming light less efficient,
c. Degradation or destruction of the outside polymer layer,
d. Fading of dyes or pigments used to produce appropriate colour in the sheeting or screen-printed graphics, and
e. Failure of bonds between layers, causing separation or delamination of the composite.

The type of sheeting failure depends on the type of sheeting used and the composition of each layer contained in the sheeting. Some failures can start or accelerate as a result of the environment. In some environments sunlight and moisture may cause the metallic reflector coat to corrode. In cases of changes between wet and dry or hot and cold conditions the polymer layers expand and contract causing problems in the optical path of the sheeting (Ketola, 1989).

Other deterioration variables of sheeting material include:

a. **Sheeting colour** causes variation in retroreflectivity.
b. **Contrast ratio** between background and legend has an effect on sign legibility.
c. **Sheeting type** is important. High-performance sheeting generally provides three to four times higher retroreflectivity than engineering grade but costs five times as much.
d. **Orientation to sun.** It has been found through research that sheeting facing north is more retroreflective than that facing south because the latter is exposed to the sun for longer periods of time.
e. **Ground elevation**. Higher solar radiation levels and severe weather cause increased sheeting deterioration in high-elevation locations.
f. **Area type**. Industrial and vehicle pollution affect sheeting deterioration.

g. **Sheeting age**. Duration of climatic exposure and other factors are determined by in-service age of the sheeting material determined by date of installation or manufacture.

4.0 Methodology

4.1 Field Investigation of Traffic Sign Retroreflectivity

Field evaluation of traffic sign retroreflectivity is required to identify signs that need replacement, according to FHWA standards. There were two-hundred-and-eighty-one traffic signs evaluated in this study. These consisted of enclosed lens (engineering grade) reflective sheeting and encapsulated lens (high intensity) reflective sheeting. This sign evaluation was done in southwestern New Brunswick due to convenience.

4.2 Equipment

Sign retroreflectivity was measured using a portable device called a retroreflectometer. This Model 920 Retroreflectometer can be used to randomly check reflective sheeting materials when they are manufactured to ensure they meet specifications. However the model 920 is designed mainly for measuring traffic signs in the field to check performance of the sheeting during the manufacturer's warranty period.

The retroreflectometer consists of two units, the 910B retroreflectometer head (a pistol-shaped assembly) and the 910 control unit (a rectangular assembly). The light source, detector, amplifier, and digital readout is enclosed in a modified 910 Optical Head. The power supply is a rechargeable 12 volt high capacity battery for longer operating life. Power is provided to the 920 through a phono cord.

The instrument is operated and stored in a temperature range of 32°F to 120°F (0°C to 50°C) and humidity from 5% to 90%. The 920 should stabilize at the ambient temperature before taking measurements, otherwise it will need to be frequently re-calibrated. Various environmental conditions do not affect the performance of the retroreflectometer.

4.3 Procedure

The retroreflectometer takes approximately sixty seconds to warm up and calibration drift is minimal. A set of calibration standards is provided by the manufacturer, to use during testing. Before taking measurements, the unit must be calibrated with these, according to the colour of sheeting being tested. A measurement is taken by pressing the retroreflectometer perpendicular against the

face of the sign in various spots to get an average reading. There is a plunger switch at the end of the barrel of the 920 which turns on the light source when the instrument is pressed against the material being tested.

The standard caps and front lens must be kept clean and the protective lens cap must be kept on the retroreflectometer when not in use. Flare and erroneous measurements will be obtained if the lens is dirty. The 920 reference standard caps should be re-calibrated every two years or more often if caps are scratched or damaged.

4.4 Data

Results of the two-hundred-and-eighty-one traffic signs retroreflectometer readings were analyzed and compared to the minimum requirements outlined in the FHWA standards. In most cases, three retroreflectivity readings were taken for a sign to obtain an average value. If the readings varied a lot, additional readings were taken. Colour, type, age, exposure, and average retroreflective reading of each sheeting as well as location (ie rural versus urban and collector versus local) were recorded. The percentage of maximum retroreflectivity was calculated and plotted on a graph versus age since the objective was to determine deterioration as a function of age. The percentage of maximum retroreflectivity gave a better representation than plotting the actual retroreflective reading versus age. The graphs contained some fluctuation and making it difficult to observe any trends therefore the data was smoothed using a moving average. This gave a better representation of any trends that existed.

5.0 Comparative Evaluation

5.1 Frequency of Sign Ages

Signs were found to range from a few months old to seventeen years old. Signs newer than a year old were considered one year old. The measurements were taken in 1994 which was used as the base year to determine the sign age. The year the signs were placed in service is put on a sticker found on the back of the sign, not to be confused with the date imprinted on the bottom front of the sign which is the date of manufacture. There were a considerable number of signs that did not have a date sticker. According to previous research, it was determined that these signs were installed between 1984 and 1986, therefore they were allocated to the ten year old category.

Signs should normally be replaced every five to seven years (Paniati and Schwab, 1991). As shown in Figure 2, the majority of signs are eight years old. Although NBDOT does not have a program to replace signs as a function of age, not many

signs were to be older than eight years. The reason is signs are usually replaced by NBDOT due to complaints or vandalism, therefore they are not used beyond their service life.

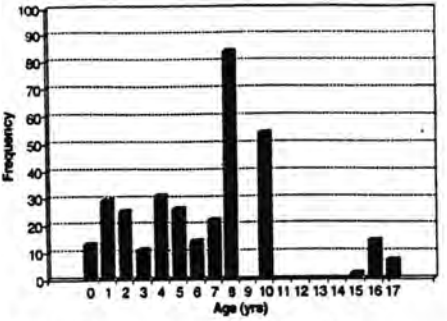


Figure 2: Frequency of Sign Ages

5.2 Deterioration of All Sign Colours

Each sheeting type and colour has a different retroreflectivity value. When a sign is new it should be 100% of its designated retroreflective rating. As the sign ages, it deteriorates due to reasons previously mentioned in this report. This results in a lower retroreflectivity value. After the sample of signs were measured a calculation was performed to determine the percentage of retroreflectivity. These values were then plotted against age to determine the effect of age on retroreflectivity. It can be seen in Figure 3 that retroreflectivity gradually decreases with age.

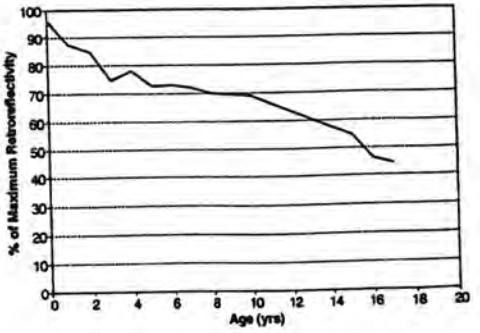


Figure 3: Deterioration of all sign sheeting colours

5.3 Yellow Engineering Grade Versus High Intensity Sheeting

Signs with yellow sheeting are used for cautionary (warning) signs. For this reason they are located in advance of the hazard. Since there is little reading time required for warning signs, the minimum retroreflective values are not sensitive to changes in speed. Warning signs include a variety of letter and symbol sizes. As a result, the amount of retroreflectivity required for legibility decreases as the size of the detail increases. There are different minimum acceptable retroreflectivity requirements as outlined by the FHWA standards depending on whether the sign falls into the fine versus bold category. The Manual on Uniform Traffic Control Devices (MUTCD) defines bold signs as those with symbols of turns, curves, arrows, chevrons, cross or side roads, T or Y intersections, lane reduction, divided highways or two-way traffic. Fine signs are those with letters or other symbols. Signs containing engineering grade yellow with fine symbols have a higher minimum acceptable requirement for retroreflectivity because they are more difficult to recognize than those containing bold symbols.

Most cautionary signs have an engineering grade sheeting rather than the more expensive high intensity sheeting. The standard retroreflectivity reading for engineering grade sheeting is 83.1 cd/lx/m^2 whereas for high intensity sheeting it is 273 cd/lx/m^2. The yellow background and black legend provide enough contrast resulting in good legibility. As shown in Figure 4, the retroreflectivity of engineering grade sheeting begins to decrease but slightly fluctuates after two years and then gradually deteriorates. This may be due to the use of different rolls of sheeting for different signs or sample variance. Each roll of sheeting may differ slightly in retroreflectivity. There is an overall decrease in the retroreflectivity of the engineering grade yellow sheeting with the minimum acceptable requirement for fine signs being reached at approximately year twelve. The minimum acceptable requirement for bold signs is never reached, even at year seventeen.

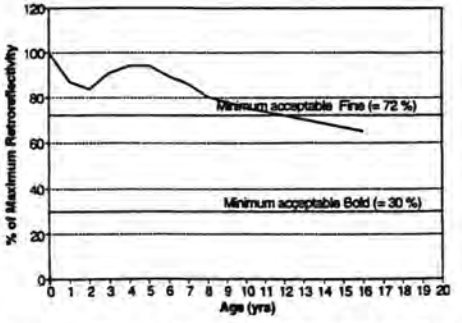


Figure 4: Deterioration of yellow engineering grade sheeting

As shown in Figure 5, the retroreflectivity of high intensity yellow sheeting gradually decreases with age but the minimum acceptable retroreflectivity requirement, as outlined in the FHWA standards, is never reached even after seventeen years. The minimum requirements in this case are lower than those for engineering grade sheeting because high intensity sheeting provides more contrast than the engineering grade therefore showing the same amount of retroreflectivity at a lower value.

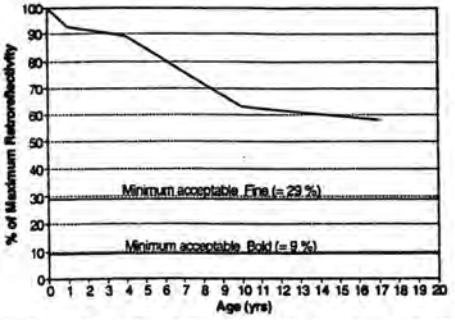


Figure 5: Deterioration of yellow high intensity sheeting

5.4 White Engineering Grade Versus High Intensity Sheeting

Signs with a white sheeting are predominantly for regulatory purpose. The signs in this category are placed at the point of regulation. Response is required prior to reaching the sign therefore traffic speed was considered with respect to retroreflectivity specifications. Two traffic speed categories were chosen to represent high speed rural and low speed urban conditions. For urban areas speed zones are generally 64 km/h (40 mi/h) or less which require less retroreflectivity than high speed rural areas of 72 km/h (45 mi/h). In a higher speed zone, sign retroreflectivity requirement must be greater because the sign must be seen and recognized at a distance further back since a motorist does not have much reaction time. The amount of retroreflectivity required for legibility decreases as sign size increases therefore size was considered with respect to retroreflectivity.

Engineering grade sheeting is used mostly for speed limit signs rather than the more expensive high intensity sheeting because the black legend on white background produces very good contrast, making the sign legible. The standard retroreflectivity reading for engineering grade sheeting is 116 cd/lx/m^2 whereas for high intensity it is 342 cd/lx/m^2. As shown in Figure 6, the percent retroreflectivity of engineering grade white sheeting is quite constant between 80% and 100% for the first ten years although there is some fluctuation. This could be a result of using different rolls of sheeting for different signs and sample

variance. Despite these fluctuations there is an overall decrease in retroreflectivity as age increases. Signs sixteen and seventeen years old actually fall below the minimum acceptable requirements as outlined in the FHWA standards.

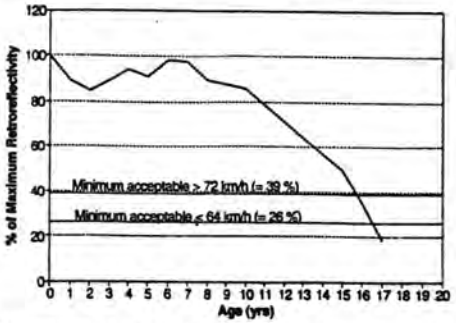


Figure 6: Deterioration of white engineering grade sheeting

The high intensity sheeting was only found on signs that also contain red, such as STOP, YIELD, and DO-NOT-ENTER signs. This is because there is not a very large contrast between red and white therefore a high intensity sheeting is required to make the sign clearly visible. The retroreflectivity of the white letters on STOP signs was not measured because it was difficult to get the lens of the retroreflectometer on the white part only. The high intensity white sheeting measured in this study was that on YIELD and DO-NOT-ENTER signs because they consisted of a larger white area. Since both the legend and background of these signs are retroreflectorized a minimum maintained contrast ratio of 4:1 is used as a general rule based on previous research (FHWA, 1993). If the retroreflectivity value for the white sheeting divided by the retroreflectivity value of the red sheeting is less than four, the sign should be replaced. The contrast ratios obtained from this study are shown in Table 2.

Table 2: Red and white contrast ratios for high intensity sheeting

Age	Red	White	Contrast ratio
1	64.8	348.6	5.4
2	65.0	323.6	5.0
4	20.1	324.1	16.1
5	22.0	372.6	17.1
6	16.5	295.1	17.9
7	14.4	319.1	25.0
8	16.5	300.4	25.5
9	44.5	279.9	6.9

The minimum retroreflectivity value for white is different when red is the background colour than if white was used as the background with a black legend,

such as for speed limit signs, due to the limited contrast. As shown in Figure 7, the high intensity white sheeting gradually deteriorates as it ages but remains much higher than the minimum acceptable requirements outlined in the FHWA standards, even at year seventeen. The minimum percent of maximum retroreflectivity reached at age seventeen is 77% which shows that white high intensity sheeting has a very good service life.

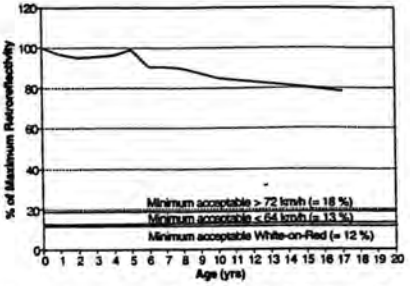


Figure 7: Deterioration of white high intensity sheeting

5.5 Red High Intensity Sheeting

Signs containing a high intensity red sheeting are STOP, YIELD, and DO-NOT-ENTER signs. The standard retroreflectivity reading for red high intensity sheeting is 68.0 cd/lx/m^2. As shown in Figure 8, retroreflectivity is high but decreases rapidly after two years and continues to gradually decrease until approximately age seven. Ultimately, retroreflectivity remains constant at 40%. Although this sheeting appears to deteriorate quickly, the sheeting at ten years is still above the minimum acceptable requirement as outlined in the FHWA standards.

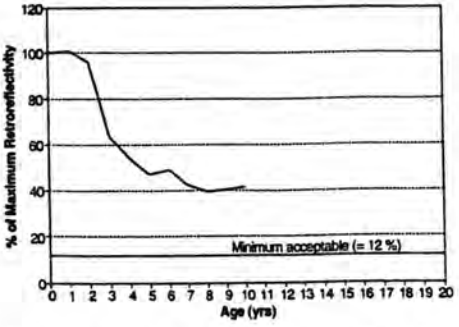


Figure 8: Deterioration of red high intensity sheeting

5.6 Red and Orange Engineering Grade Sheeting

Signs containing red or orange engineering grade sheeting material have the same standard retroreflectivity reading of 27.1 cd/lx/m^2, therefore they are grouped together. Signs with a red sheeting are usually STOP, YIELD, or DO NOT ENTER signs while those with an orange sheeting are construction signs. Only 4% of the stop signs measured contained a red engineering grade sheeting while all others had a high intensity sheeting material.

Construction signs are only temporary and often damaged or vandalized quickly due to heavy construction. As a result, most of the construction signs measured were approximately one or two years old. Construction signs must be kept new because of their importance for motorists to see them well in advance and take the necessary precautions. As shown in Figure 9, the retroreflectivity of new construction signs exceeded 100% at year one but quickly deteriorated with age. It can also be seen that they do not fall below the minimum acceptable requirement as outlined by the FHWA standards.

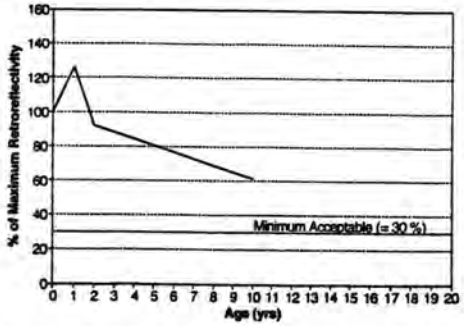


Figure 9: Deterioration of red and orange engineering grade sheeting

5.7 Blue and Green Engineering Grade Sheeting

Signs containing green or blue sheeting material have the same standard retroreflectivity reading of 19.2 cd/lx/m^2 for engineering grade sheeting and 60.1 cd/lx/m^2 for high intensity sheeting, therefore they are grouped together. Signs with a green sheeting are classified as direction or guide signs while those with a blue sheeting indicate motorist services. Not many signs were sampled in this group because green guide signs are generally mounted overhead therefore requiring the aid of a mechanical device, such as a boom truck, to gain access. Since they are generally placed overhead rather than along the roadside, sign placement was considered as a function of retroreflectivity. This is because the

lights from a vehicles' headlights shine on them at a different angle. Although guide signs do not usually require a manoeuvre prior to reaching the sign, the vehicle speed does affect the amount of time available for reading the sign and ultimately the distance at which the sign must be legible. For this reason traffic speed affects the retroreflectivity specification.

As shown in Figure 10, there is a gradual deterioration of engineering green and blue sheeting with age. The calibration lens for this sheeting colour was not very accurate. The calibration value could not be reached using the retroreflectometer therefore the readings taken were lower than they should have been but they still did not reach the minimum acceptable requirements outlined in the FHWA standards.

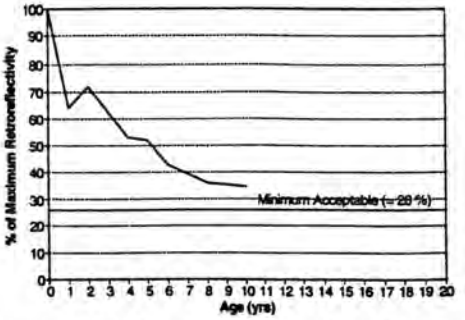


Figure 10: Deterioration of blue and green engineering grade sheeting

5.8 Exposure

The NBDOT sign shop has a test deck consisting of signs facing south in a tilted direction because according to previous research this is the direction of greatest deterioration from the sun. Visual inspections and readings are made on these signs monthly. For red on white signs a white sheeting is often used with a red ink. If the red ink is not clear coated deterioration is found to occur within eight months and the ink turns a pale pink colour.

When the signs were measured in this study, the direction they were facing was noted in order to determine the effect of exposure on retroreflectivity. As shown in Figure 11, signs facing south are less retroreflective than those facing north. This confirms that since signs facing south are exposed to the sun for longer periods of time they are less retroreflective. As shown in Figure 12, signs facing east are less retroreflective than those facing west. There is no explanation for this.

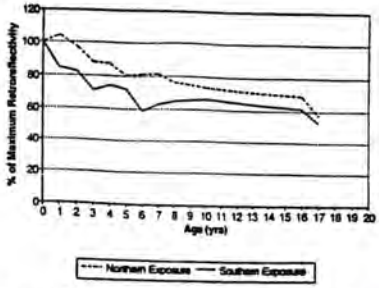


Figure 11: Deterioration of northern versus southern exposure

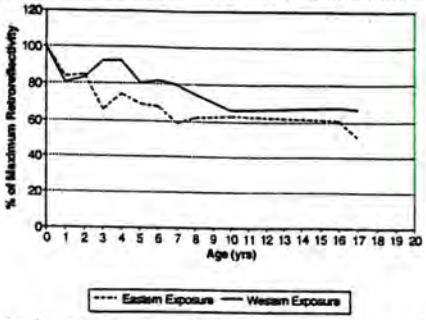


Figure 12: Deterioration of eastern versus western exposure

6.0 Sign Management System

One of the most important questions for any traffic agency is the optimum replacement time for retroreflective signs. Retroreflectivity of a sign depends on the sheeting material used which will eventually deteriorate and damage through weathering, impact from vehicles and/or snowplows, and vandalism. Sign replacement practices vary from place to place. Some signs are replaced every five to seven years which in many cases is uneconomical because the sheeting is still satisfactory while some signs are not replaced therefore causing accidents due to their loss of retroreflectivity. Larger size signs require lower levels of retroreflectivity therefore do not need to be replaced very often. The recommended optimal legend-background contrast for fully reflectorized signs is 12:1 (Sivak and Olson, 1986). For example, if the background luminance is 1 cd/m^2, the optimal luminance of the legend should be 12 cd/m^2. "Replacement (minimal) luminance" means the time when a sign does not meet its nighttime standard (Sivak and Olson, 1986). There is presently no criteria for establishing minimal sign luminance levels that is likely to be widely accepted. In this study, most signs seemed to be replaced before eight years. This is probably due to

damage or vandalism of signs rather than loss of retroreflectivity. As a result, the development of a sign management plan is unnecessary for this study.

A sign management system for traffic sign replacement could be developed allowing a sign inventory to be created that would keep track of the age and condition of signs. Some guidelines can be used for the replacement of various sheeting types and colours from the graphs shown in this report. These guidelines are based on FHWA minimum retroreflectivity standards. For example, Figure 4 indicates that yellow engineering grade sheeting should be replaced after twelve years. Figure 6 indicates that white engineering grade sheeting should be replaced after fifteen to sixteen years, while Figure 10 indicates that green and blue engineering grade sheeting should be replaced every eight years. According to this study, signs should be checked for retroreflectivity after the age of twelve years.

The development of a sign management system would allow highway agencies, such as NBDOT, to know the age beyond which specific sign types should be checked. As a result, this would enable them to locate deficient signs, use limited funds more efficiently, and estimate future budget needs.

Conclusions

At night a driver's senses are dull because of the monotony of the black-and-white world therefore this is when visibility is crucial. A motorist can only see the road as far as a vehicles' headlights reach. This increases the chances for accidents. Motorists depend on signs for critical information processed during the driving task therefore they must realize the signs are present.

New Brunswick traffic signs at the study onset were thought to be in poor condition with respect to retroreflectivity. This study has shown that retroreflectivity of signs are generally above minimum acceptable standards recently implemented in the United States by the FHWA.

The development of a sign management system would allow highway agencies such as NBDOT to know the age beyond which specific sign types should be checked. As a result this would enable them to locate deficient signs, use limited funds more efficiently and estimate future budget needs.

Recommendations

The development of minimum retroreflectivity values is not an exact process because it involves driver, vehicle, roadway, and sign factors. The recommended replacement values should provide an acceptable level of driver satisfaction

without causing highway agencies to replace a high percentage of signs. Minimum retroreflectivity values should provide agencies with objective values that can be used for implementing a maintenance schedule for traffic signs. These minimum values must also be used with engineering judgement. The user must consider the characteristics at each sign location to determine if the values provided are adequate for motorist sign visibility.

Although FHWA standards for minimum acceptable retroreflectivity were used in this study the applicability of these standards for Canada must be determined. Weather differences between the United States and Canada must also be taken into account.

A much larger area should be tested to estimate provincial traffic sign retroreflectivity and therefore provide a better representation of existing conditions. For example, an industrial city may have less retroreflective traffic signs than a non-industrial city as a result of the pollution.

References

Durand, Michel. Proposal of an international system of highway signs.

Hawkins, Gene H. Jr. 1992. Evolution of the MUTCD: Early editions of the MUTCD. ITE Journal, v 62, no 7, p 23.

Ketola, W.D. 1989. Durability testing for retroreflective sheetings. Transportation Research Record 1230, Transportation Research Board, National Research Council, Washington, D.C. p 67.

Lumia, John J. 1989. Mobile system for measuring the retroreflectance of traffic signs. Transportation Research Record 1247, Transportation Research Board, National Research Council, Washington, D.C. p 69.

Manual on Uniform Traffic Control Devices for streets and highways, 1971. U.S. Department of Transportation, Federal Highway Administration.

Minimum retroreflectivity requirements for traffic signs: summary report, 1993. U.S. Department of Transportation, Federal Highway Administration.

New Brunswick Department of Transportation Sign Shop.

Olson, Paul L. 1989. Minimum photometric properties of retroreflective signing materials. Transportation Research Record 1247, Transportation Research Board, National Research Council, Washington, D.C. p 56.

Paniati, Jeffrey F. 1989. Retroreflectivity research to enhance driver safety. Transportation research news.

Proposal for guide signs on the Interstate Highway system under your jurisdiction, 1958. American Association of State Highway Officials.

Uniform Traffic Control Devices for Canada, 1986. Roads and Transportation Association (RTAC) of Canada, Ottawa.

Colomb, Michele and Michaut, Gerard, 1986. Retroreflective read signs: visibility at night. Transportation Research Record 1093, Transportation Research Board, National Research Council, Washington, D.C. p 58.

Sivak, Michael and Olson, Paul L. 1986. Optimal and minimal luminance characteristics for retroreflective highway signs. Transportation Research Record 1027, Transportation Research Board, National Research Council, Washington, D.C. pp 53-56.

Paniati, J.F. and Schwab, R.N. 1991. Research on the end of life for retroreflective materials: a progress report. Transportation Research Record 1316, Transportation Research Board, National Research Council, Washington, D.C. p 13.

Woltman, H.L. and Szczech, T.J. 1989. Sign luminance as a methodology for matching driver needs, roadway variables, and signing materials. Transportation Research Record 1213, Transportation Research Board, National Research Council, Washington, D.C. p 21.

Hussain, et al. 1989. Effects of light sources on highway sign colour recognition. Transportation Research Record 1213, Transportation Research Board, National Research Council, Washington, D.C. p 27.

Zwahlen, Helmult T., Li, Qi and Yu, Jing 1991. Luminance measurements of retroreflective warning signs at night using the CapCalc system. Transportation Research Record 1316, Transportation Research Board, National Research Council, Washington, D.C. p 31.

Black, Kent L. Hussain, Syed F. and Paniati, Jeffrey F. 1992. Deterioration of retroreflective traffic signs. ITE Journal, vol 62, no 7, pp 16-21.

INTERSECTION ANALYSIS: BOULEVARD DECARIE 'EST' AND RUE JEAN-TALON OUEST

ABSTRACT

This paper presents an analysis of the intersection of boulevard Décarie 'est' and rue Jean-Talon ouest with particular attention to the northbound approach. The analysis of the intersection is based on the procedures set forth in the Highway Capacity Manual Special Report 209, 1985. The northbound approach is unique because it is made up of three service road lanes and one freeway exit lane. The combination of the traffic flow on this approach and its geometric characteristics has led to many accidents. The analysis shows that the intersection is nearing capacity and any increase in flow may cause the intersection to fail.

Five theoretical proposals for improvement in capacity and safety were analyzed. Only two of these proposals proved feasible: increasing the cycle length from 70 seconds to 100 seconds; and the construction of a raised median on the northbound approach to channelize the flow. While changing the cycle length theoretically increases the capacity of the intersection, a raised median could provide greater safety on the northbound approach.

1. INTRODUCTION

Don't you just hate Montréal drivers? They'll cut you off without even looking! The practice of cutting other drivers off is encouraged in Montréal by poor street design. This can be seen on the northbound approach to the intersection of boulevard Décarie 'est' and rue Jean-Talon ouest. Poor geometric design contributes to the number of accidents that occur on this approach. Here, drivers are forced to cross and merge with lanes of fast moving traffic. The objective of this paper is to propose feasible solutions to increase the safety of the

northbound approach without decreasing the capacity of this intersection.

There are several important geometric factors to consider regarding this intersection. First, boulevard Décarie 'est' describes the northbound service road lanes of boulevard Décarie, which is completely separated from the southbound lanes, or boulevard Décarie 'ouest'. Rue Jean-Talon ouest runs east and west. The intersection considered in this paper is that of boulevard Décarie 'est' and rue Jean-Talon ouest. Both of these streets are main arteries carrying heavy traffic flows. Secondly, there is an underpass between rue Jean-Talon ouest and rue Vézina which allows for trains to pass at ground level. The ascending side of the underpass has a grade of 5.50%, which levels off 350 feet (ft) before the intersection. Next, there is an off-ramp for Autoroute Décarie (Highway 15) which merges with the service road on the left of this incline. A raised median, which ends approximately 580 ft before the intersection, separates the exiting Autoroute Décarie traffic from that of boulevard Décarie 'est' service road. Finally, the eastbound approach and the westbound departure of rue Jean-Talon ouest are short overpasses (approximately 180 ft), spanning Autoroute Décarie[1].

Other features surrounding the intersection make it unique. The north-east corner of the intersection is occupied by a subway station (Métro Namur). The southeast corner has the newly opened Mangione Pizza Restaurant and Arby's Restaurant. To the immediate west of the intersection is Autoroute Décarie, a below grade freeway. Further west is the Blue Bonnets race track, with a relatively new parking lot for subway users. These features affect the number of pedestrians crossing at the intersection.

The northbound approach has many points of conflict. Approximately 43 percent of the traffic volume that turns right onto rue Jean-Talon ouest originates from the freeway (see Figure 1). These motorists must cross over two lanes of traffic and merge with a third before entering the intersection. Approximately 57 percent of the motorists turning left emerge from the service road. This

manoeuvre requires the drivers to merge with the exiting freeway traffic in the far left lane. Finally, the through traffic from both the freeway and the service road are approximately equal in volume and must merge safely together. These conflicting movements create a hazardous environment, with an average of five accidents per year on this approach[2]. In addition, these conditions cause much frustration and aggravation for the motorists.

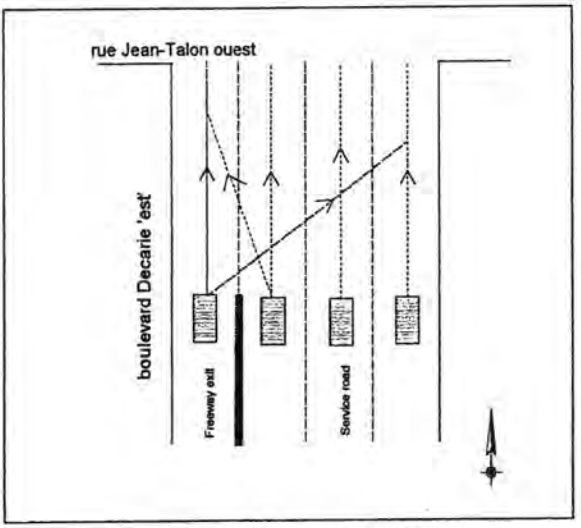


Figure 1 Points of Conflict
on the Northbound Approach

The analysis of the intersection of boulevard Décarie 'est' and rue Jean-Talon ouest was done following the procedures set forth in the Highway Capacity Manual Special Report 209, 1985[3] (HCM). The data for the analyses was supplied by le service de travaux publics, division circulation[2] and section inspection et statistiques[4] de la ville de Montréal. This data was verified by field observations of the phasing plan, traffic volumes and arrival types. The analysis was done using a home-made spreadsheet program following the HCM

guidelines. Subsequent analyses were done in an effort to find alternate solutions to the above mentioned conflicts while maintaining an efficient flow of traffic in the area.

For simplicity, boulevard Décarie 'est' will generally be referred to as Décarie, and rue Jean-Talon ouest as Jean-Talon.

2. INTERSECTION ANALYSIS

The following analysis is based on the procedures set forth in the HCM. The analysis is broken down into five worksheets: Input Worksheet, Volume Adjustment Worksheet, Saturation Flow Adjustment Worksheet, Capacity Analysis Worksheet, and Level-of-Service Worksheet.

2.1 Input Worksheet

The Input Worksheet has a diagram of the intersection with details of its geometry[5]. Boulevard Décarie 'est' is a one way artery heading north with four lanes measuring 11.3 ft each (Figure 2). Three of these lanes are the continuation of the Décarie service road, while the fourth is part of the exit ramp from the

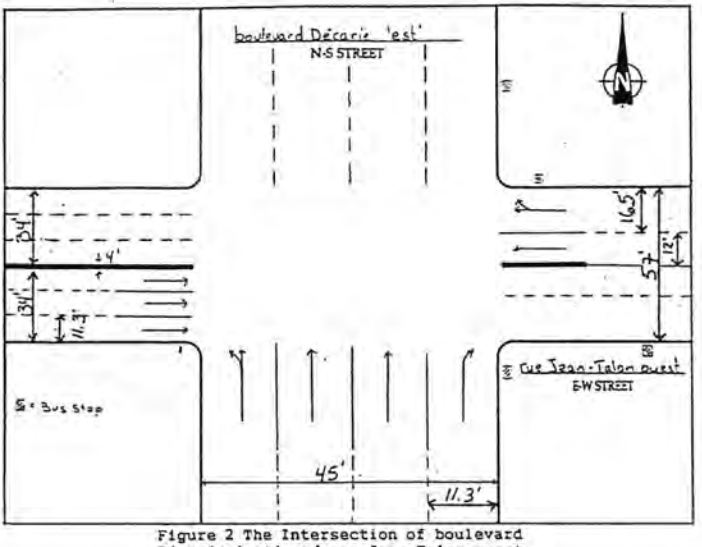


Figure 2 The Intersection of boulevard Décarie 'est' and rue Jean-Talon ouest

freeway. The far right lane is an exclusive right turn (RT) lane with permitted phasing while the far left lane is a shared left turn (LT) lane with permitted phasing. There is no median separating these flows at the intersection. There are two bus stops on Décarie, one on the approach and the other on the departure of the intersection.

Rue Jean-Talon ouest is a two way artery. There are three eastbound lanes measuring 11.3 ft each, which only accommodate through traffic. Westbound Jean-Talon has two marked lanes; one 12 ft wide and the second 16.5 ft wide[1], but motorists usually form queues in three lanes. The right lane is a shared right turn lane, while the others are for through traffic. There are two bus stops, one on the eastbound approach and the second on the westbound departure. There are medians separating the opposing traffic.

The time period analyzed was between 8:00am and 9:00am, which was found to have the highest flow values. Peak hour flows for the northbound, westbound and eastbound streets are 3130 vehicles per hour (vph), 880 vph, and 380 vph[4] respectively.

The second part of the Input Worksheet describes the traffic and roadway conditions. The intersection is on relatively flat ground, i.e. 0 percent grade, with no parking allowed on any of the approaches. Heavy vehicles are incorporated into the flow volumes using equivalent value factors. Approximately seven buses per hour pass on each approach. The flow volumes were observed in the field to be fairly constant and thus a peak hour factor (PHF) of 0.90 is used. Conflicting pedestrian movements are low, below 50 per hour per approach[4].

Flow characteristics are accounted for in the use of Arrival Types. It is approximately 1620 ft (500 m)[1] between Jean-Talon and rue Vézina, which is considered a long distance between intersections. Thus vehicles no longer travel within platoons and the approach is characterized as Type 3. Most motorists using the eastbound approach on Jean-Talon are those who turned left from

boulevard Décarie 'west' (southbound) and therefore arrive in a dense platoon during the red phase, Type 2. The distance between Décarie and avenue Mountain Sights, on the westbound approach, is approximately 377 ft[6], with the lights set in progression. The traffic should then arrive as a dense platoon during the green phase, Type 4.

The intersection has four phases as seen in the third part of the Input Worksheet. Pedestrian movement is allowed for in each phase, with a minimum of 5 seconds (sec) protected movement per direction. The phases are pretimed with a cycle length of 70 sec.

2.2 Volume Adjusted Worksheet

To calculate the adjusted flows per lane group, the Volume Adjustment Worksheet is used. The flow rate v_p is found by dividing the movement volume by the peak hour factor PHF. Next the lane groups are chosen. Two lane groups are designated for the northbound approach isolating the exclusive right turn lane as the first lane group and the through and left turn lanes as the second group. The westbound and eastbound approaches each form lane groups. A lane utilization factor of 1.05 is used for the westbound and through with left turn northbound lane groups; they fall under the category where the most heavily used lane in a group of two serves 52.5 percent of the total flow. By multiplying the flow rates in the lane groups by the lane utilization factor the adjusted flow v is obtained. The proportion of left or right turns per lane group is then recorded[5].

2.3 Saturation Flow Adjustment Worksheet

The Saturation Flow Adjustment Worksheet is used to calculate the flow rate in each lane group that would be able to be achieved if the green phase was always available, i.e. the saturation flow rate. The ideal saturation flow rate is assumed to be 1800 passenger cars per hour of green time per lane[3]. The ideal saturation flow rate is multiplied by a number of adjustment factors found in the HCM tables as follows:

Lane Width Factor (f_w): for the eastbound and northbound approaches the lane widths are approximately 11 ft with a corresponding factor of 0.97. For the westbound approach, the lane widths are averaged giving an equivalent f_w of 1.07. This is done to account for the third unmarked lane that motorists use.

Heavy Vehicle Factor (f_{HV}): this factor is set as 1 for all approaches, heavy vehicles have been accounted for by equivalent factors.

Grade Factor (f_g): the grade on all approaches is approximately 0%; therefore the factor is set at 1.

Parking Factor (f_p): there is no parking allowed on any of the approaches; therefore the factor is 1.

Bus Blockage Factor (f_{bb}): the f_{bb} is dependent on the number of buses stopping per hour and the number of lanes in the lane group. Eastbound, westbound, and northbound right turn lane groups, each having approximately 7 buses stopping per hour, have f_b of 0.99, 0.98, 0.96 respectively. The northbound lane group for through and left turn flow has a factor of 1.00 because no buses stop in these lanes.

Area Type Factor (f_a): the area in question is not in the central business district (CBD) and therefore has a factor of 1.00.

Right Turn Factor (f_{RT}): the westbound approach is categorized as case 5, shared RT lane with permitted phasing. It has a low volume of conflicting pedestrians and the proportion of right turn in the lane group is 0.341. This leads to a f_{RT} of 0.95. Case 2 categorizes the last lane group, exclusive RT lane with permitted RT phasing. The result is a factor of 0.84.

Left Turn Factor (f_{LT}): the left turn factor for the through and left turn northbound lane group is found using the right turn factor tables because there is no opposing traffic. There is insufficient conflicting pedestrian movement to lower the factor below 1.00.

2.4 Capacity Analysis Worksheet

The Flow Ratio, v/s, and the Saturation Flow Rate, v/c, are calculated in the Capacity Analysis Worksheet. The Flow Ratio is the ratio of the actual flow rate for each lane group to the saturation flow rate. There is no overlapping of phases, and thus the critical lane groups are those with the highest flow ratios per phase. In this case, the critical flow ratios are 0.286 for the westbound approach and 0.564 for the through and left turn lane group on the northbound approach. The sum of critical lane group flow ratios, $S(v/s)_{ci}$, is 0.850 (<0.90, o.k.). The green ratio, g/C, is the ratio of green time to the total cycle length. The effective green time is assumed to equal the actual green time. The lane group capacity, c, represents the maximum rate of flow that may pass through the intersection under the present geometric and signalization conditions. It is found by multiplying the adjusted saturation flow rate by the green ratio.

The value of the v/c ratio ranges from 0.00 when the flow rate is zero, to 1.00 when the flow rate equals capacity. The northbound lane groups are both over capacity, with v/c ratios exceeding 1.00. Eastbound and westbound Jean-Talon are operating below saturation, with v/c ratios of 0.247 and 0.868 respectively. The critical v/c ratio, 0.991, is dangerously high. This ratio indicates, as observed in the field, that the intersection undergoes several breakdowns (failure to accommodate flow) during peak hours and any future increases in flow volumes may cripple the intersection. The cycle length, the signal phasing plan and the geometric conditions are factors to be reevaluated.

2.5 Level-of-Service Worksheet

The final worksheet is used to find the Level-of-Service (LOS) of the individual lane groups and the intersection as a whole. In this worksheet, the average stopped-time delay is estimated for each lane group. This delay takes into account uniform delay, d_1, and incremental delay, d_2. The longest d_1 delay experienced is that for westbound Jean-Talon, a delay of 16.7 sec, where as the

longest d_2 delay is 28.1 sec experienced by the northbound exclusive right turn lane. Adding the delays together and multiplying them by the progression factor, to account for the impact of progression on delay, gives the lane group delay. Northbound lane groups have the highest delays with 43.3 and 25.6 sec/veh for the right turn only and the through and left turn lane groups respectively. The lane group LOS is found by converting these values into a representative letter using Table 1.

LOS	Stopped Delay per Vehicle (sec)
A	≤5.0
B	5.1 to 15.0
C	15.1 to 25.0
D	25.1 to 40.0
E	40.1 to 60.0
F	≥60.1

Table 1 Level-of-Service for Signalized Intersections[3]

The approach delays are found by multiplying the approach flow rates by the delays and dividing by the flow rates. LOS values are then found using Table 2. Eastbound, westbound and northbound approach LOS values are C, C, D respectively. Steps should be taken to try to lower the LOS for the northbound lane groups for they come close to being unacceptable. Any future increases in flow rates will cause failure of this approach. The intersection delay is 26.1 sec/veh and the intersection LOS is D. Once again, this creates unacceptable delays and may give rise to complete intersection failure if the demand increases slightly.

3. FIVE PROPOSALS FOR IMPROVEMENT

The analysis of the present situation shows that the intersection design and phasing are inadequate. Five proposals that make use of changes in the geometry and/or the phasing plan of the intersection have been analyzed using the HCM procedure. The first solution involves simply increasing the cycle length to 100 seconds. In the second proposal, the next lane over from the exclusive right turn lane would be designated as a through and right turn lane. The next three solutions involve extending the raised median up to the intersection on the northbound approach to physically separate the service road from the freeway exit. The first of the options with an extended median uses the present phasing plan, the second uses the cycle length of 100 sec and the third uses split phasing.

3.1 Increased Cycle Length

The first proposal deals with increasing the cycle length of the intersection. The minimum cycle length, C, is calculated to be 100 seconds. Therefore, according to the HCM, the present cycle length of 70 sec is too short. The new green ratios, g/C, calculated for the eastbound, westbound, northbound through and left turn, and northbound right turn are 0.33, 0.33, 0.57, 0.50 respectively. The saturation flow ratios, v/c, for the northbound lane groups and the critical saturation flow rate, X_c, have decreased slightly with the increase in cycle length, but the intersection is still functioning very close to capacity. The LOS for the eastbound and westbound lane groups have declined to the level of D, while a LOS of D was achieved up from an LOS of E for the northbound exclusive right turn lane. The LOS of the other northbound lane group as well as the intersection as a whole stayed at D[5].

3.2 A Second Lane for Permitted Right Turns

The second alternative is to allow for through traffic and permitted right turns from the lane next to the exclusive right turn lane. A right turn factor of 0.98[7] is used to account for the decrease in capacity of the lane group with the

permitted right turns. Although this alternative would help alleviate some of the stress on the vehicles turning right, it increases the problems for the through traffic. The v/c ratio for the through lane group is over 1.00, over capacity, for both cycle lengths of 70 sec and 100 sec. Even though the LOS of the lane group in question remains at D, the numerical value points to the higher end of the D level, nearing on E. These theoretical results are unsatisfactory and therefore will not be discussed further.

3.3 A Change in Intersection Geometry

The next three proposals deal with a change of geometry for the northbound approach. They consist of building a raised median that extends from the freeway up to the intersection. The median would taper the one lane exit ramp from the initial width of 20 ft to 14 ft at the intersection. The median itself would taper to a point of 1 ft. The service road would consist of three lanes each 10 ft wide. For these analyses, the service road lanes will be referred to as NB and the freeway exit lane will be referred to as NB' (see Figure 3).

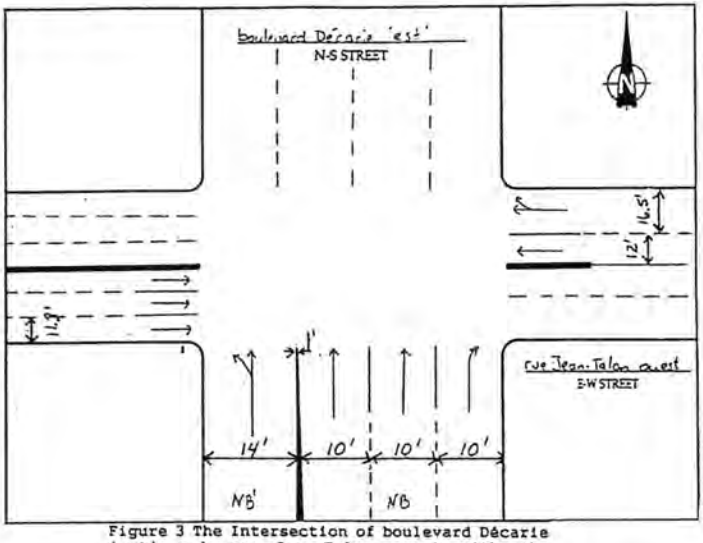


Figure 3 The Intersection of boulevard Décarie 'est' and rue Jean-Talon ouest with the Proposed Raised Median

3.3.1 Using the Present Phasing Plan

The third proposal uses the same cycle length and phase plan that presently exists, but traffic movements are greatly restricted, as seen in Figure 3. The NB' lane is designated as a shared LT lane with permitted phasing. Two of the NB lanes remain as through lanes and the third is an exclusive RT lane with permitted phasing. The HCM method is used once again to analyze this geometric layout and phase plan. The Capacity Analysis Worksheet and the Level-of-Service show the results of these calculations[5]. The flow ratios for the NB and the NB' are well below 1.00 which is a good start. The NB v/c ratios are also within reasonable levels, but the NB' lane group has a v/c ratio above 1.00 which means that it is exceeding capacity, and long queues may form. The intersection's critical flow ratio is above 1.00 and so is the critical saturation flow rate, indicating that this design is inadequate to handle the demands. On the other hand, the Level of Service ratings are good for this geometry and phasing plan. The NB lane groups have an LOS of B and the NB' has an LOS of D, whereas the intersection as a whole has an LOS of C (better than the present LOS of D).

3.3.2 Increased Cycle Length

The next proposal examines the use of a 100 sec cycle length instead of the original cycle length of 70 sec. The v/s ratios are unaffected by the change in cycle length. The eastbound and westbound saturation flow rates remain the same in these cases, but the v/c for NB and NB' decrease slightly. The Level-of-Service for the westbound and eastbound lane groups decreases from C to D, while the other lane groups experience very little change with the longer cycle length.

3.3.3 Split Phase Plan

The last proposal uses a different phasing plan. It consists of a separate phase for the NB and NB' approaches allowing for separate movements. Assuming the lost time for this phase plan is 15 sec (3 yellow phases of 5 sec

each), a cycle length of 150 sec is needed. The maximum cycle length allowed in regular circumstances is 120 sec. This value is used in the calculations. As before, the green times in each lane group are calculated having the following results: g_{EB} and g_{WB} are 33 sec (using X_i=0.247 and 0.868 respectively as in the previous cases), g_{NB} is 76 sec (assuming X_i=0.90), and $g_{NB'}$ is 64 sec. A total cycle length of 188 sec is found as a result of the three independent phases. Therefore, this combination of geometry and phasing is not acceptable because the cycle length needed is too long to be feasible and will no longer be discussed.

4. DISCUSSION

Only two of the proposals seem feasible: the increase in cycle length to 100 sec from 70 sec without any changes in geometry; and the extension of the median dividing the NB and NB' traffic with the present phasing plan. The following is an analysis of traffic flow patterns and safety considerations which may develop due to changes in the geometric and phasing plans for the intersection.

Increasing the cycle length of the intersection, as mentioned earlier, increases its capacity and decreases delay time. This is a relatively low cost solution which only requires that the pretimed control signals be adjusted for progression. On the other hand, it should be noted that there is only 180 ft between the westbound departure and the next signalized intersection. With the increase in cycle length, the number of vehicles turning left onto Jean-Talon from Décarie per cycle would increase. Traffic would likely back up on this overpass and interfere with the northbound flow.

The problem of safety on the northbound approach is not addressed with this solution. Motorists would still have to cross and merge with through and turning traffic to get to their destinations. Two types of accidents are frequently seen on this approach, rear-end and side collisions. As seen in Figure 4, a rear-end collision may be caused by a motorist, car B, who is forced to slow down or

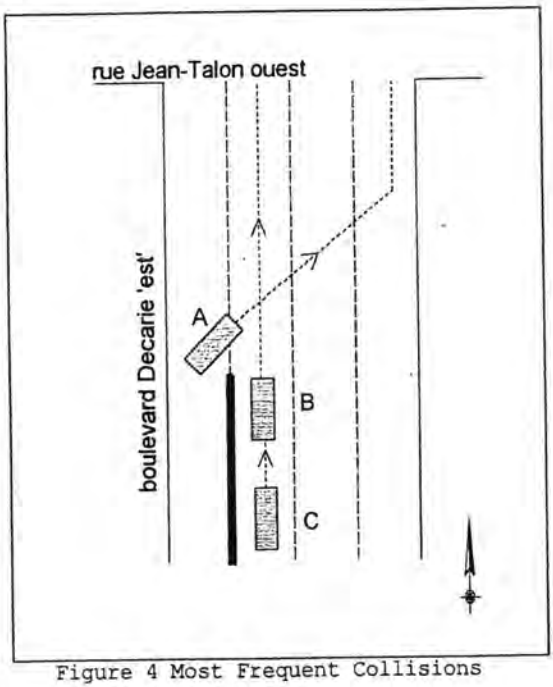


Figure 4 Most Frequent Collisions

stop in the middle of the street by a vehicle exiting from the freeway, car A, and is then hit in the rear by a tailing motorist, car C. Side collisions can be attributed to the motorists exiting from the freeway, car A, and trying to cut over to the right hand lane. They can be hit on the passenger side by motorist driving on the service road, car B, who cannot anticipate their actions. Changing the cycle length will not reduce the number of accidents on this approach, therefore geometric changes should be considered.

The second solution deals with a change in the geometry of the northbound approach (see Figure 3). A raised median would extend from the freeway to the intersection and would channelize the flow. At rush hour, under present conditions, the northbound exit at Jean-Talon experiences long queues of traffic onto the freeway. These queues can mainly be attributed to the motorists

who stop their vehicles at the end of the existing median while waiting for a window to appear in the service road traffic, which is a dangerous practice. The new geometric design suggested here resembles that of the Autoroute Décarie southbound exit at rue Paré. At rush hour, traffic is backs up considerably on both the southbound service road and the freeway, and this is admittedly undesirable. But looking at it from a different point of view, there are less accidents per year at this intersection as compared to that of the Décarie/Jean-Talon intersection[4]. Increasing the delay slightly (if at all) by adding the median would be a worthwhile exchange for decreasing the number of accidents.

Traffic flow patterns in the area would be greatly affected by this geometric layout. The northbound traffic would have limited movements at the intersection. The NB' motorists wanting to head east and the NB motorists wanting to head west must continue straight through the intersection and cross over on the stretch between Jean-Talon and rue Paré. This merging of traffic is safer than the present conditions for it occurs on relatively flat ground with good visibility.

One of the problems mentioned earlier with the 100 sec cycle length is that there will be congestion on the short westbound departure. This could be alleviated with the new median because the number of left turning vehicles would decrease due to greater limitations on movement. On the other hand, the volume of traffic on the overpass at rue Paré, already a busy intersection, would increase.

Another consequence of the channelized flow would be the increase in traffic on rue Paré, east of Décarie. The traffic which would be restricted from turning directly onto Jean-Talon from Décarie, would turn right onto rue Paré (which turns into Buchan) and then right onto either avenue Mountain Sights or avenue Victoria in order to return to Jean-Talon. These three streets do not have the capacity to handle the extra flow in their present state. Avenue Mountain Sights is a local road used extensively to access the Société de Transport de la

Communauté Urbaine de Montréal (STCUM) bus depot at the corner of Jean-Talon and avenue Mountain Sights. The intersection at Buchan and avenue Victoria already has a confusing geometric design and the added flow would cause even more hazardous conditions. If the geometry of the northbound approach of the Décarie/Jean-Talon intersection is changed, then the Buchan/avenue Victoria intersection will have to be analyzed and redesigned as well.

5. CONCLUSION

The intersection of boulevard Décarie 'est' and rue Jean-Talon ouest is operating very close to capacity, and the northbound approach exceeds capacity. Small increases in the volume of traffic may cause the intersection to fail, creating long delays for motorists.

The geometric layout of the northbound approach encourages drivers to cut across several lanes of traffic on an incline with poor visibility. This poor design has lead to many accidents. Improvements should be made.

Five proposals were set forth, of which only two are feasible. Increasing the cycle length from 70 sec to 100 sec would slightly increase capacity but will not improve the safety of the northbound approach. Constructing a raised median would improve safety, but it would increase delay for the traffic exiting from the freeway. Channelizing the flow and restricting turning from the northbound approach would make it necessary to reevaluate the geometry and phasing plan of at least four more intersections to account for the increase in traffic volume.

LIST OF REFERENCES

1. Module technique, profils et domain public, le service de travaux publics, ville de Montréal (1965). Plan no 8, route Transcanadienne, autostrade Décarie. Canada.

2. Division circulation, le service de travaux publics, ville de Montréal (1993). Sommaire complet des accidents a l'intersection des rues Décarie est et Jean-Talon. Canada.

3. Transportation Research Board (1985). Highway Capacity Manual, Special Report 209. National Research Council. U.S.A.

4. Section inspection et statistiques, le service de travaux publics, ville de Montréal (1991). Inventaire de la circulation, comptage vehicules. Canada.

5. Wiseman, S. (1995). Intersection Analysis: Boulevard Decarie 'Est' and Rue Jean-Talon Ouest. McGill University. Canada.

6. Module technique, profils et domain public, le service de travaux publics, ville de Montréal (1984). Plan no J6, Côte-des-Neiges. Canada.

7. Hirou, C. (March 1994). McGill University lecture. Canada.

APPLICATION DE LA MÉTHODE DE RECHERCHE AVEC TABOUS AU TRACÉ D'UNE LIGNE DE MÉTRO

par
HÉLÈNE DUFOURD

Article tiré du M. Sc. A du
DÉPARTEMENT DE MATHÉMATIQUES
ET DE GÉNIE INDUSTRIEL
ÉCOLE POLYTECHNIQUE DE MONTRÉAL
en collaboration avec le
CENTRE DE RECHERCHE SUR LES
TRANSPORTS (CRT)
NOVEMBRE 1994

© Hélène Dufourd, 1994.

1. INTRODUCTION

Le problème proposé

En transport urbain, il existe environ quatre-vingt-dix systèmes de métro au monde, et des projets de construction sont en cours dans un nombre important de grandes villes. D'autre part, de nombreuses villes ont opté pour la construction de trains aériens rapides ('RRT' pour 'Rail Rapid Transit'), de systèmes de trains légers ('LRT' pour 'Light Rail Transit'), ou de monorails. Le but principal de ces systèmes de transport est d'enrayer la congestion routière causée par les automobiles.

De telles installations nécessitent de gros investissements en capital ainsi que des coûts d'opération élevés, d'où l'importance d'effectuer un processus de planification long et complexe. Les décisions doivent porter sur le type de système de transport à adopter, les techniques de construction, la détermination du réseau, les connections avec les réseaux déjà existants et les installations périphériques, etc. Ces décisions sont intimement reliées et dépendent d'une multitude de critères. Certains critères de décision sont quantifiables: la population potentiellement recouverte par le réseau, le nombre de paires origine-destination parcourues par les usagers couvert, le coût de réalisation et d'opération du réseau. Il existe également un éventail de facteurs difficilement quantifiables: la qualité de vie entourant le réseau (niveaux de bruit et de qualité de l'air, esthétique), la réduction des temps de parcours des usagers, l'impact économique du réseau dans les zones touchées, la sécurité, etc. Il est presque impossible de tenir compte de tous ces facteurs. Cependant, la recherche opérationnelle peut être un outil précieux et assister de façon pertinente le processus de planification. Le cœur du problème, l'alignement et la localisation des stations d'une ligne de transport en commun, est essentiellement un problème d'élaboration de réseau. Or, de nombreux travaux (voir par exemple Magnanti et Wong [6]) traitent de la question. Avec les principaux critères quantifiables de décision fournis, une méthode de recherche opérationnelle peut être utilisée et fournir des familles de solutions prometteuses qui répondront aux objectifs du planificateur. Dans le cadre de cet article, nous proposons d'examiner l'étape de la détermination du réseau proprement dit, à l'aide de certains critères pertinents. À la section 2, nous modéliserons le problème. Ensuite, nous exposerons à la section 3 une méthodologie qui permettra de résoudre le modèle proposé: la méthode de recherche avec tabous. Nous illustrerons l'application de cette méthode au modèle proposé au moyen d'une série de problèmes-tests, à la section 4. Nous conclurons enfin par une discussion des résultats obtenus dans cet article.

2. LA MODÉLISATION DU PROBLÈME

2.1 Le problème proposé

Le problème que nous proposons est la détermination de la localisation des stations d'une ligne de transport par trains rapides, légers, aériens ou souterrains. Pour simplifier l'écriture, nous désignerons cette ligne de transport en commun par "ligne de métro".

Cette ligne pourra être ajoutée à un réseau déjà existant, ou former la base d'un nouveau réseau, selon les besoins du planificateur. Trois points doivent être considérés lors de la modélisation du problème: le territoire étudié, l'objectif recherché, les contraintes à respecter. Etant donné que nous ne disposons pas de données réelles en ce qui concerne le territoire sur lequel la ligne sera tracée, nous allons nous-mêmes simuler une population d'individus et sa répartition. Cette population sera encadrée de limites finies. Cet ensemble formera le *territoire étudié* , dont le description est faite à la section 2.1.1. Le problème de localisation des stations d'une ligne de métro est un vaste problème multicritères. Cependant, nous pouvons sélectionner quelques critères de décision pertinents lors de la modélisation du problème. Ainsi, nous croyons que la population recouverte par la ligne de métro est un facteur essentiel dans la détermination de l'efficacité de ce système de transport. Une ligne qui recouvre le plus possible d'individus est, selon nous, l'objectif à viser. Le recouvrement de population est un critère quantifiable. Nous expliquerons en détail la façon de quantifier ce critère à la section 2.1.2. Les coûts de construction et d'opération associés à la ligne de métro sont considérables et doivent être pris en compte tôt dans le processus de planification. Aussi, pour Allport et Brown [1], plus la vitesse du train est élevée, moins grands sont les temps de parcours des passagers, et plus grands sont les gains de productivité. Ces deux critères – les coûts et les gains – sont quantifiables. La section 2.1.3 décrit en détail le cheminement entrepris pour transformer ces deux critères de décision en contraintes sur la ligne de métro.

2.1.1 Le territoire étudié

Le territoire d'une ville est d'abord entouré de frontières bien déterminées. Dans notre simulation, le territoire est entouré de limites formant un carré de cent par cent unités de longueur. Ces unités de longueur n'ont pas de valeur réelle; elle peuvent être des pieds, des kilomètres, etc. À l'intérieur de ces limites, nous subdivisons le territoire de façon à former un treillis. Nous ne considérons que les valeurs entières des coordonnées, pour chaque point du treillis. La population présente sur le territoire est agrégée sur chaque point à coordonnées entières du treillis. Nous discrétisons donc le problème. Pour déterminer la population présente sur le territoire, nous assignons à chaque point du treillis un score numérique représentant un nombre d'individus, selon la méthode suivante:

1. Générer aléatoirement des carrés de dix par dix unités de longueur sur tout le treillis, ayant comme poids respectifs des scores numériques de 1, 5, et 10 individus à assigner sur chaque point du treillis visité.
2. Répéter la procédure 25 fois, en permettant la superposition de carrés s'il y a lieu.
3. Assigner aux points non visités un score numérique de 0.

Les côtés de ces carrés ne peuvent être placés que sur les coordonnées entières du treillis. La population présente aux points sur lesquels plusieurs carrés se sont superposés est la somme des poids de ces carrés. La figure 2.1 illustre la méthode utilisée pour assigner à chaque point du treillis un score numérique représentant le nombre d'individus.

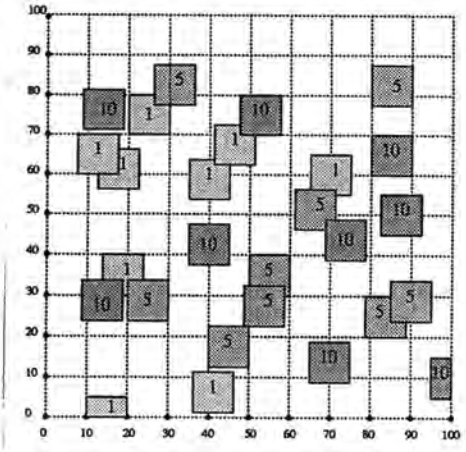


Figure 2.1: Assignation de la population

2.1.2 L'objectif recherché

L'objectif recherché est de recouvrir le plus possible d'individus sur la ligne de métro. La population recouverte par la ligne de métro est la somme des populations recouvertes par chacune de ses stations. La population recouverte par une station regroupe non seulement la population présente sur le point sur lequel est localisée la station, mais aussi une série de points qui entourent la station. Les points considérés sont tous les points tels que leur distance Manhattan par rapport à la station n'excède pas $dmax$ unités de longueur, où $dmax$ représente la distance maximale qu'un individu peut parcourir à pied de son lieu d'origine à la station. Pour Lutin et Benz [5], plus de 90% des usagers marchent moins de 6 minutes pour se rendre de leur point d'origine à la station et 50% d'entre eux marchent moins de 3 minutes. Un individu localisé à une distance de 3 unités de longueur de la station est moins porté à utiliser la ligne de métro qu'un individu localisé plus près de la station. Nous affecterons donc à la population divers facteurs de multiplication selon la distance à laquelle elle se trouve de la station. La population recouverte par une station est calculée de la facon suivante:

1. Représenter le premier terme par le score numérique (la population) du point sur lequel est localisé la station, affecté du facteur θ_0.
2. Ajouter au premier terme les scores numériques des points distants de $1, 2, ..., dmax$ unités de la station, affectés respectivement des facteurs de multiplication $\theta_1, \theta_2, ..., \theta_{dmax}$.

Géométriquement, l'ensemble des points distants de 0 à $dmax$ unités de longueur d'une station localisée au point (x, y) forme une figure de $dmax$ losanges. La figure 2.2 montre

l'ensemble des points admissibles au calcul du recouvrement pour une distance d où d varie de 0 à $dmax$ et où, dans notre exemple et dans nos tests numériques, $dmax$ est fixée à 3 unités de longueur.

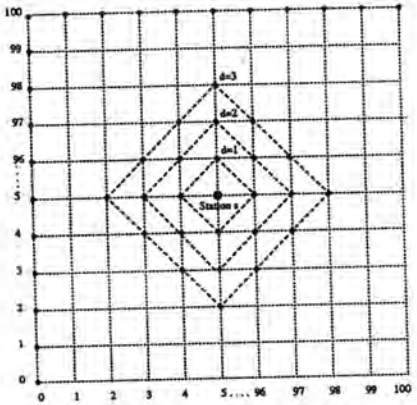


Figure 2.2: Points retenus lors du calcul du recouvrement

Le recouvrement de la station s est alors défini par:

$$R(s) = \sum_{d=0}^{dmax} \sum_{(x,y):D[(x,y),s]=d} p(x,y)\,\theta_d, \tag{2.1}$$

où $D[(x,y),s]$ est la distance Manhattan entre (x,y) et s, et $p(x,y)$ est la population au point (x,y). Lors de nos tests numériques, les facteurs de multiplication θ prendront les valeurs rapportées au tableau 2.1.

Tableau 2.1: Facteurs de multiplication

θ	Valeur
θ_0	1,00
θ_1	1,00
θ_2	0,50
θ_3	0,25

2.1.3 Les contraintes à respecter

Lorsque l'étape de réalisation de la ligne de métro est en marche, des coûts de construction sont impliqués. Ces coûts peuvent être amortis sur une longue période. Pour chaque année de service de la ligne de métro, des coûts d'opération sont impliqués, ainsi que des coûts reliés à un éventuel prolongement du réseau. On peut regrouper ces éléments en un coût global par année, qui représente la somme des amortissements et des frais courants.

En regardant certains aspects de la ligne de métro comme les distances inter-stations et le nombre de stations, on peut introduire cette notion de coût global. En effet, il est raisonnable de penser que les coûts sont proportionnels à la longueur des tunnels et au nombre de stations. Pour tenir compte de ceux-ci, nous introduisons les contraintes suivantes:

1. Limiter la distance Manhattan entre deux stations voisines l'une de l'autre à l_{min} et l_{max} unités de longueur, où l_{min} et l_{max} représentent les distances Manhattan minimales et maximales permises, respectivement.
2. Fixer a n le nombre de stations sur la ligne de métro.

En limitant à l_{max} unités de longueur la distance maximale entre deux stations voisines l'une de l'autre, on ajoute de l'efficacité à la ligne, car si les stations étaient trop éloignées les unes des autres, la population serait mal desservie.

Du côté des gains, pour optimiser l'efficacité de la ligne de métro, la distance entre deux stations voisines l'une de l'autre est fixée à au moins l_{min} unités de longueur. Cette restriction permet d'économiser sur le nombre de stations, et elle est essentielle, car le coût de construction d'une station est considérable. Aussi, une ligne dont les stations seraient trop rapprochées les unes des autres ne recouvrerait pas un territoire assez grand. Enfin, cette restriction permet au métro d'atteindre des pointes de vitesse, et de fournir ainsi aux usagers de bons temps de parcours.

Le paramètre l_{min} peut être insuffisant dans le cas où on aurait un budget serré. En effet, on ne veut pas nécessairement d'une longue ligne sur laquelle toutes les stations sont espacées les unes des autres de l_{min} unités de longueur. Le fait de limiter à n le nombre de stations permises sur la ligne de métro est donc un complément au paramètre l_{min}, pour les coûts relatifs aux stations.

Lors de nos tests numériques, les paramètres l_{min} et l_{max} prendront les valeurs de 8 et de 16 unités de longueur respectivement, alors que le paramètre n sera fixé à 15 stations. Avec une valeur de 8 pour le paramètre l_{max}, il n'y aura pas d'interférence entre deux stations dans le calcul de leur recouvrement.

2.2 L'énoncé du problème

Maintenant que le territoire étudié a été décrit et que les critères de décision ont été exprimés en un objectif et en contraintes, nous avons une modélisation complète du problème. Dans la formulation de l'objectif, nous allons donc étudier la localisation des stations d'une ligne de métro de façon à maximiser le recouvrement de population. Le problème consiste à placer 15 stations sur un territoire représenté par un treillis de 100 par 100 unités de longueur, où seuls les points à coordonnées entières sont considérés. Chaque station est localisée sur un point du treillis. La meilleure ligne de métro sera celle ayant l'objectif le plus élevé. Cette ligne aura comme particularités qu'entre deux stations voisines l'une de l'autre, la distance Manhattan ne sera jamais plus petite que 8 unités de longueur, et jamais plus grande que 16 unités de longueur.

Cette modélisation a un ensemble discret de solutions possibles. L'accent est mis sur la *localisation* des stations de la ligne de métro. Une fois les stations localisées, il nous reste à les relier les unes aux autres, pour obtenir un chemin qui constituera le tracé final de la ligne de métro. Dans cet article, compte tenu du faible nombre de stations, les tracés ont été réalisés manuellement.

3. LA MÉTHODE DE RÉSOLUTION DU PROBLÈME PROPOSÉ

3.1 La méthode de recherche avec tabous

3.1.1 Introduction

Le principe des heuristiques d'exploration locale est le suivant: à partir d'une solution initiale réalisable, on applique itérativement des modifications locales à la solution courante, tant que celles-ci permettent d'améliorer la valeur de l'objectif. L'inconvénient de ces heuristiques est que la recherche est bloquée par les optima locaux: dès que la valeur de l'objectif diminue (dans un problème de maximisation), la procédure prend fin. Or, la valeur des optima locaux peut être très éloignée de celle du véritable optimum. La puissance de ces approches est donc limitée. La méthode de recherche avec tabous a été créée dans le but d'éliminer cet inconvénient. Introduite par Glover [3,4], la méthode de recherche avec tabous est une métaheuristique: c'est une procédure utilisée pour explorer l'espace des solutions d'un problème donné, associée à une heuristique interne d'amélioration locale. Le principe de base de cette méthode est la permission de poursuivre la recherche de meilleures solutions, même si l'on a une *dégradation* de l'objectif. Cette méthode permet alors de "passer par dessus" les optima locaux et d'aller trouver de meilleures solutions. Nous croyons que la méthode de recherche avec tabous est tout à fait appropriée à notre problème: elle est souple, permet de considérer des objectifs complexes ainsi que diverses contraintes, et peut produire un éventail de solutions intéressantes. Dans notre contexte, on construira d'abord une ligne initiale que l'on modifiera graduellement à l'aide de déplacements de stations vers des emplacements voisins, dans l'espoir de découvrir une solution de bonne qualité. En partant de plusieurs lignes initiales différentes, on peut facilement produire une grande variété de solutions finales intéressantes.

3.1.2 La notion de "mouvement tabou" et de "liste taboue"

Le problème fondamental de la méthode de recherche avec tabous est le suivant: si l'on permet une dégradation de l'objectif, comment faire pour empêcher la recherche de revenir sur ses pas à l'itération suivante, dans le cas où aucune solution réalisable n'est meilleure que celle de l'optimum trouvé jusqu'à maintenant? La réponse se trouve dans la notion des *mouvements tabous*. En interdisant à la recherche de retourner sur les optima locaux trouvés durant un certain nombre d'itérations, on évite à la recherche de se bloquer. Ces

interdictions sont appelées les mouvements tabous. On espère ainsi voir la recherche se diriger vers d'autres solutions réalisables plus intéressantes que les premières. Une fois le risque de blocage passé, on relâche l'interdiction portée sur ces mouvements et on l'applique sur les mouvements suivants. Bref, les mouvements tabous sont une mémoire à court terme de la recherche.

Ces mouvements tabous sont organisés de façon structurée. On garde en main la liste des dernières transformations effectuées durant la recherche, et on interdit alors d'effectuer les transformations inverses durant un certain nombre d'itérations. Cette liste en mémoire est appelée la *liste taboue*. Elle contient un nombre m d'éléments définissant des mouvements tabous; le paramètre m est appelé la taille de la liste taboue. Le choix de la taille de la liste taboue est critique: si sa valeur est trop petite, un cyclage peut apparaitre; si sa valeur est trop grande, des mouvement attrayants peuvent être laissés de côté, amenant l'exploration à des solutions de moins bonne qualité. Dans notre cas, la durée de validité de chaque mouvement tabou est déterminée séparément et ne change pas par la suite, à moins que le mouvement inverse ne soit impliqué à nouveau. Il n'y a donc pas réellement de liste taboue, car tous les mouvements tabous sont indépendants les uns des autres. Pour le choix de la durée de validité m du mouvement tabou, nous proposons de prendre un nombre aléatoirement dans l'intervalle $[m_{min}, m_{max}]$. Lors de nos tests numériques, on prendra $m \in [5, 10]$.

3.1.3 La notion d'"aspiration"

Les mouvements tabous sont un mécanisme anti-cyclage qui peut s'avérer trop puissant:

1. En interdisant des mouvements attrayants alors qu'il n'y a pas de risque de cyclage;
2. En provoquant à long terme une stagnation de la recherche.

On introduit alors un mécanisme inverse qui permet de lever un mouvement tabou s'il n'y a pas de risque de cyclage: la *"fonction d'aspiration"*. Il existe plusieurs sortes de fonctions d'aspiration. Celle que nous utiliserons consiste à lever un mouvement tabou si cela permet d'améliorer la valeur de la meilleure solution trouvée jusqu'à ce point.

3.2 L'application de la méthode de recherche avec tabous au modèle proposé

3.2.1 Le voisinage d'une station et le domaine des solutions

Dans notre modèle, la ligne de métro est formée de n stations. Une *solution* est un ensemble d'emplacements des stations de la ligne de métro. Le principe de l'exploration est de changer les emplacements d'une station à la fois, dans le but d'améliorer l'objectif. Un échange entre les coordonnées de la station et d'un de ses voisins est alors effectué. La nouvelle solution comprend alors les emplacements des $n-1$ autres stations et l'emplacement

du voisin choisi. Une itération a été complétée. Dans notre représentation du territoire par un treillis, chaque station de la ligne de métro admet au plus quatre voisins distants d'une unité de longueur de celle-ci: à droite (voisin "1"), à gauche (voisin "3"), en haut (voisin "4") et en bas (voisin "2"). La figure 3.1 illustre le voisinage d'une station. Au

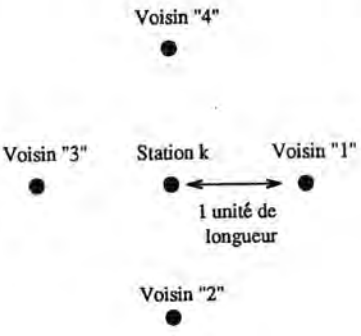


Figure 3.1: Le voisinage de la station k

total, la ligne de métro, dont le tracé initial est expliqué à la section suivante, admet au plus $4n$ voisins. Nous allons maintenant établir le domaine des solutions possibles pour l'exploration locale. Que l'on soit au début ou au cours de l'exploration, la solution courante S est:

$$S = \{s_i,\ i = 1, ..., n\}, \tag{3.1}$$

où s_i sont les emplacements des stations. Connaissant les emplacements des stations de la ligne à l'itération t, on détermine les voisins v_{ij} admissibles de chaque station. Ces voisins doivent satisfaire les contraintes de distances inter-stations. On a donc:

$$d(v_{ij}(S), s_{i+1}) \leq l_{max} \quad i \neq n; \tag{3.2}$$

$$d(v_{ij}(S), s_{i-1}) \leq l_{max} \quad i \neq 1; \tag{3.3}$$

$$d(v_{ij}(S), s_l) \geq l_{min} \quad l = 1, ..., n \text{ et } l \neq i. \tag{3.4}$$

Si le voisin "l" de la station k est candidat au remplacement de cette dernière, alors la solution possible correspondante est:

$$S_{kl}(S) = S \setminus \{s_k\} \cup \{v_{kl}(s)\} \tag{3.5}$$

et le domaine des solutions possibles à l'itération t est:

$$DOM(t) = \{S_{kl}(t),\ \ k = 1, ..., n \text{ et } l \text{ admissible pour } k\}. \tag{3.6}$$

Le recouvrement total de la solution S est:

$$z(S) = \sum_{\zeta} \qquad R(s_i). \tag{3.7}$$

3.2.2 Les tabous

Tel qu'indiqué auparavant, à chaque itération, la nouvelle solution courante est obtenue en déplaçant dans la solution précédente une seule station vers un de ses voisins. Ce mouvement est défini par deux composantes: le numéro de la station déplacée (1, 2, ..., ou n), et la position du voisin par rapport à la station (position 1, 2, 3, ou 4 selon que le voisin est placé à droite, en bas, à gauche ou en haut de la station). Si, par exemple, la station k est déplacée vers son voisin "1" de droite, alors on assiste à un mouvement de composantes $[k][1]$. On établit ensuite le voisinage immédiat du voisin choisi: son voisin de gauche prendra alors l'ancienne valeur de la station et les autres valeurs des voisins devront être calculées. Le numéro de cette station demeurera à k. Le mouvement inverse (le mouvement $[k][3]$ vers la gauche) sera alors interdit pour m itérations, où $m \in [m_{min}, m_{max}]$. À l'itération suivante, même si le mouvement $[k][3]$ est le plus attrayant, la recherche se dirigera automatiquement vers un autre voisin, si le critère d'aspiration ne s'applique pas, et si le déplacement vers ce voisin n'est pas tabou, la transformation sera permise et le mouvement inverse deviendra tabou lui-aussi. Au bout de m itérations, le mouvement $[k][3]$ pourra à nouveau être effectué, si c'est celui-ci qui améliore le plus la fonction objectif. Le domaine est donc réduit par la quantité des mouvements tabous et s'agrandit lorsque l'un de ceux-ci est annulé. On peut représenter la liste taboue par une matrice de longueur n par 4, où chaque élément de la matrice représente le nombre d'itérations pour lesquelles le mouvement est tabou. En pratique, lorsqu'un mouvement devient tabou, on assigne à la cellule correspondante de la matrice de la liste taboue la somme du compteur d'itérations courant et du paramètre m. Lors de la recherche, on compare la valeur du compteur d'itérations actuel à celle de la cellule du mouvement désiré. Si cette dernière est plus grande que la valeur du compteur d'itérations, alors le mouvement est interdit.

3.2.3 La procédure de recherche avec tabous pour le problème proposé

La procédure que nous proposons comprend une méthode de recherche avec tabous, assortie d'un critère d'aspiration permettant d'accepter une solution taboue si son recouvrement est meilleur que celui de la meilleure solution trouvée jusqu'à maintenant. De plus, la durée de la procédure va dépendre de la meilleure solution trouvée jusqu'à maintenant. En effet, on arrêtera la procédure quand on aura effectué $Maxiter$ itérations sans améliorer la valeur de l'objectif. En pratique, on réalise ceci en augmentant la valeur du nombre maximum d'itérations permis, appelé $Derniter$, à $Maxiter + t$, où t est le compteur, dès que l'objectif est amélioré. Soit t, le compteur d'itérations, S la solution courante ($S*$ la meilleure solution), $S_{ij}(S)$ une solution possible, $z(S')$, le recouvrement de S' (z* le meilleur recouvrement), T, la liste des tabous, $DOM(t)$, le domaine des solutions possibles à l'itération t, et $\overline{DOM(t)}$ le domaine des solutions non-taboues ou pour lesquelles on peut lever le tabou à l'itération t. L'algorithme de la recherche avec tabous s'expose comme suit:

1. INITIALISATION

   Poser $t := 0$ et $T := \emptyset$; choisir une solution initiale S_0; poser $S := S_0$ et $z^* := z(S_0)$; lire $Maxiter$; poser $Derniter := Maxiter$.

2. LA RECHERCHE

Tant que $t < Derniter$ faire:

- $t := t + 1$;
- Choisir $S_{kl}(S)$ tel que $S_{kl}(S) \in \arg\max[z(S_{ij}(S))]$, où $S_{ij}(S) \in \overline{DOM(t)}$, $i = 1, ..., n$ et j admissible pour i.
- Si $z(S_{kl}(S)) > z^*$:
  - $S^* := S_{kl}(S)$;
  - $z^* := z(S_{kl}(S))$;
  - $Derniter := Maxiter + t$;
- Choisir aléatoirement $m \in [m_{min}, m_{max}]$;
- Mettre le mouvement inverse du déplacement de s_k vers $v_{kl}(S)$ dans T pour m itérations;
- $S := S_{kl}(S)$;
- Rétablir le voisinage de la station k et $DOM(t)$.

L'algorithme exposé ci-haut est appelé la "procédure de recherche avec tabous".

3.2.4 L'intensification et la diversification de la méthode

Après quelques tests préliminaires, nous avons observé que la méthode de recherche avec tabous était relativement bloquée par des optima locaux. Pour contrer cet inconvénient, nous introduisons deux nouveaux concepts: *l'intensification* et la *diversification* de la recherche.
Le but de l'intensification consiste à concentrer l'effort d'exploration dans certaines régions du domaine qui ont été identifiées comme prometteuses.
De son côté, le concept de diversification vise à rediriger la procédure vers des portions de l'espace des solutions où elle n'est pas encore allée (ou peu souvent) afin d'éviter que le processus de recherche ne soit trop localisé et laisse de grandes régions du domaine totalement inexplorées.
Nous allons augmenter la puissance de la méthode au moyen d'une procédure en deux étapes, qui s'apparente de près aux deux concepts d'intensification et de diversification:

1. Un retour à l'optimum trouvé jusqu'à maintenant: pour contrer l'inconvénient qu'a la méthode à se diriger vers des directions peu attrayantes, au bout de la procédure de recherche avec tabous, nous allons permettre à la recherche de retourner à la meilleure ligne trouvée jusqu'à maintenant. C'est l'intensification. La liste taboue est remise à zéro.

2. Un élargissement temporaire du domaine des solutions: pour permettre à la ligne de se diriger vers des régions peu ou pas du tout explorées, à partir de la meilleure solution, le voisinage de chaque station sera étendu à un périmètre plus grand. C'est la diversification.

L'idée consiste à "secouer" la ligne. On lui permet alors de prendre le voisin le plus attrayant de toutes les stations de la meilleure ligne trouvée jusqu'ici, en autant que celui-ci satisfasse les conditions de distances minimales et maximales entre les stations. S'il est choisi, alors il remplace la station qui lui est voisine. Ce voisin peut dégrader ou améliorer l'objectif, ou encore le laisser tel quel. Avec un élargissement du domaine des solutions, ce voisin peut être situé à une distance de marche plus grande que d'une unité de la station correspondante. La ligne forme alors un "pic" à cette station, et l'espace exploré se trouve alors agrandi. On peut, aussi, secouer c stations à la fois ($c = 1, ..., n$). Lors de la diversification, autour de chaque station, tous les points situés à une distance Manhattan de 8 unités de longueur ou moins par rapport à la station sont considérés. La figure 3.2 montre le cas d'une ligne de départ sur laquelle on a appliqué une fois la procédure de recherche avec tabous et la phase d'intensification et de diversification, lorsque, lors de cette phase, 3 secousses à la fois ont été permises sur 3 stations différentes. Les stations secouées sont indiquées par des flèches.

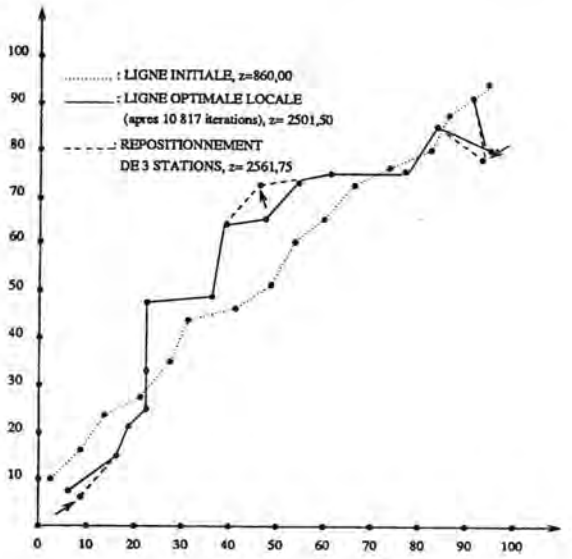


Figure 3.2: Illustration de trois secousses à la fois ($c = 3$)

Une fois la phase d'intensification et de diversification terminée, on part de la nouvelle ligne et on recommence la procédure de recherche avec tabous, avec cette fois-ci un domaine des solutions rétabli à au plus 4 voisins par station, et en ayant soin d'ajuster le paramètre *Derniter*, pour pouvoir continuer la procédure. Ici, pour ajouter à l'efficacité de la phase d'intensification et de diversification, nous allons empêcher à la station secouée de retourner vers son emplacement trouvé lors de la procédure de recherche avec tabous

précédente. Pour ce faire, la direction inverse est exprimée par un ou plusieurs mouvements tabous. Si, au cours de la procédure de recherche avec tabous, on arrive encore à améliorer l'objectif, on recommence la phase d'intensification et de diversification, suivie de la procédure de recherche avec tabous, et ainsi de suite. Sinon, le processus est arrêté. À la section 4, les tests numériques illustreront en détail l'application de ce schéma.

4. LES TESTS NUMÉRIQUES

4.1 La description des tests numériques de base

Nous allons recourir à une série de tests numériques pour déterminer la localisation des stations d'une ligne de métro. D'abord, dix treillis de 100 par 100 unités de longueur forment autant de tests numériques. Sur chacun de ces dix treillis, une population d'individus est générée aléatoirement par la superposition de carrés ayant divers scores numériques, le tout tel que décrit à la section 2. Le nombre de carrés superposés et leurs scores numériques restent les mêmes pour chaque test numérique. Puisqu'ils sont placés de façon aléatoire, c'est seulement leur emplacement sur chacun des treillis qui varie. La population totale sur chaque treillis est sensiblement la même. Sa densité diminue en fonction du nombre de carrés dont une portion dépasse les limites du treillis. Pour se donner un ordre de grandeur, la population totale sur chaque treillis est d'environ 40 000 individus. Chaque treillis est différent par la *répartition* de sa population. Voici la description complète du processus en trois étapes utilisé pour chaque test numérique:

1. La construction de 60 *lignes initiales* (30 dans la direction de chacune des deux diagonales du treillis) selon une marche aléatoire.
2. L'application de la méthode de recherche avec tabous pour 100 itérations au maximum, sur chacune des 60 lignes initiales.
3. L'application du traitement au complet, tel que proposé à la section 3, sur les 30 meilleures lignes (appelées *lignes de départ*) trouvées lors de l'étape 2. Rappelons que ce traitement comporte deux phases: la procédure de recherche avec tabous et la phase d'intensification et de diversification.

L'étape 2 de ce processus a pour but de donner à la recherche de bonnes lignes de départ. La section 4.2 exposera quelques résultats obtenus lors des tests numériques de base. La section 4.3 exposera quelques résultats obtenus sur l'effet de la durée de validité des mouvements tabous (paramètre m).

4.2 Les résultats obtenus lors des 10 tests de base

Nous allons ici exposer plusieurs données obtenues lors des tests numériques. Un tableau illustrera quelques résultats, pour chacun des 10 treillis considérés. Tous les tests ont

été effectués sur une station de travail SUN Sparc10. Lorsque des temps de calcul sont présentés, ils sont tous exprimés en secondes. Les termes suivants seront utilisés dans le tableau:

1. *La solution initiale* : la valeur du recouvrement obtenu pour la ligne initiale de départ (la fonction objectif z), qui longe une diagonale.
2. *L'optimum local glouton* : la plus grande valeur de l'objectif z, si l'on appliquait à la ligne initiale de départ uniquement l'heuristique interne, sans appliquer de méthode de recherche avec tabous, c'est à dire en utilisant un algorithme glouton.
3. *Le temps total glouton* : le nombre de secondes qu'il a fallu pour terminer l'algorithme glouton.
4. *L'optimum local tabou* : la plus grande valeur de l'objectif z, alors que l'on appliqué à la ligne initiale de départ le processus en trois étapes, tel que décrit à section 4.1.
5. *Le temps à l'optimum tabou* : le nombre de secondes qu'il a fallu pour se rendre à l'optimum local tabou.
6. *Le temps au total* : le nombre de secondes qu'il a fallu pour terminer le processus.

Au tableau 4.1, pour chaque treillis, nous avons recueilli la valeur du meilleur optimum local tabou trouvé. En d'autres termes, pour chaque treillis, 60 lignes initiales ont été construites, soit 30 le long de chacune des deux diagonales. La méthode de recherche avec tabous a été appliquée sur chacune de ces 60 lignes, pendant 100 itérations. Les données des meilleurs optima locaux tabous ont été recueillies. Sur ces 60 optima locaux tabous, les 30 meilleurs en termes de recouvrement ont été sélectionnés. Sur les 30 lignes de ces optima locaux , le traitement de l'étape 3 du processus a été appliqué. Le meilleur optimum local tabou, parmi les 30 lignes obtenues suite au processus, a été retenu pour chaque treillis.

Tableau 4.1: Meilleurs optima locaux tabous

Treillis no	Solution initiale	Optimum glouton		Optimum tabou		
		Valeur	Temps total	Valeur	Temps à l'opt.	Temps au total
1	1107,75	1984,00	0,26	2921,50	106,75	173,60
2	1575,75	2288,50	0,15	2891,75	106,65	175,02
3	970,50	1643,25	0,17	2601,25	220,20	288,73
4	505,75	1832,75	0,32	2833,00	297,55	363,77
5	572,50	1342,75	0,17	2901,50	205,65	270,68
6	999,75	1485,25	0,15	2398,50	132,17	197,38
7	798,00	1722,00	0,18	2744,75	200,75	265,93
8	1247,00	2152,50	0,20	2581,00	170,42	236,12
9	782,50	2134,25	0,28	2409,25	69,55	135,80
10	1231,75	2260,75	0,23	2897,75	83,00	147,72

Dans tous les cas, l'optimum local tabou est meilleur que l'optimum local glouton. Les optima locaux tabous sont nettement supérieurs aux recouvrements observés des lignes initiales.

En calculant la moyenne et le coefficient de variation obtenu à la suite de la procédure, pour chaque groupe de trente lignes, nous avons obtenu des moyennes variant de 2156,43 à 2721,68 individus pour le recouvrement, et des coefficients de variation variant de 3,6% à 12,1%. Ces valeurs tendent à démontrer que la méthode de recherche avec tabous donne des ensembles de lignes assez homogènes: on peut alors conclure que la méthode est passablement robuste puisque plusieurs solutions initiales permettent de trouver des optima locaux de valeurs semblables.

4.3 L'effet du paramètre m

La durée d'interdiction d'un mouvement (paramètre m) est un paramètre difficile à calibrer. Si m est trop grand, le domaine des solutions possibles est réduit et une série de mouvements attrayants peuvent être ignorés. Si m est trop petit, on peut assister à un cyclage, ou du moins voir la recherche stagner dans les mêmes zones du treillis, sans augmenter son champ d'exploration. Si le territoire étudié contient beaucoup de "montagnes", l'exploration pourrait ne jamais en sortir, bloquée par ces maxima locaux. Nous avons expérimenté l'effet de plusieurs paramètres et voici quelques observations:

1. Bien qu'il y ait une amélioration notable de l'objectif, la plupart des lignes finales obtenues lors des tests de base restent dans le même corridor que les lignes initiales;
2. Lorsqu'un grand nombre de stations sont secouées en même temps (paramètre c) au cours de la phase d'intensification et de diversification, on peut espérer obtenir des lignes dont le recouvrement est largement supérieur à la moyenne;
3. Le fait d'augmenter la distance maximale entre deux stations voisines l'une de l'autre (paramètre l_{max}) amène de bonnes solutions;
4. Le fait d'augmenter de façon significative le paramètre *Maxiter* amène de bonnes solutions, mais augmente tout autant les temps de calcul, ce qui n'est pas rentable en pratique.

Ces éléments laissent penser que la recherche est relativement bloquée par les optima locaux, et que la stratégie à adopter est de forcer la recherche à s'en sortir. En augmentant la longueur des listes taboues de façon significative, on forcerait l'exploration à sortir des montagnes, même si elles sont attrayantes, et d'aller en visiter d'autres, quitte à y revenir plus tard, quand:

1. Le mouvement tabou sera permis à nouveau;
2. La phase de secousses aura passé. (À ce moment-là, seuls les mouvements inverses reliés à la station secouée seront tabous.)

Nous proposons d'observer les effets du paramètre m lorsqu'il est compris dans les intervalles suivants:

1. $m \in [20, 40]$;
2. $m \in [25, 75]$;
3. $m \in [50, 100]$.

Nous gardons les mêmes lignes initiales que celles des tests de base, et la même méthode est employée, mais cette fois-ci avec des valeurs de m différentes. Nous avons recueilli, de la même façon qu'aux tests numériques de base, les données concernant le meilleur optimum local tabou, pour chaque treillis. Le tableau 4.2 expose les résultats obtenus pour chacun des 3 intervalles, en ce qui concerne les meilleurs optima locaux tabous pour chaque treillis, ainsi que les temps au total d'exécution, pour chaque optimum.

Tableau 4.2: Meilleurs optima locaux tabous: différentes valeurs de m

Treillis	$m \in [20, 40]$		$m \in [25, 75]$		$m \in [50, 100]$	
no	Optimum local	Temps au total	Optimum local	Temps au total	Optimum local	Temps au total
1	3672,00	188,79	3623,25	21,34	3589,75	14,33
2	3627,25	154,79	3609,00	58,39	3415,75	9,75
3	2644,50	154,08	2902,75	63,77	2653,00	22,22
4	3308,50	59,89	3298,25	50,78	3205,50	15,56
5	3878,25	130,04	3871,25	74,36	3764,50	15,02
6	3034,75	269,53	3260,25	93,65	2735,00	11,59
7	2848,00	246,30	2863,75	57,17	2785,75	33,56
8	2954,75	135,60	2997,50	79,32	2926,50	16,56
9	3140,00	155,99	3040,00	70,24	3018,00	23,35
10	2903,25	63,70	2910,75	75,90	2912,25	17,42

On observe une très grande amélioration; non seulement au niveau de la valeur des optima locaux tabous, mais aussi au niveau des temps de calcul, qui sont plus courts. Les meilleurs recouvrements se trouvent, en général, dans l'intervalle $[20, 40]$, et tendent à diminuer lorsque m augmente, mais pas de façon significative. Il est étonnant de voir la performance de la méthode, lorsque m augmente. Ainsi, on remarque que lorsque $m \in [50, 100]$, les temps de calcul sont très petits, alors que les recouvrements observés sont élevés. On pourrait penser que le fait d'augmenter la longueur de la liste taboue amène la recherche à se diriger rapidement vers un bon optimum local, tellement bon qu'aucun autre meilleur recouvrement ne peut être trouvé, à l'intérieur de *Maxiter* itérations. Ceci expliquerait la rapidité des temps de calcul. Il faut cependant prendre en compte le fait qu'étant donné que la longueur de la liste taboue est augmentée, la cardinalité de l'ensemble des solutions possibles $\overline{DOM(t)}$ diminue. Lorsqu'il n'y a plus de voisins disponibles, la procédure de recherche avec tabous est stoppée, et la phase d'intensification et de diversification est effectuée, aussi longtemps qu'il y a amélioration de l'objectif dans la procédure de recherche avec tabous. Les temps de calcul plus courts seraient donc dus au fait qu'il n'y a plus

de voisins disponibles. Les recouvrements très élevés pourraient être dus au fait qu'étant donné que la recherche se rend souvent à la phase d'intensification et de diversification à cause du manque de voisins disponibles, les lignes se trouvent fréquemment secouées, et ceci aurait un effet positif du point de vue de la diversification. Le tableau 4.3 expose les statistiques concernant les optima locaux tabous, ainsi que leurs temps totaux d'exécution, pour chacun des trois cas.

Tableau 4.3: Statistiques sur les optima locaux tabous: différentes valeurs de m

Treillis no		$m \in [20, 40]$		$m \in [25, 75]$		$m \in [50, 100]$	
		Optimum local	Temps au total	Optimum local	Temps au total	Optimum local	Temps au total
1	$\bar{x}$	3217,81	94,37	3211,32	24,31	3054,18	10,12
	cv	0,047	0,541	0,058	0,468	0,053	0,471
2	$\bar{x}$	3367,50	109,87	3287,51	34,86	2998,12	10,97
	cv	0,047	0,431	0,048	0,398	0,062	0,455
3	$\bar{x}$	2478,06	153,41	2547,23	64,64	2391,38	17,82
	cv	0,043	0,462	0,047	0,533	0,044	0,397
4	$\bar{x}$	2944,54	82,72	2976,66	37,80	2818,97	11,11
	cv	0,114	0,581	0,096	0,486	0,104	0,452
5	$\bar{x}$	3086,48	111,74	3238,40	45,96	2962,575	12,91
	cv	0,155	0,358	0,153	0,834	0,114	0,675
6	$\bar{x}$	2642,06	109,47	2623,40	38,49	2387,35	8,76
	cv	0,061	0,491	0,099	0,924	0,063	0,575
7	$\bar{x}$	2620,63	128,35	2647,32	56,97	2534,20	14,46
	cv	0,104	0,373	0,077	0,473	0,084	0,487
8	$\bar{x}$	2793,93	133,17	2774,53	48,89	2662,66	11,39
	cv	0,043	0,396	0,050	0,532	0,055	0,478
9	$\bar{x}$	2862,46	129,71	2842,08	37,79	2708,58	13,27
	cv	0,063	0,434	0,043	0,488	0,055	0,505
10	$\bar{x}$	2822,49	105,30	2813,10	41,06	2764,12	16,27
	cv	0,021	0,406	0,022	0,510	0,025	0,464

Le graphique de la ligne correspondant au meilleur optimum local tabou parmi les trois intervalles, pour un treillis sélectionné, est montré à l'annexe I. On remarque que les moyennes des recouvrements sont plus élevées dans le cas où $m \in [20, 40]$, et diminuent par la suite. Par contre, même dans le cas où $m \in [50, 100]$, les moyennes restent nettement plus élevées que celles des tests numériques de base. Avec des temps de calcul inférieurs à ceux des tests de base, et des coefficients de variation assez bas, le fait d'augmenter la durée de validité m des mouvements tabous est une bonne stratégie.

CONCLUSION

Le problème de déterminer la localisation des stations d'une ligne de métro pourrait être résolu à l'aide de la méthode de recherche avec tabous que nous avons proposée. En effet,

lorsque la longueur des listes taboues est bien calibrée, nous obtenons de bonnes lignes de métro – du point de vue du recouvrement de population –, tout en ayant des temps de calcul raisonnables, de l'ordre de deux à trois minutes.

Si des raisons extérieures (économiques, environnementales, politiques, etc.) forcent le planificateur à implanter la ligne de métro dans un certain corridor, il pourrait utiliser le processus décrit dans cet article, mais partir de solutions initiales qui se trouvent dans ce corridor, au lieu des solutions initiales qui longent la diagonale.

La méthode proposée pourrait être utilisée dans le cas ou l'on voudrait élargir un réseau déjà existant; à ce moment-là, certaines dispositions doivent être prises pour empêcher le plus possible de recoupements entre les stations de deux lignes différentes.

Bien sûr, comme nous l'avons mentionné, la ligne de métro finale sera l'aboutissement d'un processus de décision long et complexe. La recherche opérationnelle est un outil, et peut aider à rationnaliser certaines décisions qui pouvaient se prendre de façon *ad hoc* auparavant. Toutefois, son rôle est d'aider le planificateur à prendre des décisions rationnelles, et non pas à s'y substituer.

BIBLIOGRAPHIE

[1] Allport, R.J. et M. Brown (1993), "Economic Benefits of the European High- Speed Rail Network", Transportation Research Record, 1381, pp. 1–11.

[2] Gendreau, M., G. Laporte et J.A. Mesa (1994), "Locating Rapid Transit Lines", Journal of Advanced Transportation, à paraître. Aussi rapport CRT–907, Centre de recherche sur les transports, Université de Montréal.

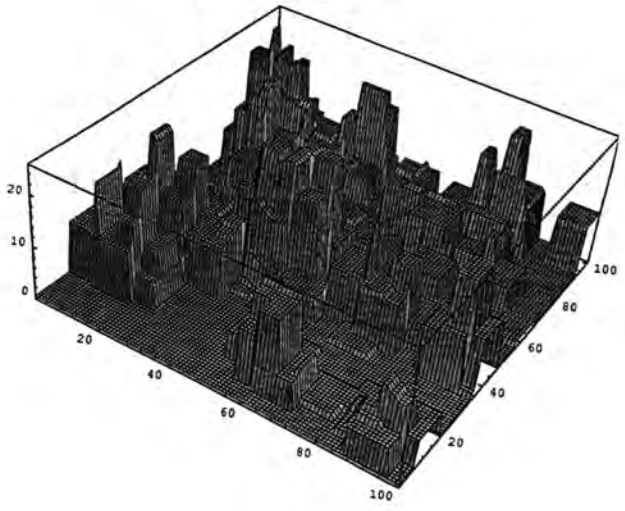
[3] Glover, F. (1989), "Tabu Search – Part I", ORSA Journal on Computing, 1, pp 190–206.

[4] Glover, F. (1990), "Tabu Search – Part II", ORSA Journal on Computing, 2, pp 4–32.

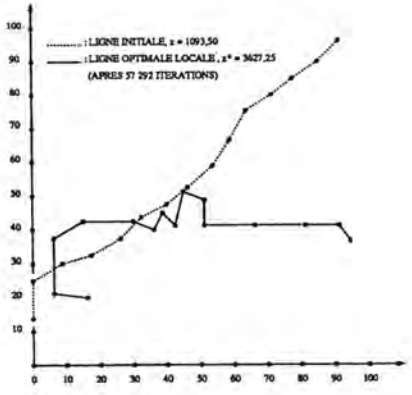
[5] Lutin, J.M. et G.P. Benz (1992), "Key Issues in Light Rail Transit Station Planning and design", Transportaion Research record, 1361, pp 117–124.

[6] Magnanti, T.L. et R.T. Wong (1984), "Network Design and Transportation Planning: Models and Algorithms", Transportation Science, 18, pp 1–55.

ANNEXE I



Territoire 2



Meilleur optimum local, treillis 2, $m \in [20, 40]$

The Journey to Work and Urban Spatial Structure: A Case Study of the Greater Toronto Area, 1986 to 1991

Darren M. Scott
Department of Geography, McMaster University

Abstract

This paper examines changes in journey-to-work flows to uncover the evolving spatial structure of the Greater Toronto Area, from 1986 to 1991. Using data from the 1986 and 1991 Transportation Tomorrow Surveys, two analyses are employed. First, principal components analysis is applied to 1986 journey-to-work flow data for 40 Greater Toronto Area (GTA) zones to delineate functional regions for which changes in flows are examined, as viewed from two perspectives: the origins of workers employed in each region, and the destinations of workers residing in each region. Second, changes in the spatial configuration of GTA zones are examined using non-metric multidimensional scaling to visualize structure in the destinations of workers residing in each zone. Findings from these analyses suggest that the spatial structure of the Greater Toronto Area is becoming more complex over time because of the changing orientation of journey-to-work flows, emphasizing the urban periphery as a place of work and residence.

Introduction

Most people will agree that although cities, since their genesis millennia ago, have undergone an extraordinary evolution in both form (the spatial pattern of land uses) and structure (the spatial organization of the urban landscape), they have remained immutable with respect to two basic functions—residence and workplace. Hanson and Pratt (1988, 299) state unequivocally that "the link between home and work is one of the cornerstones of urban geography." However, Vickerman (1984, 15) captures the significance of this linkage, stating that "a key factor in the spatial structure of metropolitan areas is the relationship between workplace and residential location." Although such basic urban functions have remained unchanged through

time, their locational relationships have not, resulting in a continual reorganization of metropolitan spatial structure. Thus, this paper examines changes in journey-to-work flows over time to uncover the evolving spatial structure of the North American metropolis. Specifically, such changes are investigated for the Greater Toronto Area, from 1986 to 1991. The reasons for this choice of spatio-temporal context are:

1. the Greater Toronto Area is representative of large, contemporary, North American urban regions. It defines a commuter shed for a central city (Toronto), and it experiences the morphogenetic processes that continuously reshape the North American urban landscape.
2. journey-to-work data are readily available for the years 1986 and 1991.

Study Area

The Greater Toronto Area is situated along the western shores of Lake Ontario (Figure 1) where it "encompasses an area slightly over 7,000 square kilometres, of which approximately 1,500 square kilometres are urban and over 5,500 square kilometres are rural" (Municipality of Metropolitan Toronto 1992, 1). It comprises five regional municipalities (Metropolitan Toronto, Durham, York, Peel, and Halton) that, in turn, consist of 30 area municipalities. Because of the high population density within its boundaries, Metropolitan Toronto is divided into 16 major planning districts. The dominant urban form in other regional municipalities is low-density, automobile-oriented, residential and commercial development. Many individuals must endure long journeys to work because of the distances between residences outside of Metropolitan Toronto and workplaces inside it.

Data

For this investigation, journey-to-work flow data were derived from the 1986 and 1991 *Transportation Tomorrow Surveys* that were conducted by the Data Management Group at the University of Toronto for the Regional Municipalities of Metropolitan Toronto, Peel, Halton, York, and Durham; the Ontario Ministry of Transportation; GO Transit; and the Toronto Transit Commission. The utility of these surveys for geographical research is found in their design—that is, for each household surveyed, information was obtained concerning the attributes of that household, the attributes of all residents older than five years, and the attributes of each trip made on the previous day by household members.

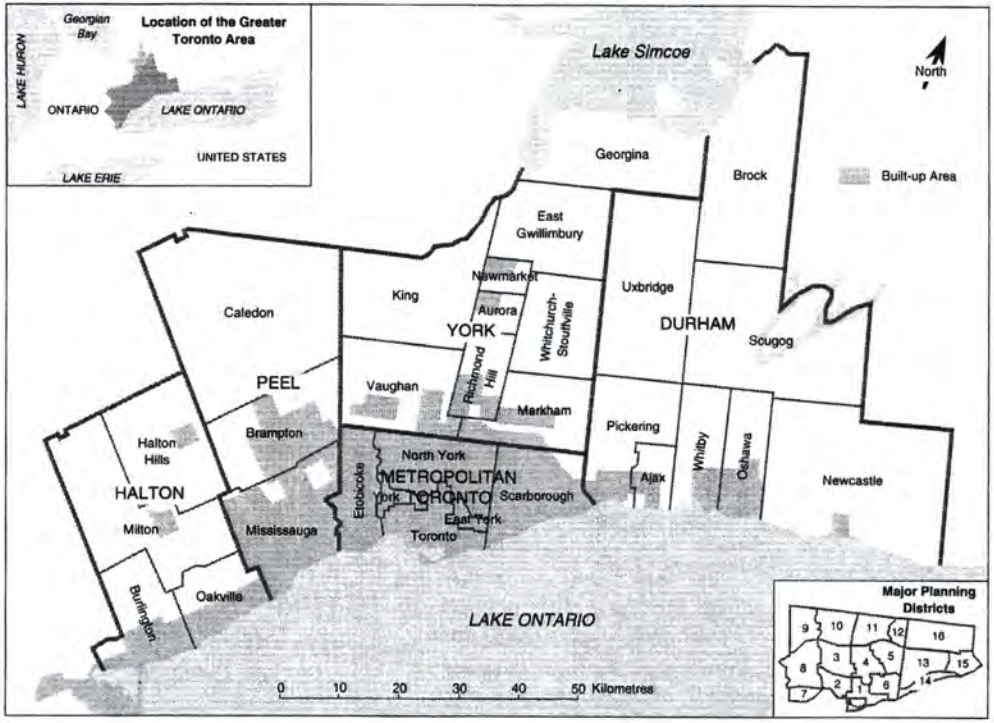


Figure 1
The Greater Toronto Area.

An important difference between the 1986 and 1991 surveys involved sampling design. In 1986, the survey consisted of a 4.1 percent random sample of all households in the Greater Toronto Area. In 1991, the sample was stratified based on low-growth and high-growth areas that were defined in terms of the estimated population change between 1986 and 1991 at the traffic zone level. Areas that experienced greater than 17 percent growth were categorized as high-growth. These areas were sampled at a rate of 4.5 percent, whereas low-growth areas were sampled at a rate of 0.5 percent. Therefore, comparisons between the two years had to be made at an aggregate level, such as Greater Toronto Area (GTA) zones—the 16 major planning districts of Metropolitan Toronto and the 24 area municipalities of Durham, York, Peel, and Halton Regions, as shown in Figure 1 (Ontario Ministry of Transportation 1992).

For the surveys, traffic zones were grouped into larger expansion zones for the purpose of expanding the sample to represent the total population of the Greater Toronto Area. Expansion factors were defined for each of these zones as the ratio of census population to sample population. Thus, the data pertaining to each individual surveyed were used to represent many people. In 1986, the difference between the census and expanded survey population was 2.2 percent, whereas, in 1991, the difference was 2.46 percent (Ontario Ministry of Transportation 1992).

Journey-to-work matrices were created for this investigation using the origin, destination, and expansion factor associated with each individual who had a work trip within the Greater Toronto Area on the day prior to being surveyed. The 1986 survey contained 66,070 such individuals, whereas the 1991 survey contained 24,771. The difference in the physical size of these samples was attributed to the sampling designs employed for the surveys. The samples were, however, assumed to be comparable because the expansion factors for low-growth areas were considerably larger for 1991 than they were for 1986.

Methods

Functional Regions

For this investigation, a functional region is defined as a set of GTA zones that exhibit similar journey-to-work destination patterns. Such regions were delineated for the Greater Toronto Area as a means of providing a general assessment of changes in journey-to-work flows, as viewed from two perspectives: the origins of workers employed in each region, and the destinations of workers residing in each region.

Principal components analysis was chosen to identify functional regions in the Greater Toronto Area because its application to flow data is a proven regionalisation

technique—that is, it has been used in numerous studies to delineate regions that comprise groups of origins with similar destination patterns (Davies and Musson 1978; Nader 1980; 1981). This technique was applied to 1986 journey-to-work flow data for GTA zones. Initially, a data matrix was created with each column representing an origin from which journey-to-work flows were derived for each of 40 destinations. This matrix was then transformed into a matrix of correlation coefficients that measured the similarity between destination patterns of origin pairs. Factoring of the correlation matrix yielded nine factors, with eigenvalues of 1.0 or greater, that were subjected to an orthogonal (varimax) rotation. Together, the nine factors explained 77.1 percent of the variance in the destination patterns of the origins. The highest factor loading for each origin was used in delineating the nine functional regions shown in Figure 2.

The functional regions identified by the principal components analysis were used in creating two matrices of journey-to-work flows—one for 1986 and the other for 1991. These matrices were viewed from the perspectives outlined above. For each perspective, a relative measure of the relationship between workplace and residence was derived using information obtained from the matrices. For each destination region, the origins of its workforce were measured as the proportion of workers residing in each origin region. For each origin region, the destinations of its workforce were measured as the proportion of workers employed in each destination region. Subsequently, these measures were compared, for 1986 and 1991, to identify changes in journey-to-work flows for functional regions in the Greater Toronto Area.

GTA Zones

Non-metric multidimensional scaling (MDS) provided a convenient means for visualizing structure among GTA zones, using 1986 and 1991 journey-to-work flows as measures of proximity. Goodchild and Janelle (1988, 13) state that "spaces of two dimensions have been used in many [MDS] applications to obtain a visual representation of matrices of interactions between objects, on the understanding that strong interaction can be equated with similarity, and therefore, with proximity in two dimensions." Thus, two-dimensional solutions were specified for each year to simplify interpretation, including changes over the five-year period. However, it was necessary to provide an initial configuration for the GTA zones, based on coordinates obtained for their centroids in absolute space, so that these spatial representations of structure could be compared. The journey-to-work flows were not preprocessed to remove the effects of zone size from them because non-metric MDS seeks solutions in which the distances between objects maintain the same rank order as their proximities in the input matrix. However, it was specified that the matrices were row conditional, meaning that the structure depicted in the spatial

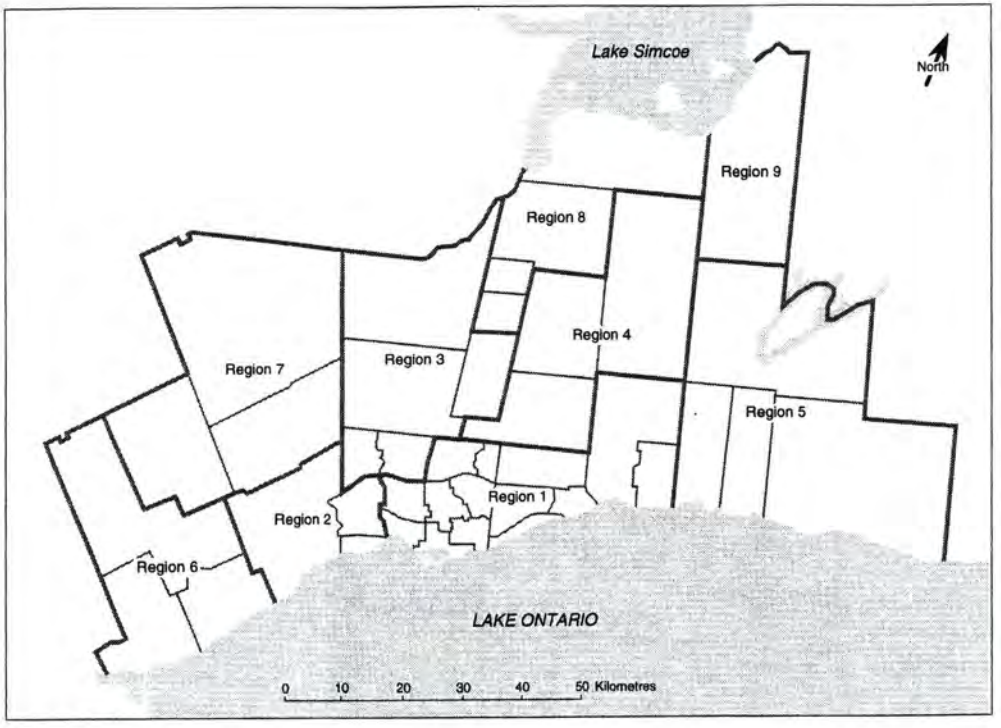


Figure 2
Functional regions in the Greater Toronto Area.

configurations represented the destinations of workers residing in GTA zones. This method was employed to reduce the amount of stress in the solutions—that is, the deviation between the proximities input to the analysis, and the relative distances displayed in the spatial configurations. Two measures of stress were used in this analysis: the coefficient of determination, which measures *goodness-of-fit* since the value increases as the match between proximities and distances increases; and Kruskal's stress, which measures *badness-of-fit* since the value increases as the match between proximities and distances decreases.

Findings

Employment and Residential Location of the GTA Workforce

The structural evolution of a metropolitan area depends on the physical growth of constituent regions in terms of their employment and residential opportunities. Changes in the relative concentration of workers employed or residing in such regions are indicative of alterations in the journey-to-work flows that define urban spatial structure. Therefore, an appropriate starting point in any investigation of the structural evolution of a metropolitan area is to examine the changing geography of workplace and residence.

Based on the functional regions shown in Figure 2, regional shifts in employment location, between 1986 and 1991, are presented in Table 1 for the GTA workforce. Although Region 1 (Central Region) dominated the Greater Toronto Area in terms of employment opportunities, the table clearly demonstrates the decline in the relative importance of the region as a focus for economic activity—in 1986, 55 percent of the workforce found employment in the region, whereas, by 1991, this had dropped to 52.8 percent. This decline was accompanied by increased economic activity in peripheral regions, particularly Regions 2 and 4. The other regions of the Greater Toronto Area remained relatively stable, experiencing little or no change in their proportions of employment opportunities.

Table 2 reveals regional shifts in the residential location of the GTA workforce. The table clearly demonstrates the dominance of the Central Region as a place of residence, despite its decline in relative importance—in 1986, 50.5 percent of the workforce resided in the region, whereas, by 1991, this had dropped to 47.3 percent. The proportion of workers residing in Region 2 also declined during the five-year period. Gains in residential activity were distributed relatively evenly among Regions 3, 4, 5, 6, 7, and 8—they increased their residential opportunities by 0.5 to 0.7 percent.

The findings presented in Tables 1 and 2 indicate that changes have occurred to the journey-to-work flows that define the evolving spatial structure of the Greater

Toronto Area. The relative decline of the Central Region and the relative growth of peripheral regions suggest that such flows were being reoriented, emphasizing the urban periphery as a place of work and residence. The following subsection examines such change.

Table 1
Employment Location of the Greater Toronto Area Workforce, 1986 and 1991

	Region									Total
Year	1	2	3	4	5	6	7	8	9	Workforce
1986	55.0	15.4	12.0	3.5	4.3	3.9	4.5	1.3	0.1	1,537,287
1991	52.8	16.2	11.9	4.5	4.2	4.1	4.7	1.5	0.1	1,691,884

Note: Employment location of the GTA workforce is measured as the proportion of workers employed in each region.

Table 2
Residential Location of the Greater Toronto Area Workforce, 1986 and 1991

	Region									Total
Year	1	2	3	4	5	6	7	8	9	Workforce
1986	50.5	16.8	9.4	3.5	5.2	5.3	6.7	2.4	0.2	1,537,287
1991	47.3	16.5	10.0	4.2	5.7	5.8	7.3	3.0	0.2	1,691,884

Note: Residential location of the GTA workforce is measured as the proportion of workers residing in each region.

Origins and Destinations of the GTA Workforce

Table 3 expresses the journey-to-work flows to each destination region as percentages of the total workforce employed in the region. The diagonal of the table represents the proportion of each destination's workforce that resides in the region—that is, the degree to which it is self-contained. This concept is central to any study concerning the structural evolution of an urban area because, in terms of spatial interaction, the relative importance of a destination region is directly related to the proportion of the metropolis that provides it with labour (O'Connor and Maher 1979).

In terms of self-containment, Table 3 reveals some interesting trends. Region 4 was the least self-contained—27.5 percent of its workforce resided locally in both

Table 3
Origins of the Workforce Employed in each Destination Region of the Greater Toronto Area, 1986 and 1991

Origin		Destination Region								
Region	Year	1	2	3	4	5	6	7	8	9
1	1986	74.3	19.4	32.9	47.1	10.1	4.2	7.8	10.5	0.0
	1991	71.5	17.6	30.2	43.1	10.2	4.4	8.1	7.4	0.0
2	1986	9.7	54.6	15.6	2.9	0.9	10.1	13.7	0.6	0.0
	1991	9.3	54.2	12.6	4.2	0.9	10.1	14.7	0.9	0.0
3	1986	6.1	6.6	35.7	10.1	0.6	0.9	5.8	8.3	0.0
	1991	6.7	6.8	37.6	9.5	0.4	1.5	5.2	10.2	0.0
4	1986	3.5	0.9	3.1	27.5	1.1	0.2	0.5	3.8	5.1
	1991	4.2	0.8	4.1	27.5	0.9	0.2	0.3	4.8	3.0
5	1986	2.3	0.3	0.4	2.4	86.1	0.1	0.4	0.8	5.3
	1991	3.2	0.5	0.6	3.9	86.9	0.3	0.2	1.3	0.0
6	1986	1.4	6.7	1.3	0.7	0.1	80.6	3.2	0.0	0.0
	1991	1.7	7.2	1.9	1.0	0.0	80.6	3.4	0.0	0.0
7	1986	1.5	10.9	7.2	2.1	0.2	3.7	68.0	0.6	0.0
	1991	1.9	12.4	8.0	1.0	0.1	2.9	67.2	1.1	0.0
8	1986	1.1	0.5	3.7	6.6	0.3	0.2	0.5	73.8	9.7
	1991	1.4	0.5	4.8	9.4	0.1	0.0	0.8	73.4	33.3
9	1986	0.0	0.0	0.0	0.6	0.7	0.0	0.0	1.7	80.0
	1991	0.1	0.0	0.1	0.5	0.5	0.0	0.0	1.0	63.7

Note: For each destination region, the origins of its workforce are measured as the proportion of workers residing in each origin region.

1986 and 1991, whereas over 40 percent, journeyed from the Central Region. This trend is not unexpected given the proximity of the two regions, and the fact that Region 4 had high ratios of employment to residential opportunities for both years (Tables 1 and 2). Region 5 was the most self-contained—86.9 percent of its workforce resided locally in 1991, an increase of 0.8 percent from 1986. Regions 1, 2, 6, 7, 8, and 9 also exhibited self-containment in excess of 50 percent for both years, in contrast to Region 3. Region 9 experienced the greatest change in self-containment over the five-year period—80 percent of its workforce resided locally in 1986, as compared to 63.7 percent in 1991. Although self-containment decreased in four other regions (Regions 1, 2, 7, and 8) the magnitude of change ranged from 0.4 to 2.8 percent.

Insights concerning the structural evolution of the Greater Toronto Area can also be gained by examining the columns of Table 3, which reveal changes in the proportion of a particular destination's workforce originating from each of the other regions. Apart from its resident workforce, the most important origin of Central Region workers was Region 2. While its contribution decreased by 0.4 percent over the five-year period, it was still providing 9.3 percent of the Central Region's workforce in 1991. In contrast, all other peripheral regions increased their respective proportions of Region 1's workforce. In turn, the Central Region was important, as a source of workers, to Regions 2, 3, 4, 5, and 8 despite the fact that its relative contributions decreased to all of these regions, except for Region 5. A comparison of Table 3 with Figure 2 reveals that contiguous regions were mostly responsible for the greatest increases in the proportion of workers journeying to destination regions.

An additional perspective concerning the structural evolution of the Greater Toronto Area is provided by Table 4, which expresses the journey-to-work flows from each origin region as percentages of the total workforce residing in the region. The diagonal of the table represents the proportion of each origin's workforce that is employed in the region. With the exception of Regions 2 and 4, this proportion decreased, over the five-year period, for all other regions. The decrease was particularly profound for Region 9—the proportion of resident workers employed in this region decreased by 16.4 percent to 30.4 percent in 1991. In both years, the Central Region boasted the greatest proportion of workers employed locally, whereas Region 4 had the lowest.

The rows of Table 4 reveal changes in the proportion of a particular origin's workforce employed in each of the other regions. Clearly, the Central Region was an important destination for workers. However, over the five-year period, its relative importance declined with respect to Regions 2, 3, and 4, which, together with the Central Region, comprised 80.2 percent of the resident GTA workforce in 1986, and 78 percent, in 1991 (Table 2). In contrast, the relative importance of Region 4, as a destination, increased for all regions, with the exception of Region

Table 4
Destinations of the Workforce Residing in each Origin Region of the Greater Toronto Area, 1986 and 1991

Origin		Destination Region								
Region	Year	1	2	3	4	5	6	7	8	9
1	1986	80.8	5.9	7.8	3.3	0.9	0.3	0.7	0.3	0.0
	1991	80.0	6.0	7.6	4.1	0.9	0.4	0.8	0.2	0.0
2	1986	31.8	50.1	11.1	0.6	0.2	2.3	3.7	0.0	0.0
	1991	29.8	52.9	9.1	1.1	0.2	2.5	4.2	0.1	0.0
3	1986	35.5	10.9	45.3	3.8	0.3	0.4	2.8	1.2	0.0
	1991	35.2	11.0	44.8	4.2	0.2	0.6	2.5	1.5	0.0
4	1986	54.4	4.0	10.4	27.4	1.3	0.3	0.7	1.4	0.2
	1991	52.5	3.2	11.7	29.4	0.9	0.2	0.4	1.7	0.1
5	1986	24.6	1.0	0.9	1.6	71.1	0.1	0.3	0.2	0.1
	1991	29.7	1.4	1.3	3.1	63.7	0.2	0.2	0.3	0.0
6	1986	15.0	19.5	3.0	0.5	0.1	59.2	2.7	0.0	0.0
	1991	15.3	20.3	4.0	0.7	0.0	56.9	2.8	0.0	0.0
7	1986	12.5	25.2	13.0	1.1	0.1	2.1	45.9	0.1	0.0
	1991	13.5	27.4	13.0	0.6	0.0	1.6	43.6	0.2	0.0
8	1986	25.1	3.0	18.6	9.8	0.5	0.3	1.0	41.2	0.5
	1991	25.3	2.6	19.1	14.0	0.1	0.1	1.3	36.7	0.9
9	1986	11.0	1.8	2.8	11.0	14.7	0.0	0.0	11.9	46.8
	1991	30.4	0.0	4.4	13.0	13.0	0.0	0.0	8.7	30.4

Note: For each origin region, the destinations of its workforce are measured as the proportion of workers employed in each destination region.

7. These findings are significant because they further emphasize regional shifts in employment location—that is, the relative decline of the Central Region and the relative growth of peripheral regions.

Tables 3 and 4 have highlighted, at an aggregate level, the changing journey-to-work flows that define the evolving spatial structure of the Greater Toronto Area. The findings suggest that urban spatial structure was becoming more complicated over time despite the high degree of self-containment for most regions. The following subsection examines changes in journey-to-work flows at a smaller scale—that of the GTA zone, as shown in Figure 1.

Structural Configuration of GTA Zones

For 1986 and 1991, Figures 3 and 4 present two-dimensional, non-metric scaling solutions for journey-to-work flows among 40 GTA zones. The coefficient of determination and stress index for the former configuration are, respectively, 0.72 and 0.23, whereas, for the latter, they are 0.64 and 0.26. These values indicate that both solutions replicated their original proximity matrices reasonably well. Comparison of the spatial configurations yielded insights concerning the structural evolution of the Greater Toronto Area.

Figure 3 reveals several interesting features concerning the spatial structure of the Greater Toronto Area in 1986. It clearly identifies the schism between zones that are primarily places of work, and those that are primarily places of residence. As expected, important employment nodes (Planning District 1, Vaughan, and Markham) are clustered near the centre of the configuration. The reason for this is that they are centrally located within the region, and therefore, receive significant journey-to-work flows from most other GTA zones. This is evidenced by their low self-containment, which is measured as the proportion of each GTA zone's employed workforce that resides locally (Table 5). In contrast, residential zones are located in the periphery of the configuration because they receive flows from few other zones. In fact, there appears to be an interaction gradient radiating outward from the centre, whereby zones are grouped according to their importance as workplaces—that is, as self-containment increases, and zones become less important, their distance from the centre also increases. For example, Richmond Hill, with 34.3 percent self-containment (Table 5), serves as an important workplace for surrounding zones, as well as an important place of residence. As expected, it is located between the centre of the configuration and the zones that comprise its labour-shed. In contrast, Brock, with the second highest self-containment in the Greater Toronto Area (Table 5), is located farthest from the centre.

With the exception of the primary employment cluster, it is possible to discern five groups of zones that are spatially contiguous. Three groups—Northern (Uxbridge, Brock, King, Georgina, East Gwillimbury, Whitchurch-Stouffville,

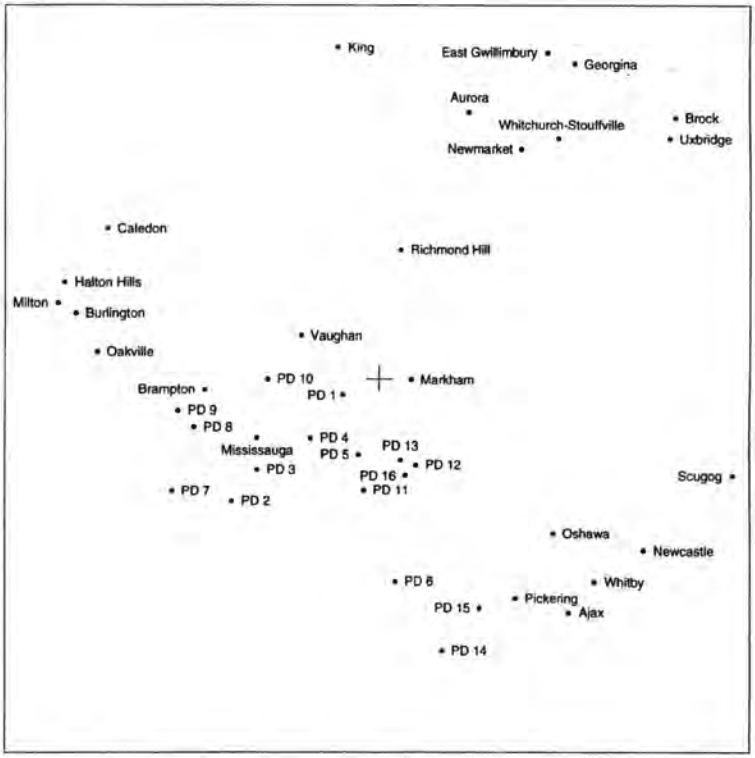


Figure 3
Destination structure of journey-to-work flows among Greater Toronto Area zones, 1986.

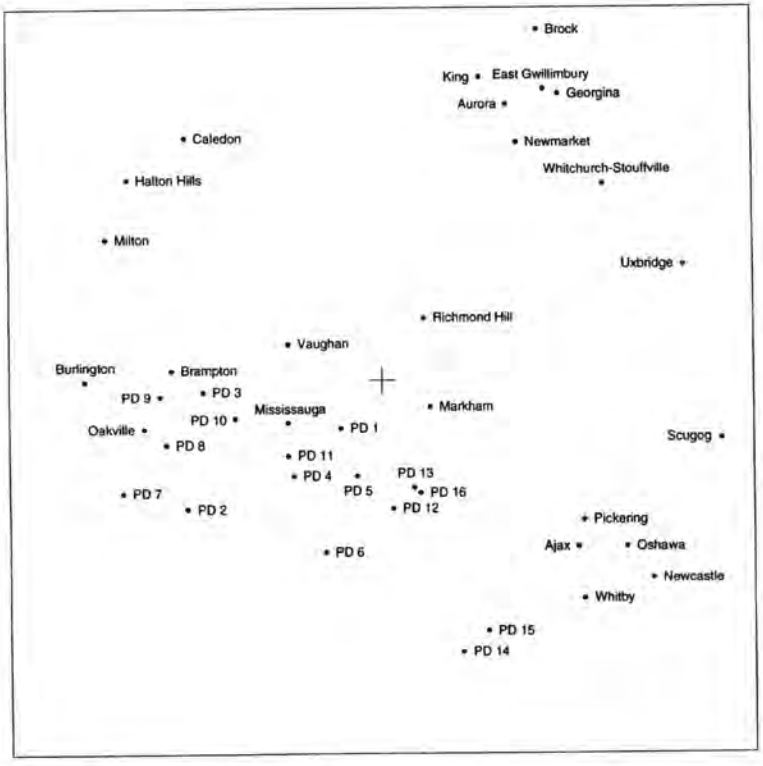


Figure 4
Destination structure of journey-to-work flows among Greater Toronto Area zones, 1991.

Table 5
Number of Jobs and Self-containment for Greater Toronto Area Zones, 1986 and 1991

	Number of Jobs			Self-containment (%)		
GTA Zone	1986	1991	Change	1986	1991	Change
Planning District 1	340,004	361,860	21,856	11.0	12.5	1.5
Planning District 2	39,115	36,633	- 2,482	26.9	33.0	6.1
Planning District 3	66,612	57,221	- 9,391	28.0	24.9	- 3.1
Planning District 4	76,816	74,307	- 2,509	19.7	18.4	- 1.3
Planning District 5	52,081	58,227	6,146	13.4	16.1	2.7
Planning District 6	36,868	32,258	- 4,610	36.3	38.2	1.9
Planning District 7	22,696	20,650	- 2,046	19.7	16.5	- 3.2
Planning District 8	59,071	58,649	- 422	23.0	24.2	1.2
Planning District 9	51,981	44,982	- 6,999	16.3	18.2	1.9
Planning District 10	75,243	78,557	3,314	24.2	20.4	- 3.8
Planning District 11	40,577	51,188	10,611	19.2	18.4	- 0.8
Planning District 12	26,132	33,874	7,742	12.1	12.6	0.5
Planning District 13	84,059	79,151	- 4,908	25.0	24.5	- 0.5
Planning District 14	7,004	8,744	1,740	21.4	20.3	- 1.1
Planning District 15	9,004	10,360	1,356	28.7	33.2	4.5
Planning District 16	43,504	57,922	14,418	23.5	22.7	- 0.8
Brock	1,717	1,439	- 278	80.0	62.7	- 17.3
Uxbridge	2,012	4,245	2,233	51.7	26.5	- 25.2
Scugog	2,246	2,802	556	71.6	62.7	- 8.9
Pickering	12,875	20,531	7,656	26.2	29.2	3.0
Ajax	10,518	12,222	1,704	32.6	32.9	0.3
Whitby	12,618	16,491	3,873	33.6	35.0	1.4
Oshawa	43,228	43,512	284	63.4	57.3	- 6.1
Newcastle	8,491	7,887	- 604	50.1	59.8	9.7
Georgina	3,188	3,894	706	78.7	72.4	- 6.3
East Gwillimbury	1,025	1,645	620	28.7	38.5	9.8
Newmarket	10,378	11,906	1,528	45.2	37.3	- 7.9
Aurora	5,933	8,032	2,099	33.8	36.0	2.2
Richmond Hill	15,036	21,255	6,219	34.3	25.2	- 9.1
Whitchurch-Stouffville	3,638	4,271	633	31.7	30.7	- 1.0
Markham	48,613	67,362	18,749	20.8	21.2	0.4
King	2,156	2,317	161	33.8	27.7	- 6.1
Vaughan	40,609	54,367	13,758	11.9	18.1	6.2
Caledon	4,137	5,359	1,222	48.4	63.6	15.2
Brampton	58,121	69,894	11,773	58.2	58.1	- 0.1
Mississauga	155,173	194,309	39,136	46.1	45.1	- 1.0
Halton Hills	7,172	5,012	- 2,160	74.0	58.2	- 15.8
Milton	7,352	8,686	1,334	63.4	54.2	- 9.2
Oakville	29,151	31,490	2,339	49.7	48.8	- 0.9
Burlington	23,113	28,901	5,788	84.5	83.7	- 0.8

Newmarket, Aurora), Eastern (Scugog, Newcastle, Oshawa, Whitby, Pickering, Ajax, Planning Districts 6, 14, 15), and Western (Caledon, Halton Hills, Milton, Burlington, Oakville)—exhibit strong internal and weak external journey-to-work flows. They comprise residential zones located in the periphery of the Greater Toronto Area. In contrast, two groups—East-central (Planning Districts 4, 5, 11, 12, 13, 16) and West-central (Brampton, Mississauga, Planning Districts 2, 3, 7, 8, 9, 10)—exhibit strong internal, but moderate external flows. They comprise zones that are centrally located, fulfilling both residential and employment functions. The proximity of the East-central and West-central groups to the primary employment cluster suggests that Planning District 1 is an important destination for work trips.

The configuration for 1991 (Figure 4) differs from that for 1986 in several respects. The movement of Richmond Hill toward the centre of the configuration suggests that it has become more important as a workplace. In fact, its self-containment decreased by 9.1 percent over the five-year period (Table 5). In contrast, the outward movements of Planning District 1, Vaughan, and Markham suggest their relative importance as workplaces has decreased, as evidenced by their increases in self-containment (Table 5). The Western group of zones no longer exists—Burlington and Oakville are associated with the West-central group, while Milton, Halton Hills, and Caledon have become more peripheral. The reason for this is that the strong internal flows that defined the Western group in 1986 weakened over the five-year period. Change in the relative location of Oshawa suggests that it has become more important as a destination for workers residing in the Eastern group of zones. In fact, the number of trips to Oshawa, from Pickering, Newcastle, and Whitby, increased substantially over time. Changes in the relative locations of Pickering and Ajax suggest that they have increased their flows to Planning Districts 12, 13, and 16. In contrast, zones of the East- central and West-central groups have experienced little change in their relative locations because, in both years, they received flows from most other zones in the Greater Toronto Area. Therefore, their freedom to vary in the configuration was limited. There are, however, three notable exceptions to this trend—Planning Districts 4, 10, and 11.

Multidimensional scaling of journey-to-work flows has proven useful for examining the structural evolution of the Greater Toronto Area at the level of GTA zones. Comparison of the configurations presented for 1986 (Figure 3) and 1991 (Figure 4) revealed considerable changes in the relative locations of many zones. As expected, such changes were more pronounced for zones located in the periphery of the region than for those with central locations. The reason for this is that an urban area's physical growth occurs outward from an established core. At first, this growth comprises residential construction because the movement of jobs follows the movement of people to the suburbs (Muller 1976; 1986). Therefore, the urban periphery can be considered chaotic in terms of spatial structure because the linkages between home and work are in a continual state of flux. In contrast, the spatial

structure of the urban core remains relatively stable over time because the high densities for economic and residential activities allow for the establishment of strong linkages between home and work.

Conclusion

This investigation has documented the structural evolution of the Greater Toronto Area, from 1986 to 1991, by investigating changes in journey-to-work flows for functional regions and GTA zones. With respect to functional regions, the relative decline of economic and residential activities in the Central Region, accompanied by their relative growth in peripheral regions, resulted in the changing orientation of journey-to-work flows. For most peripheral regions, the proportion of work trips originating from the Central Region decreased over the five-year period. This relative decline in core-to-periphery movements was replaced by increased flows among peripheral regions. As shown in Table 6, from 1986 to 1991, movements from the Central Region to other regions decreased by 0.2 percent, while within-periphery flows increased by 2.4 percent. In addition, Region 1's self-containment decreased over time. Consequently, the proportions of work trips originating from most other regions increased with respect to this destination. In fact, from 1986 to 1991, within-core movement decreased by 3.1 percent, while periphery-to-core flows increased by 0.9 percent (Table 6). Clearly, the changing orientation of journey-to-work flows emphasized the urban periphery as a place of work and residence. In turn, when comparing the two-dimensional, non-metric scaling solutions presented for 1986 and 1991, this finding was responsible for changes observed in the relative locations of many GTA zones.

Table 6
Orientation of Journey-to-work Flows for Functional Regions in the Greater Toronto Area, 1986 and 1991

	Number of Work Trips			Percentage of Work Trips		
Orientation	1986	1991	Change	1986	1991	Change
Within core	627,903	639,463	11,560	40.9	37.8	- 3.1
Within periphery	543,252	637,753	94,501	35.3	37.7	2.4
Core to periphery	148,841	160,146	11,305	9.7	9.5	- 0.2
Periphery to core	217,291	254,522	37,231	14.1	15.0	0.9
Total	1,537,287	1,691,884		100.0	100.0	

References

Davies, W. and Musson, T. (1978). Spatial patterns of commuting in South Wales, 1951-1971: A factor analysis definition. *Regional Studies*, 12: 353-366.

Goodchild, M. and Janelle, D. (1988). Specialization in the structure and organization of geography. *Annals of the Association of American Geographers*, 78: 1-28.

Hanson, S. and Pratt, G. (1988). Reconceptualizing the links between home and work in urban geography. *Economic Geography*, 64: 299-321.

Muller, P. (1976). *The Outer City: Geographical Consequences of the Urbanization of the Suburbs*. Washington, D.C.: Association of American Geographers.

Muller, P. (1986). Transportation and urban form: Stages in the spatial evolution of the American metropolis. In S. Hanson (ed.), *The Geography of Urban Transportation*. New York: Guilford Press, 24-48.

Municipality of Metropolitan Toronto (1992). *The Liveable Metropolis*. Toronto: Metropolitan Corporate Services Department.

Nader, G. (1980). An economic regionalisation of Canada: The validity of provinces as regions for the conduct of regional economic policy. *Canadian Journal of Regional Science*, 3: 117-138.

Nader, G. (1981). The delineation of a hierarchy of nodal regions by means of higher-order factor analysis. *Regional Studies*, 15: 475-492.

O'Connor, K. and Maher, C. (1979). Change in the spatial structure of a metropolitan region: Work-residence relations in Melbourne, 1961-1971. *Regional Studies*, 13: 361-380.

Ontario Ministry of Transportation (1992). *The 1991 Transportation Tomorrow Survey, Version 2.1 Data Guide*. Toronto: Ontario Ministry of Transportation.

Vickerman, R. (1984). Urban and regional change, migration and commuting: The dynamics of workplace, residence and transport choice. *Urban Studies*, 21: 15-29.

Acknowledgements

I am indebted to my Master's supervisor, Dr. Donald G. Janelle, whose support and suggestions were instrumental in completing this research. Additionally, I would like to thank the Canadian Transportation Research Forum for scholarships received throughout my Master's training in transportation geography.

USING STATE DEPENDENT UTILITY FUNCTIONS IN THE CONTEXT OF ROAD SAFETY: THEORY AND EVIDENCE

El Bachir BELHADJI

Center for Research on Transportation
Department of Economics
University of Montréal

Introduction

Since Arrow (1974), an increasing number of economists have studied the question of the optimal amount of insurance. This amount is in fact obtained when the marginal utilities in the different states of the world are equal. If state independent utility functions are used, then complete insurance is optimal. If marginal utility functions are altered by the different states of the world, then the former solution is not optimal; therefore we face two situations: first, if the "bad" state decreases the marginal utility, the optimal solution is less than complete insurance; the second situation pertains to the case where the "bad" state increases the marginal utility, hence optimality is obtained when insurance is more than complete.
Therefore, if we have to deal with states that could possibly influence the marginal utilities, determining the direction of the movement (increases or decreases) is central to understanding the questions related to optimal insurance.

A Simple Insurance Model

In the economic literature, we have long been carrying analyses supposing the preferences of the agents to be state independent; the central point was the result (or the consequence) attached to the state but not the state as such. Therefore, a bet which yields an ice-cream if the day is hot and a cup of soup if the day is cold will be the same according to this literature as a bet which yields an ice-cream if the day is cold and a cup of soup if the day is hot. Nonetheless, common sense tells us that these bets are not identical because an "ice-cream in a hot day" is a different consequence than an "ice-cream in a cold day". We can see in this example that it is the state of nature that makes all the difference. In a similar fashion, a life insurance indemnity in case of death is a different

consequence than the same indemnity in case of life. We think that analyses where preferences are state dependent should be more frequent. This should apply specifically to all cases which contain an undesirable physical (and/or psychological) risk. In general, the undesirable events which involve physical risk, affect both the physical and the financial conditions of the victim. State independent approaches take into account only the financial aspect of the consequence and ignore the physical aspect. The state dependent approach takes into account these two aspects simultaneously.

Notation of the insurance model

Two states of the world are possible in the model: the state "accident" will be obtained with probability p, and the state "no accident" has a probability of occurrence of $1-p$. It is a one period model, where the individual is concerned not only with the financial outcome associated with the state, but also with the realization of the state itself. Thus, for the same level of wealth, the individual will prefer the state "no accident" to the state "accident". After admitting that the monetary amounts in case of "accident" are different consequences than the same monetary amounts in case of "no accident", we define two utility functions: $V(S_1)$, for the accident event and $U(S_2)$ for the "no accident" event. If the state "accident" occurs, the individual will incur a monetary loss equivalent to an amount $l \succ 0$; it is the case for example of a property damage to a vehicle caused by a traffic accident. This same individual face the following choice problem: on one hand, he chose to protect himself against the "accident" event by buying quantities of protection $x \succ 0$; on the other hand, he chose a level of insurance by buying from his insurer a quantity of insurance coverage $q \succeq 0$, against a premium $y \succ 0$. The utility level will then be U(w-y-x) in case of a "no accident" and V(w-y-x-l+q) in case of an "accident", where w is the initial wealth. The behavior of the individual is supposed to obey the rationality axioms. The individual maximizes the expected utility function defined by:

$$EU = p(x)\ V(S_1) + (1 - p(x))\ U(S_2)$$

where
S_1 = w-y-x-l+q (final wealth in case of an "accident")
S_2 = w-y-x (final wealth in case of a "no accident")

Hypotheses of the insurance model

The model is based on a certain number of fundamental hypotheses upon which the results are derived. Primarily, they concern the behavior of the utility

functions and the accident probabilities. It is supposed that the utility functions U and V are twice differentiable, increasing (U'(S) > 0 and V'(S) > 0) and strictly concave (U''(S) < 0 and V''(S) < 0). The accident probability p depend on selfprotection activities x. We suppose that p(x) is decreasing (p'(x) < 0) and strictly convex in x (p''(x) > 0). Also, the individual is supposed to prefer the "no accident" state to the "accident" state, such that U(S) > V(S) $\forall S > 0$.
If we suppose actuarial insurance, we will have:

$$q(y,p) = \frac{y}{p}$$

After these different notations, the expression of EU takes the form:

$$EU = p(x)\ V[w-y-x-l+q] + [1-p(x)]\ U[w-y-x]$$

and

$$y = p(x)\ q$$

Although, we give the solution for the optimal amount of selfprotection, we will focus on the solution for the optimal amount of insurance.

$$\underset{x,q}{MAX}\ EU = p(x)\ V[w-y-x-l+q] + [1-p(x)]\ U[w-y-x] \qquad (2.1)$$

under the constraint:

$$y = p(x)q \qquad \text{(nul profits)}$$

The first order conditions relative to selfprotection and insurance respectively are:

$$p'(x)\ [V-U] - p'(x)q\ EU' - EU' = 0 \qquad (2.2)$$

$$p(x)\ [1-p(x)]\ [V'-U'] = 0 \qquad (2.3)$$

<u>The optimal amount of insurance</u>

The optimal amount of insurance is given by equation (2.3), respecting thus the equality:

$$V'\ [w-y-x-l+q] = U'\ [w-y-x]$$

i) First possible solution:
if $V'(S) = U'(S) \quad \forall S$ then $q^* = l$

ii) Second possible solution:
if $V'(S) < U'(S) \quad \forall S$ then $q^* < l$[1]

iii) Third possible solution:
if $V'(S) > U'(S) \quad \forall S$ then $q^* > l$

The proof is analogous to that of the previous case

While in the first case, $q^* = l$, the individual is compensated up to the total financial loss he incurs, in the second case, $q^* < l$, he receives only a fraction of that loss. However, in the third case, he receives not only the total amount of the financial loss, but also an additional amount to compensate him for the loss of utility following the realization of the "bad" state ($q^* > l$).

To test for the relative magnitudes of the marginal utility functions (i or ii or iii above), we developped a model where an individual is willing to pay a certain amount (*WTP amount*) to improve his safety; this is done in the first section. To know these amounts (and the values of other variables), we developped a questionnaire to elicit *WTP* (willingness to pay) amounts to decrease accident probability; this is achieved in the second section. The results of the simple model are reported in the third section. Finally, extentions of the initial model are put forward and the same hypotheses are tested again using the data from the questionnaire.

1. THE SIMPLE MODEL: USING STATE-DEPENDENT UTILITY FUNCTIONS

In order to measure the relative magnitudes of the utility functions in the different states of the world, we begin our analysis with a simple two states

---

[1] We have: $V'(S) < U'(S)$. Let us put:
$S = w\text{-}y\text{-}x$
$\Rightarrow \quad V'(w\text{-}y\text{-}x) < U'(w\text{-}y\text{-}x)$
As the marginal utility is decreasing, we need only to subtract a positive quantity $(l\text{-}q)$ from the arguments of V' to have the following equality:

$$V'[w\text{-}y\text{-}x\text{-}(l\text{-}q)] = U'(w\text{-}y\text{-}x)$$

where:
$l\text{-}q > 0 \Rightarrow l > q \Rightarrow q^* < l$

model ("accident" and "no accident").
First, we begin by taking into account the values of the accident probabilities (first circles in the appendix), and omit those of the conditional probabilities (middle circles in the appendix). As a consequence, we will have two utility functions; one in the event of accident named V, and the other associated to a "no accident" event named U.

1.1 First-order Taylor's series expansions

As a first step we will be using a first-order series expansions, where utility functions are implicitly assumed to be linear. Later, we will complicate the model by introducing second-order series.

1.1.1 Theoretical model:

Suppose a representative individual facing the following choice problem: in the first situation, the accident probability P_1 prevails; he keeps his income Y in case of a "no accident" or holds an income less than Y in case of an accident. The second situation is an active one, where he is willing to pay a certain amount (d) for the probability to decrease to P_2 (such that $p_2 < p_1$); in the two states of the world, his income will be decreased by this amount because payment is made *ex-ante*. At the equilibrium, he will be indifferent as to which state of the world occurs. Formerly, this can be expressed as:

$$(1-P_1)\,U(y) + P_1 V(yr) = (1-P_2)\,U(y-d) + P_2 V(yr-d) \qquad (4.1)$$

where V and U are respectively the utility functions in case of "accident" and "no accident"; "r" is he compensation rate offered by the insurers (SAAQ: Société de l'Assurance Automobile du Québec, in our case): it is less than unity. While the left hand side of the equality represents the first situation above, the right hand side represents the second one. If we operate first order Taylor's series around Y in the utility functions, we obtain:

$$\begin{aligned} V(yr) &= V(y) + (yr-y)\,V'(y) \\ U(y-d) &= U(y) - d\,U'(y) \\ V(yr-d) &= V(y) + (yr-d-y)\,V'(y) \end{aligned} \qquad (4.2)$$

Replacing those expressions in equation (4.1), we have:

$$\begin{aligned} (1-P_1)\,U(y) + P_1[V(y)+(yr-y)\,V'(y)] &= (1-P_2)\,[U(y)-dU'(y)] \\ &\quad + P_2[V(y)+(yr-y)V'(y) - dV'(y)] \end{aligned} \qquad (4.3)$$

While the variables of this equation are known, the parameters of the utility functions are not. These parameters are to be estimated using the following notation:

$$\begin{aligned} \beta_1 &= U(y) - V(y) \\ \beta_2 &= U'(y) \\ \beta_3 &= V'(y) \end{aligned} \tag{4.4}$$

Since the dependent variable is represented by the *willingness to pay* "d", solving equation 4.3 yields:

$$d = \frac{(P_2 - P_1)\left[\beta_1 - (yr - y)\,\beta_3\right]}{P_2\left[\beta_2 - \beta_1\right] - \beta_2} \tag{4.5}$$

1.1.2 Tests

Adding an error component ε and a subscript "i" for the i^{th} individual, the above equation becomes:

$$d_i = \frac{(P_{2i} - P_{1i})\left[\beta_1 - (y_i r_i - y_i)\,\beta_3\right]}{P_{2i}\left[\beta_2 - \beta_1\right] - \beta_2} + \varepsilon_i \tag{4.6}$$

where ε_i is an *i.i.d.* error term. This equation will be estimated using non linear least squares. Due to the structure of the equation, we can only estimate two of the three parameters; therefore, we set:

$$\beta_2 = U'(y) = 1, \tag{4.7}$$

we keep the same relative conclusions if we compare the estimated parameters β_1 and β_3 to 1. The question of whether to use state dependent utility functions is answered through the test:

$$\beta_1 = U(y) - V(y) > 0 \tag{4.8}$$

We expect β_1 to be positive and significantly different from zero, because one experiences more utility in case of a "no accident" than in case of an accident (with the same income in both states). This test is equivalent to a dependency test with respect to utility functions applied to road safety.
The second test to be used is whether the "accident" event reduces (or rises) the marginal utility; this is attained by testing:

$$\beta_3 \geq (or <) \beta_2 \qquad (4.9)$$

or equivalently:

$$\beta_3 \geq (or <) 1 \qquad (4.10)$$

While there is a consensus relatively to the sign of β_1, it is not so for the case of the sign of β_3: earlier models treating "bad" events as being equivalent to a loss in income, admit implicitly that marginal utility in case of an accident is greater than that in case of "no accident" (or, that β_3 is greater than β_2). Let us mention that there is no theoretical basis sustaining the relative magnitude of both β_3 and β_2.

2. DEVELOPMENT OF THE QUESTIONNAIRE

To estimate equation (4.6) we need to know the dependent variable d_i, as well as some other (independent) variables such as income, age, probabilities...etc. We have developped a questionnaire where people were asked their *WTP* (or d_i) for an additional safety (reduction in accident probability). Before reporting the results of the estimation, we want to shed some light on the sampling procedure and the presentation of the probabilities.

2.1 Sampling procedures

The questionnaire we developed have been confided to a search firm to be run in the Montreal area. We will describe in this section our target population, the sampling base and the visual aids (probability circles).

2.1.1. Target population

The basic goal of the project was to elicit and study *willingness to pay* amounts for road safety. We also wanted to know if people living close to a "black spot"[2] are willing to pay an amount greater (less) than that of the rest of the population to improve their safety.
The target population is the one living in the urban part of Montreal; it has been divided into two sub-populations: one living near the "black spots", and the other

[2] A black spot is an intersection in the territory of the Montreal urban community, which recorded at least 21 traffic accidents per year during three consecutive years (1984-1985-1986). More detail on these black spots is found in the work of Messier (1989).

living elsewhere in the Montreal Urban Community (M.U.C). Geographically, this territory corresponds to the center part of the Montreal Island and contains the following towns: Côte St-Luc, Hampstead, LaSalle, Montreal, West-Montreal, Outremont, Ville St-Laurent, Ville St-Pierre, Verdun and Westmount.

2.1.2 Sampling base

Two sampling bases were carried out, each of them corresponds to one of the two sub-populations described above. The survey around the black spots was realized by enumeration of the residential blocks. We have identified four black spots, and asked the firm to complete 50 interviews in each of them. The identification of the population around a particular black spot was limited inside a circle within a ray of 500 meters. Twelve starting points were chosen inside the circle, each of them has two coordinates: an angle and a distance (for example 40° and 0.15 km). The closest address serves as a starting point for a cluster of 5 interviews. The interviewer enumerate residential units from right to left, and from lower to upper apartments inside a building. The first selected unit will be number *i*, where *i* is a random number selected (by computer) at the same time as the coordinates. The following units were selected by incrementing this number by 5. Then, the interviews were conducted just after the enumeration, and the absent households were contacted later. Concerning the survey outside the black spots, the methodology was completely different: a random sample of telephone numbers was selected (by computer) from the telephone exchanges list in the concerned area. The households were then contacted by phone to arrange for appointments. Considering our budget constraint, we have asked the search firm to complete 200 interviews in this territory (outside the black spots), bringing the total to 400 for the entire population.

We have developed three versions of the questionnaire, which we called "designs". Each design point is a combination of two types of risk: a base risk (or a probability of an accident) and a conditional risk (or a probability of dying if the accident has occurred). Table I shows the distribution of the different design points into the sub-samples E1, E2 and E3. Recall that each household receives only one of the designs. While the E2 design has been constructed using actual probabilities, the other two (i.e E1 and E3) have been designed to reveal the importance of the base and conditional risks. To have a sufficient number of observations in the black spots, all the residents around these spots have received the E2 design. As for the remaining respondents, they have received (randomly) one of three designs. As shown in table I, the two levels of risk per vector mean that each respondent reveals two valuation amounts (or *willingness to pay,"WTP"*, amounts). For example, the respondent receiving version E2 (in vector II) will supply one *WTP* for a reduction in his base risk from 1/10 to 1/20 and another for a reduction from 1/20 to 1/80, with a conditional risk being equal in the two cases to 1/500.

TABLE I: Distribution of the sample across the different designs

Reduction in the base (or accident) risk			Conditional risk of dying if an accident occurs	
Vector	From	to	1/500	1/10
I	1/5 1/10	1/10 1/40	—	E1 n=82
II	1/10 1/20	1/20 1/80	E2 n=256	E3 n=63

2.2 Some information on the probability circles

The scenarios we designed imply handling probabilities at two different levels: accident probability and a probability of dying if an accident occurs. From these probabilities we deduce the probability of dying from an accident (traffic accident), which is the multiplication of the previous two probability levels. We have presented the probabilities in the form of circles, because we thought they would be easy to understand given the similarities with a hazard game wheel (see cards in the appendix). A text was read by the interviewers at the same time the circles were handled to the respondent[3]. In this text, we can read: "We can think of risk as an arrow spun inside a circle. The larger the dark portion, the bigger the chances are that the arrow will stop in this section". The respondents were then given examples for each circle on the card. The interviewers also explained the circles, emphasizing on the second and third circles, more difficult to understand than the first one. The second circle was explained as a conditional risk by giving examples such as the health state of the individual at the time of the accident, the resistance to the impact...etc. The third circle was explained as a combined risk because it implied the combination of the two previous risks. It was also mentioned to the respondents that it is obtained by multiplying the probabilities of the first two circles. Each circle has two headers: one is linked directly to the meaning of the risk, the other giving more information on the risk. To the left of the base risk, the conditional risk and the final risk, we can read respectively "risk of accident", "risk of death if there is an accident" and "combined risk: accident and death". The title "possibilities" related to the first circle reminds the respondent that the risk of a traffic accident is among other risks and can affect anybody. The title "behaviour, physical condition and resistance" related to the second circle, explains that the mortality that happens

[3] The questionnaire material is available upon request from the author.

after an accident depends on factors we call "conditional". The third circle has as a second title "personal", meaning that this circle combine the first two circles in a global personal risk.

3. RESULTS

Table II presents the non-linear least square estimates of equation (4.6), with β_2 equals to 1. The three rows relate to different scenarios which are combinations of compensation rates and different design points (for these design points see table I). Concerning the value of "r", we assumed two scenarios: it can be either equal to 90% (compensation rate offered by the SAAQ in case of an injury), or to a percentage (r_{SAAQ}) determined by the SAAQ in case of death. The amount r_{SAAQ} is computed based on income, age and family characteristics.

Table II: the simple model:
non linear estimation of first order series expansions

Description	n	R^2	U-V (β_1)	V' (β_3)
r=90% All Designs	163	0.10	976 (8) [a] [735-1218][b]	0.055 (5) [0.033-0.076]
r=90% Designs E2 and E3	130	0.40	961 (6.5) [669-1254]	0.15 (9) [0.12-0.18]
r_{SAAQ} All Designs	148	0.18	488 (2.7) [132-844]	0.044 (6) [0.03-0.058]

(a): asymptotic "t"
(b): 95% asymptotic confidence interval

For the three scenarios (row 1 through 3), we obtain the same qualitative results. As expected, the parameter β_1 (or U-V) is positive and significantly different from zero. In the first scenario (first row), the value of the asymptotic t of β_1 is quite high, offering a good precision of the parameter β_1. This result explains that the individuals experience more well-being in case of a "no accident" than in case of an accident, which is very intuitive. Therefore, in dealing with models involving road safety we must use state dependent utility functions. Regarding

the order of magnitude of the parameter β_1, we cannot say anything about the difference in values between U and V, because the Von Neumann-Morgenstern utility functions are defined up to a positive linear transformation. Only the sign and the level of significance of β_1 will matter. The parameter β_3 (or V') is less than 1, positive and significantly different from 0 at very demanding levels (asymptotic t ranges from 5 to 9 according to different scenarios). The sign of this parameter confirms the economic theory where marginal utility ought to be positive. Reading the 95% confidence interval of this parameter, leads us to state that it is significantly less than 1. This result indicates that V' is less than U', thus rejecting the monetary loss equivalent approach. According to this approach, an accident (with negative health effects) is equivalent to a drop in income, which would enhance the marginal utility of income; the result expected would then be a value of marginal utility in case of an accident greater than that in case of a "no accident" event. Obviously, here the opposite result has obtained, and the need to use state-dependent utility functions has become necessary.
Although the design points (1 versus 2 and 3) are very different (regarding the values of the accident probabilities), we observed quite similar *WTP* amounts between them. To look for any impact of the divergence of the design points, we have used all the sample on the one hand, and a subset of this sample-namely designs 2 and 3- on the other. Because designs 2 and 3 contain less variability (in terms of accident probabilities) than that in all the sample, we observed an R^2 related to designs 2 and 3 greater (0.40) than the one related to all the sample (0.10 and 0.18). At last, the differing number of observations in the three cases (rows) in table II is due to the number of designs included, and the socio-economic variables necessary for the computation of the compensation rate "r" (such as age, family status and income). Since some of these variables were missing, we have precluded the corresponding cases from the sample.

In conclusion of this section, we can say that in using first-order series, we have obtained general results such as $U(y) > V(Y)$, and more specific results such as $U'(y) > V'(y)$ which apply to road safety. Since the most feared consequences of accidents are death and poor health, we can generalize this last result to the class of events affecting the health of the people. As we have noted earlier, this result dictates less than full insurance as the optimal solution, and rejects the monetary loss equivalent approach.

4. EXTENSIONS

We begin by using second-order series expansions to test for some non linearity in the utility functions. In addition to the accident probability, we will use the conditional probabilities in the subsequent section. To test for the effect of protection on the utility functions and their derivatives, we will introduce self-protection activities in the simple model.

4.1 Second-order series

In using first-order approximations in the previous section we have implicitly assumed that the utility functions are linear because we have omitted second (and subsequent) order terms. To take into account the possible non linearities, we are going to consider second order terms in Taylor's equation. If we operate second order expansions around Y, we will have the following approximations to the utility functions:

$$\begin{aligned} V(yr) &= V(y) + (yr-y)\,V'(y) + \frac{1}{2}(yr-y)^2\,V''(y) \\ U(y-d) &= U(y) - d\,U'(y) + \frac{1}{2}(-d)^2\,U''(y) \\ V(yr-d) &= V(y) + (yr-y-d)\,V'(y) + \frac{1}{2}(yr-y-d)^2\,V''(y) \end{aligned} \tag{4.11}$$

As the last two expressions contain quadratic terms (in d), we expect to solve a quadratic expression in "d", of the form:

$$Ad^2 + Bd + C = 0$$

Replacing the right hand sides of equation (4.11) into equation (4.1) yields (after some lengthy algebra):

$$\begin{aligned} 0.5\left[(1-P_2)\,U'' + P_2V''\right]d^2 - \left[(1-P_2)\,U' + P_2V' + \alpha P_2V''\right]d \\ + (P_1-P_2)\left[(U-V) - \alpha V' - 0.5\alpha^2V''\right] = 0 \end{aligned} \tag{4.12}$$

where the argument of U, V, U' and V' is "y" and $\alpha = yr-y$.

The solution to the above equation is obtained through solving:

$$d = \frac{[-B \pm \sqrt{\Delta}\,]}{2A} \tag{4.13}$$

where

$$\Delta = B^2 - 4AC$$

Numeric calculations led us to keep only one root:

$$d = \frac{[-B - \sqrt{\Delta}\,]}{2A} \tag{4.14}$$

the other root was removed because it yielded negative Δ's at the time of the iterations.

As with the previous case (first-order series), we have set $\beta_2 = 1$. Retrieving the values of the *willingness to pay "d"* from the questionnaire, we estimated equation (4.14) using non-linear regression. Table III reports these results.

Table III: The Simple Model with Second Order Series.

Description	n	R^2	U-V	V'	U"	V"
r_{SAAQ}	148	0.18	420	0.05	-4.5 E-11	-1.3 E-10
All Designs			(6.4)	(11)	(-8)	(-8)

Second-order expansions allow us to introduce second-order terms of the utility functions, which reflect the non linearity of these functions. Economic theory predicts that both U" and V" should be negative, but it remains silent when it comes to predict their relative magnitude.

In reading table III, we observe that the parameters β_1 and β_3 are very close to those derived with first-order approximations if we consider the same sub-sample (see third row in table II). Nonetheless, the values of the asymptotic t-ratios corresponding to the "non linear" model are much greater than those of the first model. This would presume that the specification of the second-order model is better adapted to the data we have and the problem we are studying. The parameters reflecting the non linearity -U" and V"- are close to zero, but significantly different from it. This could explain that the functions exhibit little curvature, which could be the result of small values of *WTP* amounts. The other possibility is that the *WTP* (d) amounts are better explained by first-order terms (U' and V') than by second-order terms (U" and V").

4.2. The model with conditional probability

Until now, we have used model specifications which took only into account the accident probability (first circles of the cards in the appendix); in this section, we will be using conditional probabilities as well (middle circles of the cards in the appendix). These conditional probabilities are expressed as probabilities of death conditional upon the realization of an accident. This probability will be noted P_c in the following model.

Due to the large number of parameters to be estimated, we restricted our model to the first-order approximations. Because we have introduced conditional probabilities, we will have three state-dependent utility functions: one in case of a "no accident", noted U(y), the other in case of a "non-fatal accident", noted

W(y) and the last one noted M(y) corresponds to the "fatal accident" state. The representative individual faces two situations: the first is where he keeps his levels of risk, namely P_1 and P_c (accident probability and conditional probability respectively); the second is where he engages in reducing his accident probability (from P_1 to P_2) in exchange of an amount "d". At the equilibrium, he will be indifferent between the two situations. While the first situation is reflected in the left hand side of the equation below, the second one is represented by the right hand side of this same equation.

$$\begin{aligned}(1-P_1)\,U(y) + P_1\left[(1-P_c)\,W(yr) + P_c\,M(yf)\right] = \\ (1-P_2)\,U(y-d) + P_2\left[(1-P_c)\,W(yr-d) + P_c M(yf-d)\right]\end{aligned} \tag{4.15}$$

where "r" represents the compensation rate offered (by the SAAQ) in case of a non-fatal accident, and "f" the compensation rate in case of death. Using "y" as the point of expansion, the first-order approximations of the utility functions will generate the following equations:

$$\begin{aligned} U(y-d) &= U(y) - dU'(y) \\ W(yr) &= W(y) + (yr-y)\,W'(y) \\ W(yr-d) &= W(y) + (yr-y)\,W'(y) - dW'(y) \\ M(yf) &= M(y) + (yf-y)\,M'(y) \\ M(yf-d) &= M(y) + (yf-y)\,M'(y) - dM'(y) \end{aligned} \tag{4.16}$$

Replacing the expressions of the above equations into equation (4.15), and solving for "d" yields (after some simplifications):

$$d = \frac{(P_2-P_1)\left[(U-M) - (1-P_c)(W-M) - (1-P_c)RW' - P_cFM'\right]}{P_2\left[U' - P_cM' - (1-P_c)W'\right] - U'} \tag{4.17}$$

where:
R = yr - y
F = yf - y

The parameters to be estimated in equation (4.17) are those of the utility functions and their derivatives. This equation will take the form:

$$d = \frac{(P_2-P_1)\left[\beta_1 - (1-P_c)\beta_5 - (1-P_c)\,R\beta_4 - P_cF\beta_3\right]}{P_2\left[\beta_2 - P_c\beta_3 - (1-P_c)\beta_4\right] - \beta_2} \tag{4.18}$$

where:

β_1 = U-M; β_2 = U'; β_3 = M';
β_4 = W'; β_5 = W-M.

Given the structure of equation (4.18), and the nature of the data, we can only estimate three of the five parameters. Thus we set $\beta_2=1$ as in the previous models, and $\beta_5=1$. Introducing subscripts to denote the i^{th} individual, equation (4.18) will be of the form:

$$d_i = f(X_i, \beta_1, \beta_3, \beta_4) + e_i \,, \qquad (4.19)$$

where f is a non-linear function and X_i is a vector of socio-economic variables of the i^{th} individual (containing y_i, τ_i and f_i). Given the form of equation 4.18, we will estimate it using the non linear least squares method. The results are reported in table IV.

Table IV: The model with conditional probabilities

Description	n	R_2	U-M β_1	M' β_3	W' β_4
r_{SAAQ} and r=90% All designs All sample	148	0.12	955 (6.4)	0.12 (1.3)[a]	0.045 (2.7)
r_{SAAQ} and r=90% All designs "decision-makers"	112	0.17	521 (2.1)	0.030 (0.18)[a]	0.047 (3.8)

(a) non-significant at the 0.1 level

Because we have noted some inconsistencies in the responses (*wtp* amounts) of some individuals with rather small (even nil) personal income, we constructed a sample called "decision makers", which is composed of only the respondents who have higher revenues than their spouses. Therefore, we have two sub-samples corresponding to the two rows in table IV. Because of this build-up, we lost 36 observations, but gained a greater power of explanation through a greater R^2. This result confirms the inconsistencies observed before. The two samples in table IV offer the same qualitative results: the parameters β_1 (or, U-M) and β_4 (or, W') are positive and significantly different from zero in both cases. However, the value of β_1 in the second sample is half the value of the corresponding parameter in the first sample. The values of the parameter β_4 across the samples are quite the same, with added significance level in the second sample. Yet, we cannot reject the hypothesis that the parameter β_3 (or, M') is zero in both cases. Recall that this parameter represent the marginal utility in case of death; for some authors, this result reinforces the idea that marginal utility in case of death should be equal to zero. For some others, we should use a kind of a bequest function in

problems involving death. We think that the explanation of this result could be found in the family status of the respondents (singles, married, with dependents...) and the financial situation of the family (spouses and dependents).
Finally, we could not apply second-order Taylor's series to the model with conditional probabilities because of the large number of parameters to be estimated and the structure of the equation of the *WTP*.

4.3. The model with self-protection activities

We have introduced in the questionnaire questions relating to the frequency of self-protection activities. One of the variables that seemed interesting concerned the frequency of pedestrian compliance to the green signal, noted "*f*". In order to determine the influence of self-protection activities (variable f) on the parameter estimates β_1 (or U-V) and β_3 (or V'), we have respecified the model of equation (4.6) to include this type of interaction in the following manner:

$$\beta_i = \beta_{i0} + \beta_{i1} f \qquad \textit{for } i = 1 \textit{ and } 3 \tag{4.20}$$

We estimated equation (4.6), including the above equation to take into account the differing values of the parameter estimates of the respondents who have different "quantities" of self-protection activities. Thus, the equation to be estimated is:

$$d = \frac{(P_2 - P_1)\left[(\beta_{10} + \beta_{11} f) - (yr - y)(\beta_{30} + \beta_{31} f)\right]}{P_2\left[1 - (\beta_{30} + \beta_{31} f)\right] - 1} \tag{4.21}$$

Because variable "*f*" (*compliance to the green signal*) contains six categories, and each one involves only few observations, we have decided to transform this variable into a binary variable, instead of working with five categorical variables. Thus, this variable was defined to take the value 1 if the respondents were *always* (code 1), *very often* (code 2) or *often* (code 3) compliant with the signals; it takes the value 0 if the respondents said they *sometimes* (code 4), *rarely* (code 5) or *never* (code 6) respect the signals.
Interpreting the parameters of equation (4.20) within (4.21) sums up to: if the persons who upkeep self-protection activities (f=1) incur a larger (smaller) drop in utility than the others (f=0) following an accident, then the sign of the interaction parameter β_{11} will be positive (negative). Similarly, if the individuals who use self-protection incur a larger (smaller) drop in marginal utility following an accident (than the others), then the sign of the parameter β_{31} will be negative (positive). The results of estimating equation (4.21) are illustrated below in table V.

Table V: The model with self-protection activities

Parameters and statistics	Parameters (asymptotic t)	95% Confidence Interval
β_{10}	496 (3)	163 - 828
β_{11}	539 (2.5)	110 - 969
$\beta_1^{(a)}$	781 (4.43)	428 - 1134
β_{30}	0.20 (9.7)	0.16 - 0.24
β_{31}	-0.18 (8)	-0.23 _ -0.14
$\beta_3^{(a)}$	0.10 (5.26)	0.062 - 0.138
n	161	
R^2	0.37	

(a) evaluated at sample averages

Let us compare the parameters β_1 and β_3 estimated as averages across the sample, in the table above with those appearing in table II (first row). We observe that the value of β_1 appearing in table V is contained in the 95% confidence interval of the β_1 value in table II, which imply that the two values are pretty close to each other. On the opposite, the value of the parameter β_3 in table V is not contained in the confidence interval of the β_3 parameter in table II; this difference would suggest that the inclusion of self-protection activities will raise -in average- the value of the marginal utility of income V'. All parameters in table V are significantly different from zero, including the constants B_{i0}. The sign of β_{11} is positive, which implies that the respondents who use self-protection suffer a greater loss of utility in case of an accident than those who do not (or rarely do) use safety measures. The parameter estimate β_{31} is negative, which implies that the more (less) one uses self-protection, the less (more) will be the value of his marginal utility after an accident. This supposes that the people who undergo the disadvantages (monetary and others) of self-protection are also those who are willing to take more insurance coverage after an accident, than those who do not use self-protection.

Finally, in reading table V, we can say that the quality of the regression has improved, as the value of R^2 equal to 0.10 in table II, has reached 0.37.

Conclusion

Since the parameter estimates β_1 (or, U-V) found are positive (and significant) across the different scenarios, we think that the problems related to road safety must be addressed with models where utility functions are state-dependent. This methodology should be generalized to all random events that substantially affect health. The fact that the marginal utility after an accident (β_3) is less than that in case of a "no accident" event (β_2), suggests that the accident event cannot be equivalent to a monetary loss. Applied to the insurance theory, this result dictates less than complete insurance as an optimal solution (for road hazards). Although second order terms of the utility functions are small, they passed the tests of significance. Their negative signs confirm the theory which assumes concavity and risk-aversion. As the individuals differ in their protective activities, we have introduced self-protection to detect possible links with the utility functions and their derivatives. We have found that -following an accident- the individuals who protect themselves more incur more disutility than the others. Likewise, individuals who use more protection have lower marginal utilities than their peers who use less frequently self-protection activities.
After having found the optimal compensation (insurance) rate to be less than unity, it would be interesting to know its intervals; this could be a good research avenue in the future.

REFERENCES

Arrow, K. J. "Optimal Insurance and Generalized Deductibles". Scandinavian Actuarial Journal. 1974; 1: pp. 1-42.

Cook, P. J.; Graham, D. A. "The Demand for Insurance and Protection: The Case of Irreplaceable Commodities". Quarterly Journal of Economics. 1977: pp. 143-156.

Cummings, R. G..; Brookshire, D. S.; Schulze, W. D. Valuing Environmental Goods: A State of the Arts Assessment of the Contingent Valuation Methods. Monograph; Laramie, Wyoming: U.S. Environmental Protection Agency; 1984; Cooperative Agreement No CR-811077-01-0.

Dehez, P.; Drèze, J. H. "State-Dependent Utility, The Demand for Insurance and the Value of Safety". In Jones-Lee, M. W. (ed). The Value of Life and Safety. North Holland Publishing Company; 1982: pp.41-65.

Dionne, G. "Moral Hazard and State Dependent Utility Functions". Journal of risk and Insurance. 1982; 49: pp. 405-422.

Ehrlick, I.; Becker, G. "Market Insurance, Self-Insurance and Self-Protection". Journal of Political Economy. 1972; 80 (Jul-Dec): pp. 623-648.

Jones-Lee, M. W. "The Value of Changes in the Probability of death or Injury". Journal of Political Economy. 1974 (jul-Dec): pp. 835-849.

Jones-Lee, M. W.; Hammerton, M.; Philips, P. R. "The Value of Safety: Results of a National Sample Survey". Economic Journal. 1985; 95 (march): pp. 49-72.

Lawson, J. J. The Valuation of Transport Safety. Ottawa: Département des Finances et Administration; Transports Canada; 1989; TP No 10569. (Evaluations économiques et recouvrement des coûts).

Messier, S. . Les traumatismes de la route sur l'Ile de Montréal, 1984-1986: identification et caractéristiques des sites dangereux. [mémoire.] Montréal: Département de Démographie. Université de Montréal; mars 1989; c1989. 186 p.

Smith, V. K.; Desvouges, W. H. " An Empirical Analysis of the Economic Value of Risk Changes". Journal of Political Economy. 1987; 95(1): pp. 89-114.

Société de l'assurance automobile du Québec. Bilan Routier: janvier 1991 - décembre 1991. Montréal. (Dossier statistique).

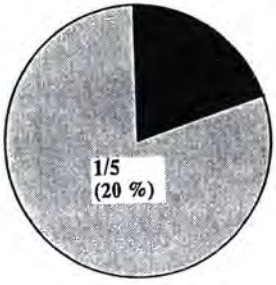
Viscusi, W. K.; Evans, W. N. "Utility Functions that Depend on Health Status: Estimates and Economic Implications". American Economic Review. 1990; 80(3): pp. 353-374.

APPENDIX: Example of designs -Design E1

CARD A

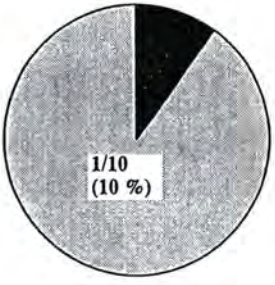
September 1991

Risk of Accidents



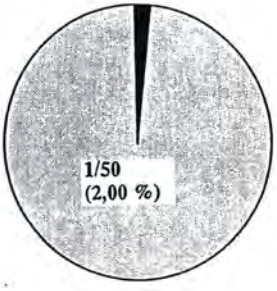
Possibilities

Risk of Death
in Case of Accident



Behavior, Physical
Condition and
Resistance
(Heredity...)

Combined Risk:
Accident and Death



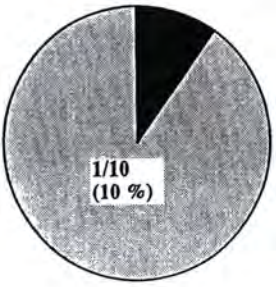
Personal Risk

APPENDIX: Example of designs -Design E1

CARD B

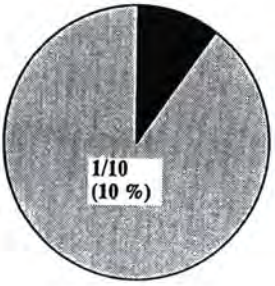
September 1992 : Improved Network

Risk of Accidents



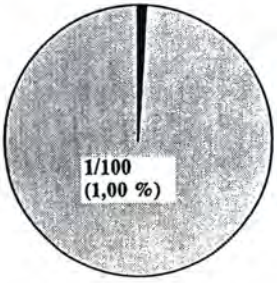
Possibilities

Risk of Death
in Case of Accident



Behavior, Physical
Condition and
Resistance
(Heredity...)

Combined Risk:
Accident and Death



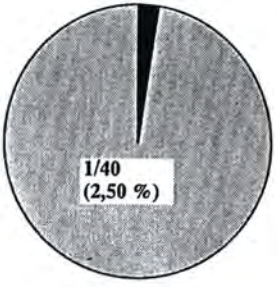
Personal Risk

APPENDIX: Example of designs -Design E1

CARD C

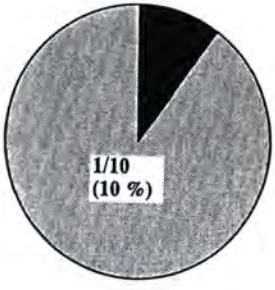
September 1993: Network Improved Upon Even More

Risk of Accidents



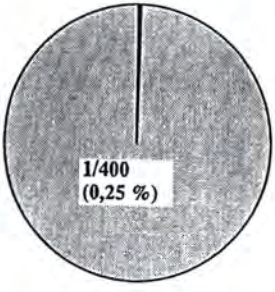
Possibilities

Risk of Death
in Case of Accident



Behavior, Physical Condition and Resistance (Heredity...)

Combined Risk:
Accident and Death



Personal Risk

Numerical Load Impact Analysis of Typical Saskatchewan Roads

Curtis F. Berthelot
Assistant Professor
Department of Civil Engineering University of Saskatchewan

1.0 Introduction

Saskatchewan's road infrastructure network is comprised of approximately 185,000 centerline kilometers of structural paved, gravel and non-structural roads [1]. Saskatchewan's road infrastructure constitutes one of the largest public investments in the province and provides vital support for the Saskatchewan economy. It is therefore the mandate of Saskatchewan Highways and Transportation (hereafter called SHT) to ensure that the road infrastructure is optimally maintained and provides the maximum socioeconomic benefits to the province, however, recent years of decreasing road maintenance budgets, increasing traffic demands and the immanent aging of the road infrastructure has made it increasingly difficult for SHT to optimize the Saskatchewan road infrastructure and guarantee optimal socioeconomic benefits for the province. To augment these current difficulties, the Canadian Federal Government abolished national transportation subsidies in efforts to rationalize transportation in Canada.

In Saskatchewan, rationalized transportation will probably result in rail branchlines to be abandoned resulting in more Saskatchewan resources (grain, timber, oil, etc) to be moved by truck over the Saskatchewan rural road infrastructure. As a result, SHT must implement road management strategies that optimize the use of the Saskatchewan rural road infrastructure to accommodate these changing traffic demands, while at the same time, ensure the road infrastructure is preserved under the increased traffic loading. To do this, efficient and equitable heavy vehicle weight regulations under a broad spectrum of changing conditions must be implemented. This will require SHT to specify heavy vehicle weight restrictions based on accurate quantification of the impact different heavy vehicles have on different road structures under varying climates and conditions with adequate reliability. Because quantifying road performance as a function of varying conditions is extremely complex and highly interrelated with several factors, mechanistic based road analysis methods are required.

1.2 Research Objectives

The specific objectives of this research are.

1. To identify factors to be considered when quantifying road impacts based on mechanics.
2. To illustrate the capabilities of three mechanistic road performance analysis software programs through simple examples.

1.3 Research Scope

The following scope was employed to satisfy the objectives of this research:

1. The Saskatchewan road infrastructure was categorized into four primary road structure types: 1) structural pavement; 2) thin membrane surfaced road (TMS); 3) gravel/grid road; and 4) farm access/municipal road.
2. The effect changing climatic seasons has on road structures was incorporated into the analysis (i.e. summer-fall dry conditions, winter frozen conditions and spring thawing conditions).
3. Different heavy vehicle configurations commonly found in Saskatchewan were considered.
4. Variations in axle group weights were employed in the analysis (50%, 100% and 150% of maximum GVW).
5. The effect of varied tire pressures (i.e. 680 kpa-100 psi to 407 kpa-60 psi) on thawing low volume roads was analyzed.
6. Three industry standard road structure analysis software programs were compared in the analysis: ELSYM5 (linear elastic all layers), KENLAYER (visco-elastic asphalt concrete layer and linear elastic all other layers) and MICHPAVE (non-linear 2-D finite element-all layers).

2.0 Mechanistic Road Performance Parameters

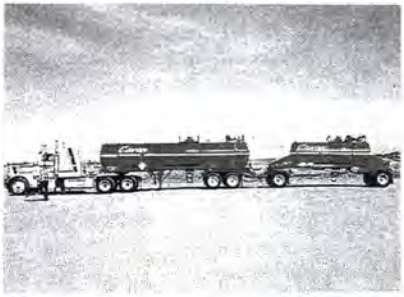
The performance of a road structure is dependent on the interrelationships of three primary factors: 1) traffic, 2) climate, and 3) road material properties and condition.

2.1 Traffic

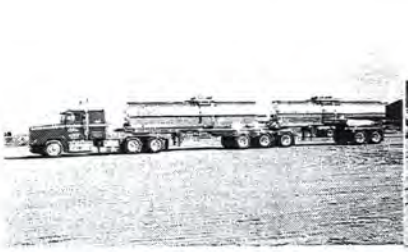
Traffic is one of the primary factors that influence the performance of a road. Heavy vehicle configuration, axle loads and vehicle properties greatly influence the damage inflicted onto a road structure. To facilitate economic growth and diversity in Saskatchewan, SHT has been proactive in permitting larger and heavier trucks as heavy truck technology has evolved with increased power and improved safety designs (i.e. b-train configuration). As a result, heavy vehicles in Saskatchewan have evolved from 5-axle semi trailer units with a gross vehicle weights of 39,500 kgs to the 9-axle B-Trains with a gross vehicle weight of 72,500 kgs. Figure 2.1 illustrates the typical heavy vehicles operating within Saskatchewan.



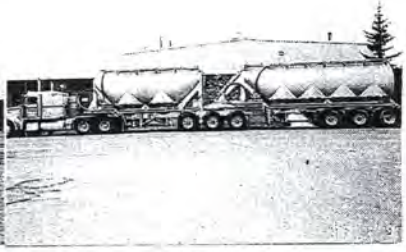
5-Axle Semi



7-Axle Rocky Mountain Double



8-Axle B-Train



9-Axle B-Train

Figure 2.1 Common Heavy Vehicle Configurations in Saskatchewan

In the past, the impact that different axle loads and configurations have on the road structure have been "normalized" through the use of load equivalencies. The most common load equivalency used is the equivalent single axle load (ESAL) derived during the AASHO Road Test during the late 1950's and early 1960's. The ESAL is based on an empirical correlation of road damage resulting from different axle loads to an 18-kip single axle load on road test sections placed and tested during the AASHO Road Test (asphalt and concrete pavement structures). The AASHO Road Test generated a fourth power relationship of incremental damage inflicted onto the pavement structure with respect to incrementally applied loads. However, because the ESAL is an average load equivalency for primary structural pavements, the ESAL is not readily applicable to determining seasonal weight limits for the non-structural roads commonly found in Saskatchewan, especially under adverse climate and load conditions. Therefore, mechanistic load equivalencies based on the actual response of the road for different heavy vehicle configurations, axle weights, road structures and

climatic conditions are required to determine seasonal heavy vehicle weight limits that optimize the socioeconomic value of the Saskatchewan road infrastructure. Figure 2.2 illustrates the dimensions and maximum allowable winter primary weights of the four common axle groups found on Saskatchewan heavy vehicles: 2-tire single steering axle, 4-tire single axle, 8-tire tandem axle and 12-tire tridem axle [2].

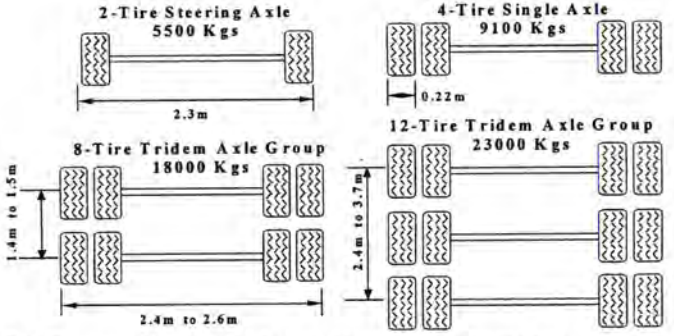


Figure 2.2 Common Heavy Vehicle Axle Groups and Dimensions

Closely spaced axles result in overlapping stresses within the road structure that cause higher stresses to be driven farther into the road structure but they also introduce confining pressures onto the road soils which increases the strength of the soil. Figure 3.3 illustrates the effect of overlapping confining stresses that occur in the road structure due to closely spaced axles.

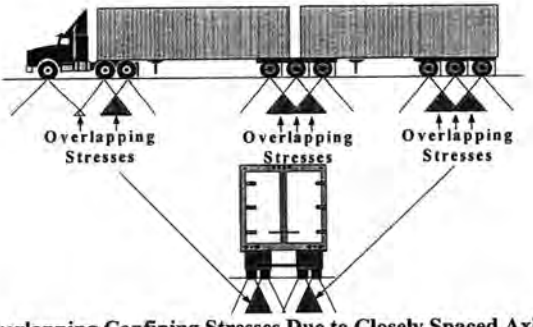


Figure 2.3 Overlapping Confining Stresses Due to Closely Spaced Axles

When a heavy vehicle is in motion, road roughness combined with vehicle harmonics result in dynamic loads at the tire/pavement interface. If the vehicle cannot readily absorb and dissipate the energy of these forces, this energy magnifies the loading of the road structure. Research conducted by the US Army Corps of Engineers for the US

Forest Service indicates reductions in the dynamic load effects due to reduced tire pressures can reduce the damage inflicted onto road structures [3]. Tire footprint area and contact pressure are dependent on four primary factors: 1) internal tire pressure, 2) applied load on the tire, 3) road stiffness, and 4) tire type (radial vs. bias ply). Figure 3.4 illustrates the effect reduced tire pressure has on load distribution and tire footprint area for a typical radial tire (the majority of heavy vehicles employ radial tires).

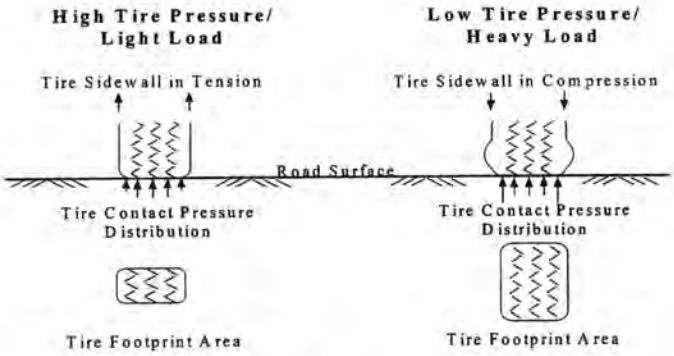


Figure 2.4 Tire Footprint and Contact Pressure Distribution

As can be seen in Figure 2.4, the effective width of a radial tire footprint does not change as the tire pressure or applied load changes. This is because the steel belts in a radial tire do not allow the tire footprint to deform in width as does occur with bias ply tires. For high tire pressures and/or light loads, the tire sidewalls are in tension, resulting in a reduced tire footprint area and a contact pressure that is less than the internal tire pressure. For low tire pressures and/or heavy applied loads the tire sidewalls are in compression, increasing the tire footprint area and the applied contact pressure approaches the internal tire pressure. In most mechanistic analyses, tire contact pressure on the road surface is assumed to be equal to the internal tire pressure. This assumption may be appropriate for rigid and flexible pavement structures, however, further research is needed to asses the impact of changing footprint pressure for non-structural pavements.

Commercial central tire inflation systems can vary the tire pressure of a heavy vehicle to as low as 50 psi while the vehicle is underway. Research conducted by the US Army Corps of Engineers indicates that decreased tire pressures can significantly reduce traffic related road damage by up to 50% on soft non-structural roads. Subsequent research performed in Nevada, Mississippi and Alabama by the USDA Forest Service determined central tire inflation to offer the following advantages.

- Reduced road distress and required maintenance.
- Eliminates wash-boarding on gravel roads.

- Increases road's load carrying ability.
- Reduces wear and vibration to the vehicle chassis.
- Improves vehicle traction in diverse conditions.
- Decreases vehicle rolling resistance in soft roads.
- No significant difference in tire wear.
- Decreases loss of aggregate on gravel roads.
- Reduced tire punchouts and tears on rough roads.
- Decreases vehicle maintenance and repair.

Research conducted by Scullion and Bush [4], showed decreasing the tire pressure of a heavy vehicle from 100 psi to 40 psi resulted in decreased vertical strain on the subgrade of 10%. Subsequent research conducted by Tayyeb, Scullion and Smith [5] revealed that at highway speeds, subgrade strains were primarily a function of vehicle speed, axle load and asphalt temperature. Tire pressure was found not to be a significant player in reducing vertical strain at the top of the subgrade, however surface shear forces were significantly reduced with lower tire pressures which would preserve the surface integrity of the road. This correlates well with the elementary theory of resilient behavior and viscoelastoplastic flow of materials in that rate and duration of load is clearly the dominant factor that affects the amount of plastic deformation realized by a material. In other words, the benefits of reducing the tire pressure of a heavy vehicle can be drastically compromised if the speed of the vehicle is reduced to accommodate the physical constraints of lower tire pressures.

The suspension system employed by a heavy vehicle can significantly improve the road "friendliness" of the vehicle and reduce the dynamic load effects of the vehicle. Therefore, heavy vehicle suspension type should be considered when determining mechanistic load equivalencies for different road types under different conditions. Heavy vehicle suspension systems range from rubber block and leaf-spring mechanisms to variable response time/load dependent air-spring suspensions. Figure 2.5 illustrates typical heavy vehicle suspension systems.

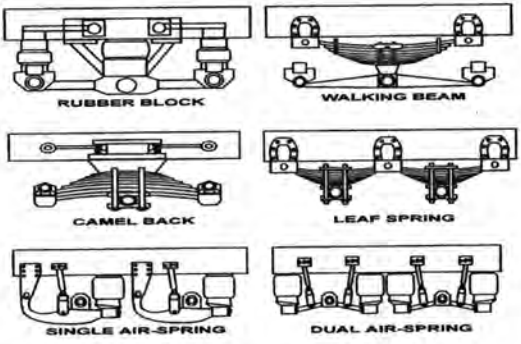


Figure 2.5 Heavy Vehicle Suspension Systems

Weigh in motion (WIM) systems have been used by SHT to quantify the dynamic load effects of different suspension systems [6]. In research conducted by Churko et al, vehicles that employed air-spring suspension systems produced much lower WIM scale system variance than other types of suspension systems. Therefore, it can be concluded that air-spring suspension systems produce less dynamic load effects over other common types of heavy vehicle suspensions. Considerable research is currently being conducted by heavy vehicle manufacturers to produce improved heavy vehicle suspensions.

2.2 Climatic Effects

Severe seasonal temperatures in Saskatchewan result in the annual freezing and thawing of Saskatchewan roads. Variations in moisture and freeze thaw can significantly effect the load carrying capacity of the road structure. Frost susceptibility of the road soils will occur if three conditions coexist: [7, 8, 9, 10].

1) Frost susceptible soils are present in the road structure (soils comprised of silty sands or silty clays);
2) Significant seasonal variations in temperature, and
3) The soil has access to free water.

Freeze thaw cycles alter the micro structure of soils within the first few freeze thaw cycles, permanently reducing the density (therefore the strength) of frost susceptible soils. Figure 2.6 conceptually illustrates the seasonal mechanics of an annual freeze thaw cycle [from 7].

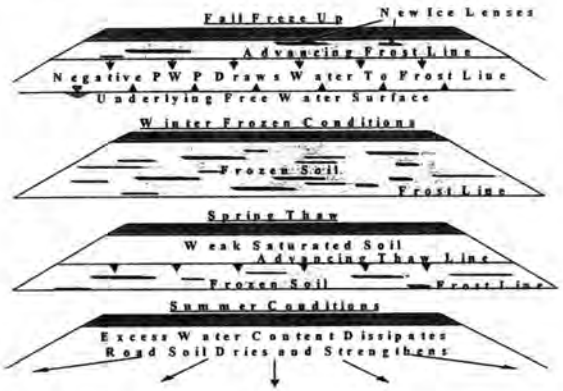


Figure 2.6 Seasonal Freeze-Thaw Process

The two primary freeze thaw mechanisms that effect road structure soils are: 1) changes in the micro-structure of the soil caused by the 9% expansion of freezing water within the soil matrix due to the opening of the water's hexagonal crystal lattice structure, and

2) inability for the soil to dissipate excess soil water content during the spring thaw [9, 10]. As illustrated in Figure 2.6, when freeze up occurs in the late fall/early winter, heat slowly escapes the road structure causing water in the soil matrix to form ice lenses. Because water and ice have a strong affinity for each other, the ice lenses draw water from the surrounding soil, causing negative pore water pressure. The negative pore water pressure draws water from underlying free water sources providing additional water for the ice lenses to grow. In severe cases where the soil is susceptible to capillary action and significant underlying free water is readily available, ice lenses may grow very rapidly causing significant frost heaves in the road surface.

During the winter months in Saskatchewan, road structures are a completely frozen matrix of soil and ice resulting in increased shear strength that is able to withstand significantly higher traffic loads without experiencing any increased damages. As a result, the province allows slightly heavier vehicle winter weights on most roads. When spring thaw occurs, heat from the road surface thaws the road structure from the surface downward. As the thaw line advances into the road structure the thawed soil above the thaw line becomes saturated with water from the melted ice. This excess water often cannot drain until the frozen underlying soil thaws therefore reducing the road's shear strength rendering it vulnerable to damage from heavy trucks. During the late spring and throughout the summer, the road structure thaws and the excess water drains from the road soils enabling the road structure to regain density and it's unsaturated state, increasing the strength of the road structure.

The degree to which frozen road structures regain its original strength varies depending on the road materials and site conditions. Spring rains can result in moisture infiltrating into the road structure and may also surge the ground water table. In Saskatchewan, it is standard road construction practice to maximize surface drainage away from the road structure and to maintain a minimum overburden of 2.4 to 3.7 meters from the road surface to the top of the ground water table.

2.3 Road Structure and Material Properties

Granular and fine grained soils typically respond to applied repeated stresses in a non-linear and resilient manner. When stress is applied to granular soils, the resilient modulus of the material increases as a function of stress because when granular material is loaded by a normal stress, the material is more resistant to shear failure because the particles are held closer together by the normal stress. For failure in shear to occur in granular materials, the particles must fracture or crawl up and over each other (dilatancy). Therefore, strong granular soils that are densely packed and have a high degree of internal friction will dilate more than weaker materials, resulting in a material that has a higher resistance to shear. Figure 2.7 illustrates granular material dilatancy under shear.

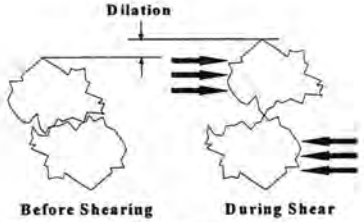


Figure 2.7 Granular Material Dilation Under Shear

When stress is applied to fine grain cohesive soils, the resilient modulus of the material decreases as a function of applied stress. For all road materials, the resilient modulus may be written as follows.

$$E_R = k_1 \rho_a \left[\frac{\theta}{\rho_a}\right]^{k_2} \left[\frac{\tau_{oct}}{\rho_a}\right]^{k_3} \quad \text{.....1}$$

Where: E_R = Resilient modulus of soil, k_i = material constants, θ = fist stress invariant, and τ_{oct} = octahedral shear stress.

Mohr-Coulomb failure theory is usually used to represent the shear strength of soils. Mohr-Coulomb assumes that the shear strength of the soil is a function of how well the soil particles hold together as a mass (cohesion), what the confining stresses of the material is, and the angle of internal friction of the soil particles. The Mohr-Coulomb equation used to express the shear strength of a material may be expressed as.

$$\tau = c + \sigma \times \tan\phi \quad \text{.....2}$$

Where τ = shear strength of soil, c = cohesion of soil, σ = normal stress to which the soil is subjected, and ϕ = angle of internal friction of the soil. Because granular material has very little inter-particle cohesion, the coefficient of cohesion for granular soils is assumed zero. However, another important factor to account for in mechanistic analysis of road materials is the effect soil suction has in strength of soils, especially fine grained soils (fine grained soils can produce suction pressure that exceeds 100,000 psi [12]). As a result, the Mohr-Coulomb relationship for shear strength of soils may be written as follows to incorporate the role soil suction, pore air pressure and pore water pressure has on soil strength properties [12].

$$\tau = c' + (\sigma_f - \mu_a)\tan\phi' + (\mu_a - \mu_w)\tan\phi^b \quad \text{.....3}$$

Where τ = shear strength of soil, c' = effective cohesion of soil, $(\sigma_f - \mu_a)$ = net normal stress state on failure plain, μ_a = pore air pressure, $(\mu_a - \mu_w)$ = matric suction, μ_w = pore water pressure, ϕ' = material property indicating angle of internal friction of the soil associated with normal stress state, ϕ^b = angle indicating material property representing

rate of increase of shear strength as a function of matric suction. Therefore it can be seen that if the pore water pressure decreases and becomes negative, the shear strength of the soil will increase as a function of the ϕ^b of the soil.

3.0 History of Mechanistic Road Analysis

Road structure performance is directly related to the mechanical responses of the road structure (i.e. strains, stresses and deflections) resulting from applied traffic loads [7]. However, road performance models have traditionally been based on empirical information correlated to general overall road performance. The most popular empirical road performance models have been formulated through extensive test road projects such as the WASHO Road Test and the AASHO Road Test. Although empirical based road performance analyses have been widely employed, road performance prediction credibility and the inherent costs of employing empirical road performance models are concerning because of several factors.

1. Collecting the required data to formulate empirical road performance models.
2. Empirical data inference space restrictions.
3. Inherent errors in empirical data collection.
4. Maintenance of empirically formulated models.
5. Variability of data collection methods (i.e. manual vs. automated road condition data collection).
6. Model instability when applied to conditions other than those for which the model was developed.

Although empirical road analysis methods have been widely used over the past few decades, recent developments in mechanistic road analysis methods is evolving the art of road design and construction from pure empiricism to more scientific based methodologies. Recent advancements in mechanistic modeling techniques, improved computational capabilities, and a better understanding of the fundamental behavior and properties of road structure materials has resulted in the development of new mechanistic based road analysis methods that replace empirical based analysis methods. Figure 3.1 illustrates the framework of mechanistic based road analysis. As illustrated in Figure 3.1, mechanistic based road performance models can be readily applied to different road types and conditions by recalibrating the model for site specific conditions. Iterative calibration cycles or Bayesian methodologies are typically used in conjunction with scientific analysis and/or knowledge regarding the fundamental performance of roads. As a result, mechanistic performance models are much less intensive to maintain and much more applicable. Mechanistic models provide an audit trail of the decision framework of which is essential when performing credible scientific analyses.

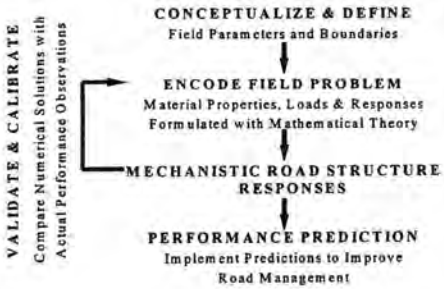


Figure 3.1 Mechanistic Modeling Framework

3.1 Linear Elastic Road Structure Mass (Boussinesq)

Mechanistic based road analysis methods have their roots in the late 1800's. Linear elastic theory was originally applied to road structures by Boussinesq in the late 1800's. Boussinesq's original formulations assumed the entire road structure to be comprised of a homogeneous isotropic linear elastic mass of material that obeys Hooke's law when subjected to an applied load:

$$E=\frac{\sigma}{\varepsilon} \quad \text{.....4}$$

where E = modulus of elasticity, σ = applied stress and ε = resultant strain. Elastic materials also exhibit a constant Poisson's ratio (ν). The Poisson's ratio is the ratio of strain normal to the applied stress divided by the axial strain of the applied stress. For elastic materials Poisson's ratio is between 0 and 0.5. Boussinesq employed the theory of an elastic half space to model stresses and strains at any location within a road structure. An elastic half space is comprised of an infinite 2 dimensional road structure with a top planer surface where traffic loads are applied as illustrated in Figure 3.2.

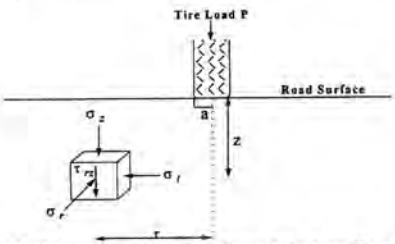


Figure 3.2 Elastic Half Space under Axi-Symetric Loading

As can be seen in Figure 3.2, any cylindrical element with center at depth z below the road surface and radius r from the vertical axis running through the center of the applied load experiences the same stresses and strains from an applied load. The primary stresses resulting from an applied load are normal stresses σ_z, σ_r and σ_t and shear stresses τ_{rz} and τ_{zr}. Boussinesq formulated the vertical stress σ_z at any depth z below the road surface and at any radial distance r from the point of load [12,13].

$$\sigma_z = \frac{3P}{2\pi z^2} \frac{1}{[1+(\frac{r}{z})^2]^{5/2}} \quad \text{.....5}$$

By integrating the vertical stress equation over a circular tire footprint area, the maximum vertical stress that occurs directly under the center of the tire footprint may be defined as:

$$\sigma_z = P[1 - \frac{z^3}{(a^2+z^2)^{3/2}}] \quad \text{.....6}$$

As can be seen from equations 5 and 6, the maximum vertical compressive stresses, strains and vertical deflection occur directly under the center of the applied load where the shear stresses are zero. These stresses and strains can be expressed from linear elastic theory as a function of the tire load (P), tire foot print radius (a) and depth (z) from the road surface. Because of symmetry $\sigma_r = \sigma_t$.

$$\sigma_t = \sigma_r = \frac{P}{2}[1+2\nu - \frac{2(1+\nu)z}{(a^2+z^2)^{0.5}} + \frac{z^3}{(a^2+z^2)^{3/2}}] \quad \text{.....7}$$

Vertical strain ε_z, radial strain ε_r and tangential strain ε_t may be expressed as:

$$\varepsilon_z = \frac{(1+\nu)P}{E}[1-2\nu + \frac{2\nu z}{(a^2+z^2)^{1/2}} - \frac{z^3}{(a^2+z^2)^{3/2}}] \quad \text{.....8}$$

$$\varepsilon_r = \varepsilon_t = \frac{(1+\nu)P}{2E}[1-2\nu - \frac{2\nu z}{(a^2+z^2)^{1/2}} + \frac{z^3}{(a^2+z^2)^{3/2}}] \quad \text{.....9}$$

Surface deflection as a function of depth and footprint area may be expressed as:

$$w = \frac{(1+\nu)Pa}{E}\{\frac{a}{(a^2+z^2)^{0.5}} + \frac{1-2\nu}{a}[(a^2+z^2)^{0.5} - z]\} \quad \text{.....10}$$

Shear stresses $\tau_{rz} = \tau_{zr}$ may be expressed as:

$$\tau_{rz} = \tau_{zr} = \frac{-3P}{4\pi E}\left[\frac{6rz^2}{R^5}\right] \quad \text{.....11}$$

3.2 Burmister's Multi-Layered Linear Elastic Analysis

Road structures are typically constructed in layers of different bound and unbound materials. As a result, Boussinesq's assumption that roads are constructed of a uniform homogeneous mass of material with half space boundary conditions is not readily applicable to modern road structures. In 1943, Burmister expanded Boussinesq's linear elastic formulations to determine stresses, strains and deflections of a two layered airport runway structure, whereby each layer has a distinct modulus, Poisson's ratio and finite layer thickness[14]. Burmister's multi-layer theory may be readily applied to road structures comprised of n-layers of different materials as illustrated in Figure 3.3.

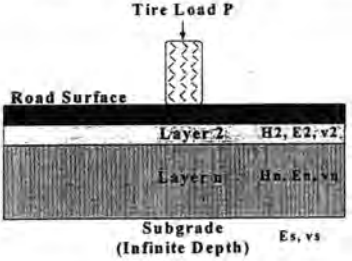


Figure 3.3 Linear Elastic n-Layered Road Structure

To apply Burmister's linear elastic theory to a layered road structure system, some basic assumptions are made:

1. Each road structure layer is homogeneous, isotropic and linearly elastic with an elastic modulus E and Poisson's ratio v.
2. Each structure layer is comprised of material that is weightless and infinite in horizontal extent.
3. Each structure layer has a finite thickness with the subgrade layer being infinite in thickness.
4. A uniform pressure P is applied to the road surface over a circular tire footprint.
5. Continuity conditions exist across the boundaries of each layer.

3.3 Non Linear Stress Dependent Material Properties

Plastic behavior in road materials is a primary cause of distress in road structures. Plastic deformation may be represented as a non-linear elastic as illustrated in Figure 3.4 for the uniaxial stress strain behavior of non-linear elastic and elastoplastic materials.

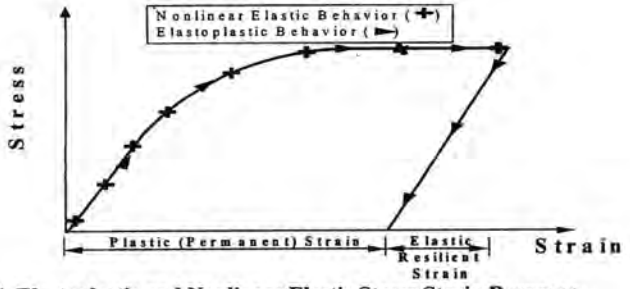


Figure 3.4 Elastoplastic and Nonlinear Elastic Stress-Strain Response

As can be seen in Figure 3.4, the loading of the two materials is identical, however, the unloading response of a non-linear elastic and elastic-plastic material differ significantly. The non-linear elastic material unloads along the exact same path to which it was loaded but the elastic-plastic material exhibits irreversible plastic behavior and unloads instantaneously on a slope parallel to the Young's modulus of the loading cycle. As a result, for any given strain in an elastoplastic material, two different stresses may be acting on the material depending if the material is being loaded or unloaded. Most roads exhibit failure over time as a function of repeated loads. Miner's cumulative damage concept was derived in 1945 for the time dependent cumulative damage inflicted onto metals under cyclical loading. In the context of roads, Miner's cumulative damage concept can be applied to quantify the effect different repeated loads under different environments have on the road structure.

$$D = \sum_{i=1}^{q} \sum_{j=1}^{r} \sum_{k=1}^{s} \frac{n_{i,j,k}}{N_{i,j,k}} \quad \text{.....12}$$

Where D = total cumulative damage, n = total applied load, N = maximum allowable load for load magnitude I, load type j and environmental condition k. Figure 3.5 illustrates a typical relationship of repeated load and viscoelastoplastic behavior.

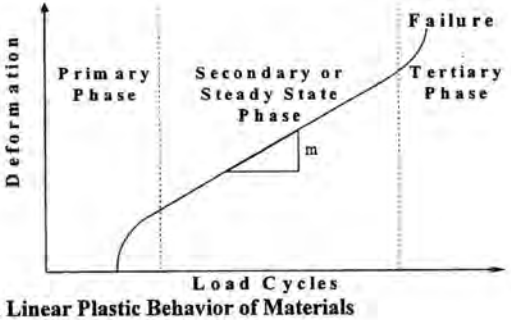


Figure 3.5 Non Linear Plastic Behavior of Materials

4.0 Analysis of Typical Saskatchewan Roads

Three mechanistic analysis software programs were examined in this analysis.

ELSYM5: Linear Elastic (all road structure layers);
KENLAYER: Viscoelastic (asphalt concrete layer) and Linear Elastic (granular base/subbase layers and subgrade);
MICHPAVE: Linear Elastic (asphalt concrete layer) and Non-Linear Finite Element (all base/subbase layers and subgrade).

4.1 ELSYM5

ELSYM5 is a linear elastic multi-layered road structure analysis program developed at the University of California at Berkeley with the capability of assessing a road structure with up to five material layers. ELSYM5 was used in this analysis to demonstrate the calculation of shear strains at the surface of the subgrade of the four road structure types under a dual tire load at maximum allowable weight limits. Shear strain at the top of the subgrade was chosen because shear is one of the primary mechanisms of load related critical failures of road structures, especially lower volume rural roads. ELSYM5 assumes all road structure layers are comprised of separate linear elastic material layers, each with a distinct modulus of elasticity and Poisson's ratio. Because seasonal variations in climate significantly effect the load carrying capability of Saskatchewan road structures, ranges of soil material strengths were used to illustrate varied soil strength during the spring thaw and dry/frozen conditions. Table 4.1 summarizes the seasonal road material properties used in ELSYM5.

Road Structure Material	Dry/Frozen Modulus (MPa) & Poisson's Ratio	Spring Modulus (MPa) & Poisson's Ratio
Asphalt Concrete	3390; 0.15	3390; 0.15
Granular Bases	344; 0.40	344; 0.40
Subgrade	68; 0.45	34; 0.45

Table 4.1 Seasonal Linear Elastic Road Structure Material Properties

As can be seen in Table 4.1 the linear elastic analysis incorporated seasonal variations in road structure performance by reducing the strength of the subgrade soil during the spring thaw.

4.2 KENLAYER

KENLAYER is a non-linear elastic road structure analysis software package developed at the University of Kentucky that employs viscoelastic theory to analyze the asphalt concrete layer. KENLAYER extends linear elastic theory to include non-linear or stress sensitive materials such as course and fine grained soils. KENLAYER employs fundamental soil properties to determine the resilient modulus of soil layers in the road

structure: (K_0, K_1, K_2, K_3, K_4, μ, deviator stress, soil cohesion, ϕ, and soil density). KENLAYER also offers the advantage of allowing individual road structure layers to be sub-layered to account for changes in resilient modulus due to changing conditions with depth (i.e. silty soil materials that are encroaching near an underlying water table).

4.3 MICHPAVE

MICHPAVE is a linear/non-linear finite element analysis program developed at the University of Michigan to asses load impacts on different road structures. MICHPAVE is very similar to ILLEPAVE in that it employs linear elastic theory to estimate stresses, strains and deflections in the asphalt concrete layer and non-linear finite element analysis techniques to analyze the granular and subgrade soil layers. MICHPAVE can also estimate the fatigue life and rut depths that will occur in the asphalt concrete, granular and subgrade layers through empirical correlations to actual observed performance of road structures.

MICHPAVE is similar to the mainframe based finite element analysis program ILLE-PAVE developed at the University of Illinois, however, MICHPAVE assumes a finite flexible boundary at some depth into the subgrade, therefore the number of elements required in the finite element analysis is significantly reduced. MICHPAVE assumes each road structure layer to be infinitely continuous in the horizontal direction and the last layer is assumed to be infinitely deep. The user must however specify a flexible boundary layer at a finite depth in the last layer (maximum 60 inches) and all stresses, strains and deflections are determined within the user specified depth of the flexible boundary. The material under the flexible boundary layer is assumed to be a homogeneous half space with a defined stiffness matrix. All stresses, strains and displacements are determined based on the application of a single circular load.

4.4 Saskatchewan Road Structures

Four road structures were used to represent the rural Saskatchewan road network: 1) structural pavements, 2) thin membrane surfaced (TMS) roads, 3) gravel/grid roads and 4) main farm access/municipal roads. Table 4.2 summarizes the design layer thicknesses for each road structure considered in the analysis.

Road Structure Type	Asphalt Concrete (mm)	Granular Base/Subbase (mm)	Subgrade
Structural Pavement	100	300	∞
Thin Membrane Surface	25	200	∞
Gravel Grid	0	200	∞
Farm Access/Municipal	0	100	∞

Table 4.2 Road Structure Layer Thicknesses

4.2 Heavy Vehicle Weights

Three ranges of heavy vehicle weights were used in the analysis: 50%, 100% and 150% of maximum allowable primary weight limits. Table 4.3 summarizes the maximum allowable heavy vehicle weights used for this analysis.

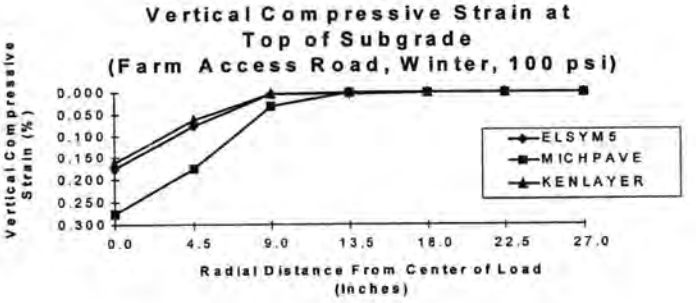
Axle Group	50% Allowable Weight (Kgs)	100% Allowable Weight (Kgs)	150% Allowable Weight (Kgs)
Steering	4500	5500	5500
4-Tire Single Axle	5000	10000	15000
8-Tire Tandem	9000	18000	27000
12-Tire Tridem	12000	24000	36000

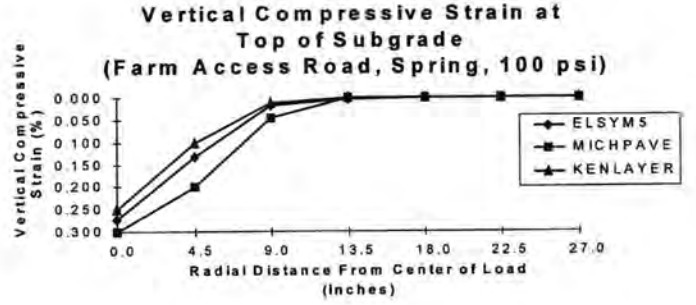
Table 4.3 Heavy Vehicle Axle Weights

4.4 Analysis Results

For illustrative purposes, an analysis determining the vertical compressive strains that results at the top of the subgrade due to an applied traffic load was determined for different scenarios using ELSYM5, MICHPAVE, and KENLAYER. Figures 4.1 to 4.3 illustrate the vertical compressive strain that results at the top of the subgrade as calculated by ELSYM5, MICHPAVE and KENLAYER for the different scenarios examined. The analysis was performed across three different road structures 1) farm access road, 2) gravel road, and 3) thin pavement structure. The analysis was performed using a standard heavy truck tire load at the maximum allowable weight limit of 4500 lbs per tire and tire pressures used were 100 psi and 60 psi.

As can be seen across Figures 4.1 to 4.3, vertical compressive strains at the top of the subgrade decrease dramatically as the structural strength of the road increases from a thin graveled farm access road to a paved road structure. This would indicate that road structure strength is highly sensitive to the road structure design and construction. The strength of the subgrade was decreased in the analysis to simulate spring thaw conditions. For purposes of his analysis, the strength of the subgrade was reduced by 50% for illustrative purposes. It can be seen in the results that decreasing the strength of the subgrade increased the resultant strain at the top of the subgrade. If the heavy vehicle tire pressures are decreased from 100 psi to 60 psi for spring road conditions to simulate the effect central tire inflation systems would have on the inflicted strain experienced by the road structure. As can be seen from the results, decreasing the tire pressure from 100 psi to 60 psi significantly reduces the resultant strain experienced by the road structure. This would suggest that central tire inflation systems may result in significant reductions in road impacts for certain road conditions, however vehicle speed must remain constant to realize the benefits of reduced tire inflation pressures. Given the significant amount of low volume non-structural roads in Saskatchewan, the use of central tire inflation systems could significantly reduce heavy truck load impacts on low volume road structures.





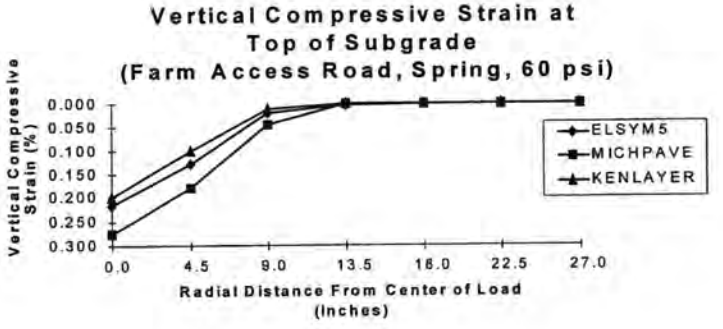
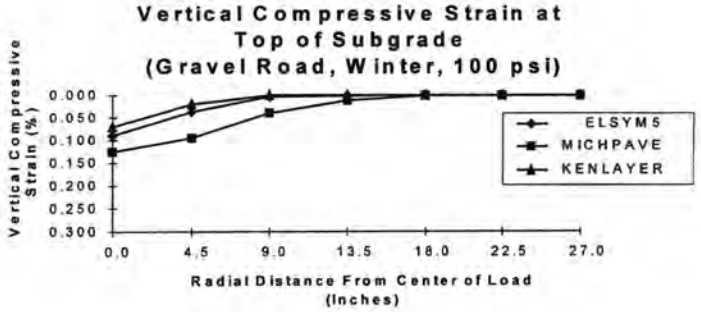
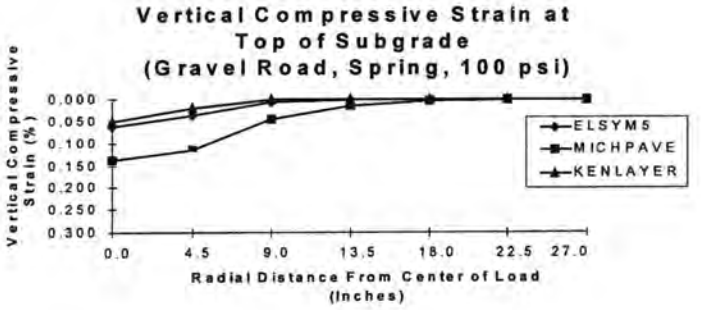


Figure 4.1 Strain at Top of Subgrade of Farm Access Road Structure





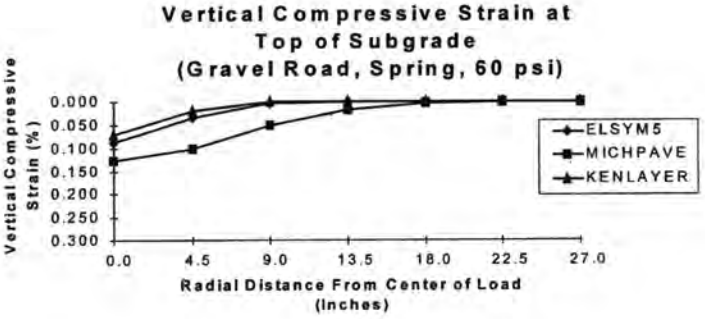
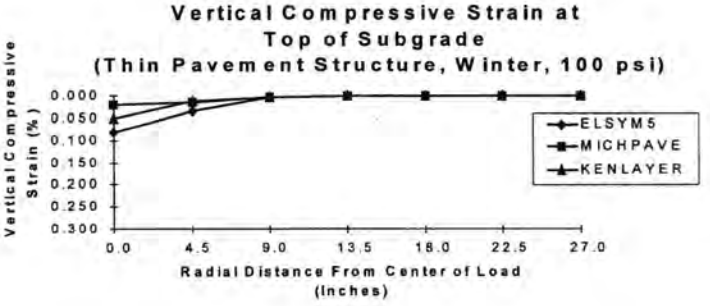
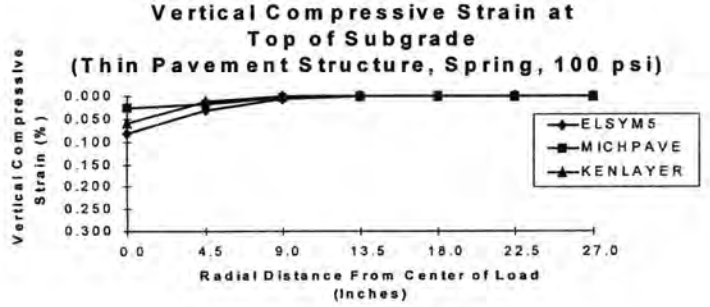


Figure 4.2 Strain at Top of Subgrade of Gravel Road Structure





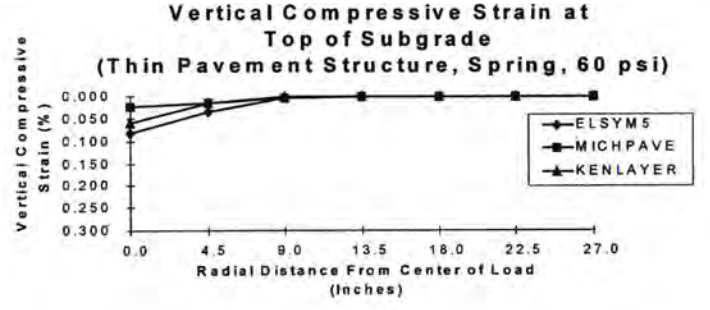


Figure 4.3 Strain at Top of Subgrade of Thin Pavement Road Structure

5.0 Summary and Conclusions

Optimized road transportation is essential for the economic well being of Saskatchewan. A key component to optimizing road transportation is first to optimize the design, construction and life cycle maintenance of the road infrastructure through sound scientific and engineering methodologies. Given recent advancements in computational abilities and a better fundamental understanding of the mechanics of road structure behavior, road infrastructure engineers are developing improved analysis tools that compare different road structures under various climatic and traffic conditions in a credible and defensible manner. These new mechanistic "scientific" based road performance analysis methodologies are replacing the traditional empirical based road assessment and design procedures. Based on the above, the following conclusions are presented:

1. Numerical road structure analysis methods exist and are an effective road infrastructure management tool.
2. Different road management and/or design strategies can be analyzed using numerical methods that specifically consider the most critical factors that effect the impact realized by the road structure.

References

1. Widger, A., Nixon, D., Sparks, G.A., Sauer, E.K.. "The Performance of Subgrades, A Research Strategy". A Presentation to Saskatchewan Department of Highways and Transportation, 1994.
2. Saskatchewan Highways and Transportation, A Guide to Saskatchewan Weight and Dimension Regulations, 1993-1994.
3. USDA Forest Service. "Central Tire Inflation- Where the Rubber Meets the Road", Video of the San Dimas Technology & Development Program, 1994.
4. Scullion, T., Bush, A.J. Use of Multidepth Deflectometer for Deflection Measurements.
5. Tayyeb, A., Scullion, T., Smith, R.E. Using Multidepth Deflectometers to Study Tire Pressure, Tire Type, and Load Effects on Pavements. TTI Report 1184, 1992.
6. Churko, B.M. et al. Study of Various Truck Suspensions on Rough Pavements Using WIM Scale. 1985.
7. Sparks, G. A., Sauer, E. K., et al. "An Analysis of Spring Load Restrictions on Yukon Highways". A Study for the Transportation Engineering Branch, Yukon Community and Transportation Service, 1991.
8. Yoder, E.J. Principles of Pavement Design. Wiley & Sons. 1959.
9. MacKay, M.H., Hein, D. K., Emery J. J. Evaluation of Frost Action Mitigation Procedures for Highly Frost-Susceptible Soils. Transportation Research Record 1362. 1994.
10. Othman, M. A., Benson, C. H. Effect of Freeze-Thaw on the Hydraulic Conductivity of Three Compacted Clays from Wisconsin. Transportation Research Record 1369. 1994.
11. Fredlund, D.G., Rahardjo, H. Soil Mechanics for Unsaturated Soils. Wiley, 1993.
12. Huang, Y.H. Pavement Analysis and Design. Prentice Hall. 1993.
13. Lytton, R. L., et al. Clusters of Terms Relevant to Pavement Performance Prediction. SHRP Contract 89-P-020 Technical Memorandum.
14. Burmister, D.M. The Theory of Stresses and Displacements in Layered Systems and Applications to the Design of Airport Runways. Highway Research Board, 1943.

.....ANNOUNCING.....

1997 CTRF ANNUAL CONFERENCE
TO BE HELD IN TORONTO
PARK PLAZA HOTEL
MAY 3-6, 1997

WATCH FOR CALL FOR PAPERS COMING IN SEPTEMBER 1996

For further information contact:

David Cape, Ryerson Polytechnic University, (416) 979-5000 phone,
(416) 979-5273 fax
Janet Weiss, CP Rail System, (905) 803-3202 phone, (905) 803-3300 fax
e.mail: ctrf@acs.ryerson.ca

CANADIAN TRANSPORTATION RESEARCH FORUM
209 - 15 Innovation Blvd.
Saskatoon, Saskatchewan
S7N 2X8
Telephone: (306) 668-2828 Fax: (306) 668-7603

INVOICE FOR 1996 CTRF MEMBERSHIP

CTRF Membership for Calendar Year 1996 ($60.00) $_____
GST at 7% ($4.20) $_____
TOTAL $_____

GST Registration No. R126274562 (If GST exempt, please indicate)

Please make cheque payable to Canadian Transportation Research Forum and mail to the address noted above. **PLEASE RETURN A COPY OF THIS INVOICE WITH PAYMENT IN ORDER THAT OUR DATABASE WILL CONTAIN ACCURATE INFORMATION RELATING TO YOUR MEMBERSHIP.**

TERMS: Net 30 days. An interest charge of 2% per month may be applied to overdue accounts.

Please complete the following to ensure accurate membership data. The information below will appear in the next published CTRF Directory.

Name: _____ Title: _____

Company: _____

Address: _____

_____

_____

Telephone: _____ Fax: _____

e.mail: _____

Please indicate your language preference: ☐ English
☐ French

Le Groupe de Recherches sur les Transports au Canada
209 - 15 Innovation Blvd.
Saskatoon, Saskatchewan
S7N 2X8
Téléphone: (306) 668-2828 Télécopieur: (306) 668-7603

FACTURE 1996

Membre du GRTC pour l'année 1996 ($60.00) $ _____
TPS ($4.20) $ _____
TOTAL $ _____

Numéro d'enregistrement pour la TPS R126274562. (Si vous êtes exempté de la TPS, veuillez l'indiquer.)

Veuillez libeller votre cheque au nom du Groupe de Recherche sur les Transports au Canada et postez le tout a l'adresse ci-haut. **VEUILLEZ RETOURNEZ UNE COPIE DE CETTE FACTURE AVEC VOTRE PAIEMENT AFIN QUE NOTRE BASE D'INFORMATION CONTIENNE UNE INFORMATION EXACTE SUR VOUS COMME MEMBRE DE L'ASSOCIATION.**

CONDITIONS: paiement effectué dans les 30 jours. Un taux d'intérêt de 2 p.cent sera ajouté sur tout montant non payé.

Veuillez fournir l'information demandée ci-dessous afin de nous permettre d'avoir une information complète sur nos membres. Cette information va apparaitre dans le prochain répertoire du GRTC.

Nom: _____ Titre: _____

Compagnie: _____

Adresse: _____

_____

_____

Téléphone: _____ Télécopieur: _____

Courrier Électronique _____

Veuillez indiquer la langue de correspondance de votre choix.

☐ anglais ☐ français



ISSN# 1183-2770