# Bayesian Linear Regression Performance Model for the Libyan National Road Network without an Initial Database on Its Condition Based on Expert's Knowledge

[1]Abdussalam Heba, Ecole de Technologie Supérieure, 1100 Notre-Dame west, Montreal QC, Canada, abdussalam.heba.1@ens.etsmtl.ca, +1(438)875-4338

[2]Gabriel J. Assaf, Ecole de Technologie Supériere, 1100 Notre-Dame west, Montreal QC, Canada, gabriel.assaf@etsmtl.ca, +1(514)396-8924

## Abstract:

Accurate prediction of rates of road deterioration is important in road management systems, to ensure efficient prioritization and for setting budget levels. Libyan roads (research target) facing increasing damage resulting from the absence of regular maintenance reinforces the need to develop a system to predict the deterioration of roads in order to determine the optimal pavement intervention strategies (OIS) to support the decision makers. In fact, in a PMS, pavement deterioration could be modeled deterministically or probabilistically. This paper proposes a Bayesian linear regression method to develop a performance model with the absence of the historical data depending on experts' knowledge as a prior distribution. As such, Libyan road experts who have worked for long time with Libyan Road and Transportation Agency will be interviewed to assist and support the input data to feed the Bayesian Model. Then, posterior distribution will be computed using a likelihood function depending on few inspections. The expected results will be the pavement future conditions based on experts' knowledge and few onsite inspections.

**Keywords**: Pavement management systems, pavement performance, International roughness index (IRI), Bayesian linear regression.

## 1. Introduction:

A road pavement deteriorates under the combined actions of traffic loading and environment, thus reducing quality of ride. Useful models should be able to quantify the contribution of relevant variables such as strength of pavement materials, traffic, and environmental conditions to pavement deterioration. Pavement Management Systems (PMS) are commonly used to select maintenance strategies that result in lower project life cycle cost (Haas 1994).
Modeling the performance of pavements is an absolutely essential activity of pavement management, and many highway agencies have developed a variety of pavement performance models for use in their pavement management activities, sometimes paying attention to one type of performance or one type of model to the exclusion of others. However, all types of performance are important and all types of models are useful in predicting at least one kind of performance. I will give a brief but comprehensive review of the types of performance, the concepts underlying pavement performance prediction models, the data required as input to them, their uses and their limitations (Shahin 2005).

There are two basic kinds of performance model: deterministic and probabilistic. While the deterministic models predict a single number for the life of a pavement or its level of distress or other measure of its condition, the probabilistic models predict a distribution of such events. Deterministic models include Mechanistic, Empirical-Mechanistic, Polynomial Constrained Least Squares, and S-Shaped Curve models. Probabilistic models include Bayesian and Markov process models (Li 2005).

### 1.1. Bayesian Model:

The basic principle of Bayesian statistics lies in combining prior probabilities (and likelihoods) with experimental outcomes to determine a post-experimental or posterior probability. The posterior distribution expresses what is known about a set of parameters based on both the sample data and prior knowledge (Han, Kaito, and Kobayashi 2014). In frequentist statistics, it is often necessary to rely on large sample approximations by assuming asymptomatic normality. In contrast, Bayesian inferences can be computed exactly, even in highly complex situations. We first give an account of basic uses of Bayes' theorem and of the role and construction of prior densities. We then turn to inference, dealing with analogues of confidence intervals, tests, approaches to model criticism, and model uncertainty (Gongdon 2003).

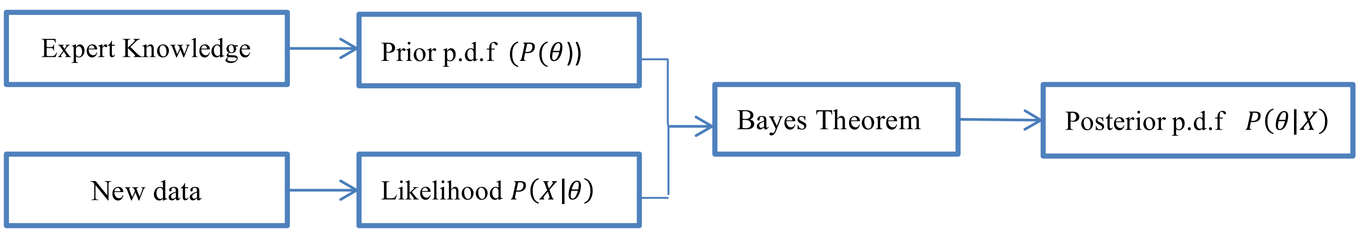


**Figure1. Bayes General Concept**

Using the probability density function, Bayes' model can be expressed as follows:

$$\boldsymbol{P(\theta|X) = \frac{P(X|\theta)P(\theta)}{P(X)} = \frac{P(X|\theta)P(\theta)}{\int P(X|\theta)P(\theta)d\theta}} \quad (1)$$

#### 1.1.1. *Prior knowledge* $\boldsymbol{P(\theta)}$

A fundamental feature of the Bayesian approach to statistics is the use of prior information in addition to the (sample) data. A proper Bayesian analysis will always incorporate genuine prior information, which will help to strengthen inferences about the true value of the parameter and ensure that any relevant information about it is not wasted (Lunn et al. 2000).

#### 1.1.2. *Maximum Likelihood Estimation (MLE)* $\boldsymbol{P(X|\theta)}$:

The maximum likelihood estimation (MLE) approach is one of the most important statistical methodologies for parameter estimation. It is based on the fundamental assumption that the underlying probability distribution of the observations belongs to a family of distributions indexed by unknown parameters (Ralph Haas 1977). The MLE estimator of the unknown parameters is the maximizer of the likelihood function, corresponding to the probability distribution in the family that gives the observations the highest chance of occurrence. The MLE method starts from the joint probability distribution of the n measured values $\boldsymbol{x_1, x_2, \dots, x_n}$. For independent measurements this is given by the product of the individual densities $\boldsymbol{p(x|\theta)}$, which is:

$$\boldsymbol{P(X|\theta) = p(x_1|\theta)p(x_2|\theta) \dots . p(x_n|\theta) = \prod_{n}^{i=1} p(x_i|\theta)} \quad (2)$$

### 1.1.3. *Posterior distribution $\boldsymbol{P(\theta|X)}$:*

It expresses what is known about a set of parameters based on both the sample data and prior information. Bayes theorem works as a mechanism for generating a posterior of any parameter mixing the prior knowledge with the likelihood. The 1st iteration production of the prior knowledge and the MLE will then be divided by $P(X)$ (normalizing factor) to normalize the distribution.

When the posterior distribution $P(\theta|X)$ is in the same family as the prior probability distribution $P(\theta)$, the prior and posterior are then called conjugate distributions. Nonconjugate prior distributions can make interpretations of posterior inferences more difficult.

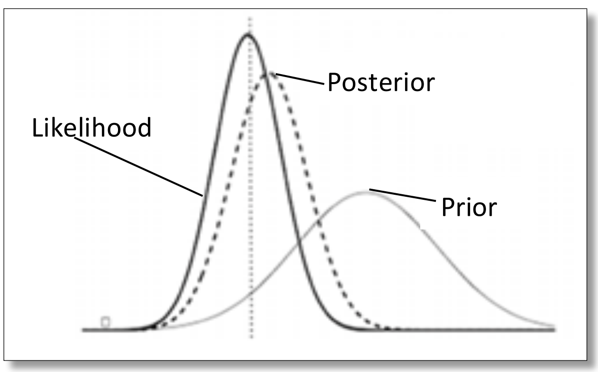


**Figure 2. Prior, likelihood, and posterior using arbitrary data.**

## 1.2. The International Roughness Index (IRI)

The International Roughness Index (IRI) is an international standard for measuring road roughness longitudinally. The index measures pavement roughness in wheel path in terms of the number of metres per kilometre that a laser, mounted in a specialised van, jumps as it is driven along a road. The lower the IRI number at given speed, the smoother the ride felt by road user. Moreover, roughness statistic is suitable for any road surface type and covers all levels of roughness (Ralph Haas 1977). IRI can be treated as a random variable so; it can be described as a probability distribution (Shahin 2005).

The main advantages of the IRI are that it is stable over time and transferable throughout the world. IRI can also be used as a measure of pavement serviceability and it can be directly related to vehicle operating costs (Shahin 2005).

## 2. Methodology:

The technique that will be used to estimate the road network deterioration is by developing a Bayesian Expert-Based probability matrices of deterioration of road classes in Libya. This method depends on combining observed data and expert experience using Bayesian linear regression techniques. Bayesian prediction approach is the process of analyzing statistical models by using prior knowledge and observations (Equation 1) (L. A. Jiménez 2012).

Bayesian linear regression is adding more accuracy to the parameters estimation of the International Roughness Index (IRI) because it recovers the whole range of inferential solutions, rather than a point estimate and a confidence interval as in classical regression (Davison 2008).

The research methodology consists of three major steps which are: experts interviewing in order to set up the prior distribution, road network inspection to estimate MLE and then computing the posterior and predictive distributions for the IRI:

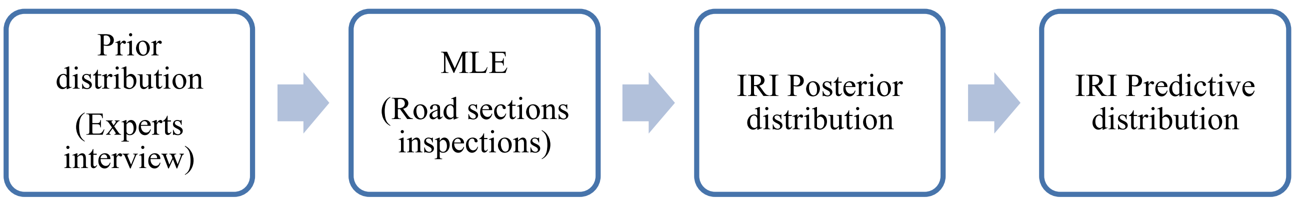


**Figure 3. Research methodology steps.**

### 2.1. Experts interviewing (Prior distribution):

Since loadings and soil conditions are the most important factors that affect damages of most pavement sections, they are often used as independent variables in developed condition prediction equations. In many cases, they combined with age as an independent variable. Since in most circumstances, agencies want to know when in years, the pavement will need intervention, in some models loads and types of soil are used as a factor that affects the rate of condition change as a function of time which is considered the independent variable.
In this research, 3 loading levels, 3 soil conditions and 2 climate zones interacting with each other and producing 18 pavement families. Then, these pavement families are used to develop a Bayesian linear regression prediction models for each family (Tables 1 and 2). Road sections are selected using precise sampling technique to avoid any biased estimations. Initial data will be collected by interviewing Libyan experts some of them are Dr. Mohamed Emmbark, Dr. Mohamed Eshtewi, and Dr. Mohamed Khalifa who have worked for many years on Libyan road network development. Experts interviewing has been conducted using a standardized open-ended interviews technique which is the most structured and include a set protocol of questions and probes. Table 1 shows the distribution of the 18 road conditions families (Pandis 2015).

### 2.2. Pavement condition inspections (MLE data)

As a part of this research, inspecting selected road sections are conducted. Sectors those are inspected are the same sectors which have been investigated during the experts interviewing. The road deterioration is measured by the IRI which is considered a standard for pavement roughness measurements and analysis for a number of agencies (Ralph Haas 1977).

| | | North Zone | | | South Zone | | |
|---|---|---|---|---|---|---|---|
| | | Load Level | | | Load Level | | |
| | | Low | Medium | High | Low | Medium | High |
| Soil Strength | Low | *Dataset1* | *Dataset2* | *Dataset3* | *Dataset10* | *Dataset11* | *Dataset12* |
| | Medium | *Dataset4* | *Dataset5* | *Dataset6* | *Dataset13* | *Dataset14* | *Dataset15* |
| | High | *Dataset7* | *Dataset8* | *Dataset9* | *Dataset16* | *Dataset17* | *Dataset18* |

**Table 1: Road network will be divided into Zones (North and south) interacting with traffic loads and soil conditions during a sequence of years.**

**Table 2. Each dataset consists of road sections condition under same factors.**

| ***Dataset i*** |
|---|
| *Section 1* |
| *Section 2* |
| ⋮ |
| *Section n* |

| ***Zone: North*** | | ***Soil strength: Low*** | | | ***Load level: Low*** |
|---|---|---|---|---|---|
| **Section** | **Year 1** | **Year 2** | **Year 3** | **Year 4** | **Year 5** |
| **1** | 1.5 | 1.9 | 2.5 | 3.3 | 3.9 |
| **2** | 2 | 2.3 | 2.6 | 3.2 | 3.7 |
| **3** | 1.8 | 3.0 | 3.2 | 4.0 | 4.5 |
| **4** | 1.7 | 2.3 | 2.9 | 3.5 | 3.9 |
| **5** | 2.3 | 2.5 | 3.0 | 3.7 | 4.1 |
| **6** | 2.0 | 2.3 | 3.1 | 3.9 | 4.2 |
| **7** | 1.9 | 2.4 | 3.3 | 4.1 | 4.3 |

**Table 3. A sample of a family pavement condition (IRI data).**

### 2.3. RI estimations

Roads deteriorate and drop their ride quality gradually over time. This relationship can be represented using linear regression but practically, road network sections under the same zone, age, load, and soil strength conditions could have a different quality of ride. Therefore, Bayesian linear regression is the appropriate technique where basic Bayesian philosophy is applied.

Moreover, in Bayesian inference, MLE is considered point estimations. However, in Bayesian linear regression, productive probability around each inference of the IRI is probabilistically investigated (L. A. Jiménez 2012).

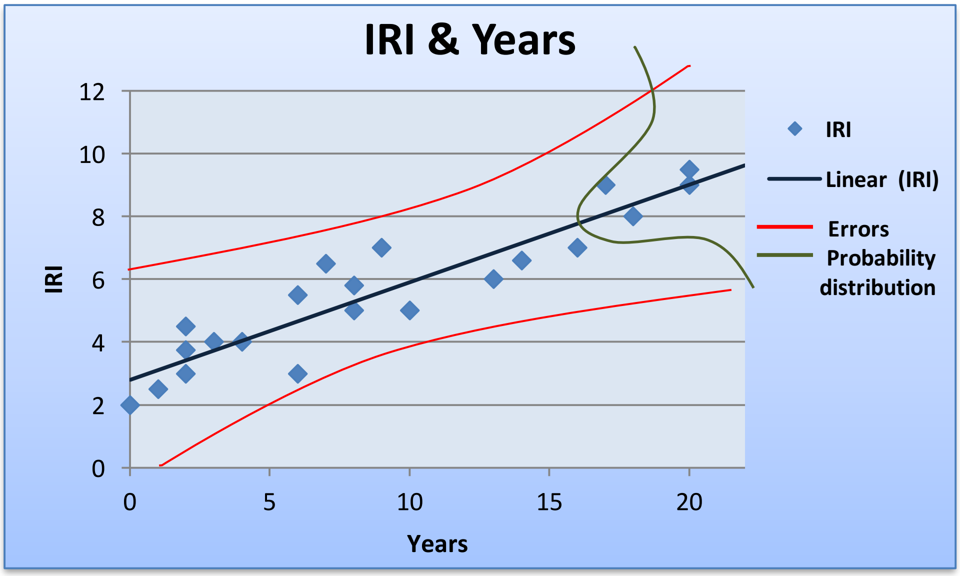


**Figure 4. The differences between linear regression and Bayesian linear regression**

The research data which is required to estimate the IRI has been divided into two main categories, the first category was extracted from the interviewing process of the experts and the second type (MLE data) has been collected from the road sections inspections process to measure the road deterioration using the IRI.

The MLE data is extracted and summarized as pairs of $(t_i, IRI_j)$ *where IRI* represents the road section condition and *t* indicates the time.

$$Data = \left((t_1, IRI_1), \dots \dots, (t_n, IRI_n)\right), 0 \leq t_i \leq 20, 0 \leq IRI_j \leq 14 \quad (3)$$

Then, we model $IRI_j$ to be conditionally independent given the *w* vector which will be the prior distribution.

$$IRI_j \sim N(w^T t_i\,,\ a^{-1}), a > 0 \quad (4)$$

$$w \sim N(0, b^{-1}I), b > 0, w = (w_1, \dots \dots, w_d) \quad (5)$$

*Where*

- $a = \frac{1}{\sigma^2}$ is the precision factor,
- *b is* the covariance matrix,
- *a* and *b* are known and,

- *w* is *a* parameter vector with a Gaussian multivariate density.

### 2.3.1. *The Posterior distribution:*

The next step is to compute the posterior distribution on *w* given data. $t_i$ will be written as $\varphi(t_i) = (\varphi_1(t_1), \dots., \varphi_n(t_1))$ in order to be able to model the nonlinearities of $t_i$ . To compute the posterior we need to calculate the MLE and then the predictive distribution.

#### *Maximum Likelihood Estimation (MLE).*

Given data

$$D = (IRI_1, \dots., IRI_n), IRI_i \in (0{,}14) \quad (6)$$

D represents a sample from the *IRI* statistical population which has been collected from road sections inspections. Then, the MLE is computed using the following formula:

$$P(D|w) \propto \exp\left(-\frac{a}{2}(IRI - Aw)^T(IRI - Aw)\right) \quad (7)$$

Where A is the design matrix and IRI is a value that we are going to predict which is in a column vector form.

$$A = \begin{pmatrix} - & t_1^T & - \\ \vdots & \vdots & \vdots \\ - & t_n^T & - \end{pmatrix}, \quad IRI = (IRI_1, \dots\dots, IRI_n)^T$$

#### *Posterior*

From the classical Bayesian definition, the posterior is proportional with the prior

$$P(w|D) \propto P(D|w)P(w) \quad (8)$$

After that we blog the MLE expression into the posterior which will be as;

$$P(w|D) \propto \exp\left(-\frac{a}{2}(IRI - Aw)^T(IRI - Aw) - \frac{b}{2}w^T w\right) \quad (9)$$

With few calculus steps we can express *w* to be in a form of Gaussian distribution and called the precision matrix:

$$P(w|D) = N(w|\mu, A^{-1})\ Where\ \mu = a\Lambda^{-1}A^T iri\ ;\ \ \Lambda = aA^T A + bI \quad (10)$$

That shows us the Maximum a Posterior (MAP) and MLE estimations of *w* which are:

$$w_{MAP} = (A^T A + \frac{b}{a}I)^{-1}A^T iri \quad (11)$$

$$w_{MLE} = (A^T A)^{-1} A^T iri \quad (12)$$

### ***Predictive distribution***

The predictive distribution is the conditional distribution of unobserved observations (prediction) given the observed data. Our unobserved observation is the experts interviewing data and the observed data, is the collected data from road condition inspections which can be expressed in the following format.

$$P(iri|t, D) = \int P(iri|t, w)(w|t, D)dw \quad (13)$$

$$= \int N(iri|w^T t, a^{-1}) N(w|\mu, A^{-1})dw \quad (14)$$

$$\propto \int exp\,(-\frac{a}{2}(iri - w^T t)^2 exp\,(-\frac{1}{2}(w - \mu)^T \Lambda(w - \mu)dw \quad (15)$$

This formula is then factored and put in a quadratic form in *w* in a formula similar to the following general expression: $\int N(w| \dots \dots) g(iri) dw$ and then, since $g(iri)$ is not depending on *w,* it comes out of the integral and $\int N(w| \dots \dots)dw$ integrates to 1. After several algebraic steps, finalization of the predictive distribution is:

$$P(iri|t, D) = N\left(iri\middle|u, \frac{1}{\lambda}\right) \quad where \quad u = \mu^T t \quad and \quad \frac{1}{\lambda} = \frac{1}{a} + t^T \Lambda^{-1} t \quad (16)$$

Finally, using mathematical expectation and equation (16) in all road sections families, IRI parameters will be estimated depending on:

- *iri* which is the experts interview data,
- *t* is the time corresponding with road condition and,
- *D* is the data collected from road inspections.

## Conclusion

Estimating road roughness in order to measure the pavement performance using Bayesian linear regression technique has many advantages some of them are the ability to include in the statistical model the prior knowledge as well as the existing data (evidences). Secondly, it has the ability to make inferences and predictions by including the complex probability density distributions of prior model results. Moreover, productive probability around each inference of the IRI can be probabilistically investigated. Consequently, based on previously mentioned features, this technique is highly recommended when developing a model to estimate pavement performance with the absence of historical data.

This method is not exclusive to the Libyan network roads, but is applicable in any road network when the circumstances were similar especially in developing countries. After estimating the value of the IRI, the agency will be able to develop appropriate intervention strategy for maintenance or rehabilitation of the road sections and estimate the costs of the chosen strategy.

## References


Davison, A. C. 2008. *Statistical models*. New York: Cambridge University Press.

Gongdon, Peter. 2003. *Applied bayesian Modelling*. West Sussex, England: John Wiley & Sons Ltd.

Haas, Ralph. 1994. *Modern pavement management*. Malabar Florida: Krieger Pub Co

Han, Daeseok, Kiyoyuki Kaito, and Kiyoshi Kobayashi. 2014. "Application of Bayesian estimation method with Markov hazard model to improve deterioration forecasts for infrastructure asset management." *KSCE Journal of Civil Engineering* 18 (7):2107-2119. doi: 10.1007/s12205-012-0070-6.

L. A. Jiménez, Donath Mrawira. 2012. "Bayesian Regression in Pavement

Deterioration Modeling: Revisiting the

AASHO Road Test Rut Depth Model." *Infraestructura Vial* 25:28-35.

Li, Zheng. 2005. "A Probabilistic and Adaptive Approach to Modeling Performance of Pavement Infrastructure." Ph.d, Faculty of the Graduate School, University of Texas at Austin.

Lunn, David, Andrew Thomas, Nicky Best, and David Spiegelhalter. 2000. "WinBUGS - A Bayesian modelling framework: Concepts, structure, and extensibility." *Statistics and Computing* 10 (4):325-337. doi: 10.1023/A:1008929526011.

Pandis, Nikolaos. 2015. "The sampling distribution." *American Journal of Orthodontics and Dentofacial Orthopedics* 147 (4):517-519. doi: http://dx.doi.org/10.1016/j.ajodo.2015.01.009.

Ralph Haas, Road and Transportation Association of Canada. 1977. *Pavement management guide*. Ottawa, ON.

Shahin, M. Y. 2005. *Pavement Management for Airports, Roads and Parking Lots*. New York, USA: Springer Science.