Capturing Variability in Pavement Performance Models from Sufficient Time-Series Predictors: A Case Study of the New Brunswick Road Network

Luis Esteban Amador-Jiménez[1] S.M. ASCE & Donath Mrawira[2]

ABSTRACT

This paper proposes the use of multi-level Bayesian modeling for calibrating mechanistic model parameters from historical data while capturing reliability by estimating a desired confidence interval of the predictions. The model is capable of estimating the parameters from the observed data and expert criteria even in cases of missing data points. This approach allows rapid generation of several deterioration models without the need to partition the data into pavement families. It estimates posterior distributions for model coefficients and predicts values of the response for unobserved levels of the causal factors. A case study from New Brunswick Department of Transportation is used to calibrate a simplified mechanistic pavement roughness progression model based on six-year IRI observations. The model incorporates the effects of pavement structural capacity in terms of deflection basin parameter (AREA) in place of the modified structural number, traffic loading (ESAL) and environmental factors. The results of the model showed that, as expected, chipseal roads have higher as built roughness and deteriorate faster that asphalt roads. Sensitivity analysis of the deterministic (the mean predictions) part of the model showed that in New Brunswick where traffic is relatively low the environment is the most important factor.

Key Words: Performance Model, Multilevel Bayesian Regression, Missed data, Calibration.

[1] Assistant Professor,, Dept. of Building, Civil and Environmental Engineering, University of Concordia Montreal, PQ, Canada, H3G 2W1, Email: amador@encs.concordia.ca

[2]Associate Professor and DC Campbell Chair in Highway Construction and Pavement Research, University of New Brunswick. Fredericton, NB, Canada, E3B 5A3, Email: donath@unb.ca

INTRODUCTION

Performance deterioration modeling and decision making tools (such as mathematical optimization) are at the core of any transportation asset management system. Current performance deterioration modeling lacks measures to explicitly show the associated uncertainty of the predicted response. State of the practice in performance deterioration models is based mostly on deterministic or mechanistic relationships which only give the mean condition at any time for any family of roads. While such predictions are acceptable as a starting point, their use within a transportation asset management systems precludes the capability to analyze works programs while exploring the impact of pavement predictions on the resulting plan, for example.

The exercise of allocating resources in any management system is commonly related to three major questions: what asset to treat, when to treat and which treatment to employ. The concept of operational windows is normally used to narrow the applicability of a given treatment to a candidate section depending on the condition at a given time. Trade-off decisions among competing assets are based on the overall efficiency obtained after selecting the best path of treatments and assets across time. Under such scheme, the performance deterioration modeling is required to provide accurate predictions of the mean expectation along with an expected range of variation (i.e., confidence interval).

Among existing methods for deterioration modeling, perhaps the most adequate are the mechanistic forms which directly relate observable causal factors with measurable responses. The process of adapting road deterioration models to local conditions requires the estimation of model parameters that reflect local conditions of environment, traffic loading, pavement strength, and observed distresses and damages.

OBJECTIVES

(1) To capture uncertainty in pavement performance modeling.

(2) To calibrate mechanistic model parameters from locally observed data

(3) To address practicalities on pavement performance modeling such as; incorporating expert criteria and fully utilizing all available information even if there is missing data.

(4) To generate several deterioration models for various pavement families in one step.

(5) To correct biased performance models by borrowing information across groups.

REVIEW OF PERFORMANCE MODELING

This section reviews historical advancements on performance deterioration modeling, focusing more on identifying deficiencies and drawbacks of classical approaches and then moving towards the identification of Multilevel Bayesian Modeling as the suggested approach for capturing uncertainty while being able to address a range of practicalities on pavement performance modeling.

Selection of Model Formulation

According to Haas et al. (1994) and TAC (1997) there are four basic types of performance deterioration models:

- Purely mechanistic: "the predictive measure is some primary response such as stress, strain or deflection," (TAC 1997).
- Mechanistic – empirical: a dependent variable (response) is related to a "measured structural or functional deterioration" (TAC 1997) through a functional form (C-SHRP 1997).
- Empirical or regression: a dependent variable (response) is related to one or more independent variables "like subgrade strength, axle load applications, pavement layer thickness and properties, environmental factors and their interactions" (Haas et al. 1994).

- Probabilistic: “experience is captured in a formalized or structured way” (Haas et al. 1994) using transition probabilities matrices (TPM) from the Markov Chain approach.

Probabilistic methods such as the Markov Chain produce predictions in a way that shows associated variability. However, its very formulation lacks of robustness because it only mimics the changing dynamics from observed historical trends. It lacks the sensitivity to changes in the causal factors. Another approach is the use of regression. This approach has been criticized because of its limited transferability to regions where environmental conditions and traffic loading differs from the original one. Although this point of view of classical regression does not extends to Bayesian regression since the calibration of the regression coefficients from locally observed data can be used to adapt models to local conditions. This research demonstrates how functional forms of the mechanistic type (that explicitly relate measurable responses to observed factors), can be combined with probabilistic distributions and simulation in order to generate predictive models for dissimilar groups of pavement families, including reliability measures.

The Multilevel Bayesian Regression Model

Bayes theorem is a probabilistic relationship that combines prior knowledge with observed information to produce an adjusted probability of any event. The theorem can be applied to the specific case of regression (TAC 1997). In such cases the theorem (Equation 1) works as the platform for estimating improved expectations of parameters or their predicted responses by combining prior knowledge (expert criteria or results from previous experiments) with observed data of the response and the causal factors (likelihood). The adjusted expression of the event probabilistic distribution is called the posterior. The process is embedded in a probabilistic expression in which the mean is represented by the classical mechanistic equation and the variance is represented by the variability of the model predictions for any specified confidence interval.

Equation 1 shows Bayes theorem in its continuous form composed of three terms: the posterior which is given in terms of the likelihood of the data given a vector of model parameters (*θ)*, *P(data/θ)* times the prior knowledge *P(θ)* over the summation of all joint marginal probabilities. Hong and Prozzi (2006) explained that the denominator (known as the normalization constant) ensures that the sum of the probabilities adds to one (100%). Choosing the right prior has been a matter of debate (Spienhalter et al. 2009, Bishop 2006). In general, the likelihood is given by the available data and the prior from either previous investigations or expert criteria. Priors can be informative or non-informative. Non-informative priors are preferred whenever little is known about the process one is modeling, although in such cases the posterior will tend to look like the likelihood. Informative priors in the form of expert criteria or results from independent research will get mixed with the likelihood and produce an enhanced posterior distribution.

$$P(\theta / Data) = \frac{P(data/\theta)\cdot P(\theta)}{\int P(data/\theta)\cdot P(\theta)\cdot d\theta} \qquad [1]$$

Although Bayesian statistics has been known to be theoretically helpful they remained inapplicable for decades because of the lack of close form solutions to solve the integral in the denominator of Equation 1, until sampling techniques such as the Markov Chain Monte Carlo simulation became available.

As explained by Freitas (1999) sampling can be used as a technique of approaching the true value of complex integrals [e.g., area under a given probabilistic distribution *p(x)*] by generating random values and counting their frequency within the limits of *p(x).*

An alternative to solving complex functions such as the integral on Bayes theorem denominator, is that of sampling. Several techniques for sampling have been tested in the literature including rejection sampling, importance sampling, sampling importance re-sampling, and Markov

Chain Monte Carlo (MCMC). The most comprehensive MCMC is the Metropolis-Hasting; all others are said to be derived from it (Andrieu et al. 2003).

Gamerman and Lopez (2006) explained that Gibbs sampling is a particular case of Metropolis-Hasting for acceptance (A), either 1 or 0. Gibbs sampling is a very attractive method of setting up an MCMC algorithm for getting the joint posterior distribution of all (causal factors) parameters. "The idea behind Gibbs sampling is that we can set up a Markov chain simulation algorithm from the joint posterior distribution by successfully simulating individual parameters from the set of *p* conditional distributions" (Albert 2007). The procedure goes by simulating in turn one value of each individual parameter (called one cycle of Gibbs sampling).

The idea of multilevel regression is that of having sub-populations of individuals sharing certain characteristics among the overall set of observations. This idea is close to that of homogeneous groups or families of pavements (Butt et al. 1989). Classical regression analysis lacks a formal mechanism to consider the generation of sub-models for such groups. It is generally accepted that the treatment of pavements within homogeneous groups produces more reliable predictions as they are closer in behaviour (George et al. 1992). However, the mechanism for creating such groups has not been clearly addressed.

A classical regression model fits observed data to a functional form which in turn can be used to predict non-observed values. Regression models always implicitly include uncertainty due to unobserved variables, which may lead to unreliable outcomes. Traditional regression models treat all data as individuals belonging to the same population and sharing the same characteristics, therefore, they are said to pool all data and obtain values for the regression parameters that produce the best fit (pooled model). The most popular approaches for fitting such models are minimum least square distances and maximum likelihood (Bishop 2006).

The use of a regression model for the entire population, containing fixed parameters (known as complete pooled model) often produces bad fits with high dispersion translated into unreliable predictions. It is also possible to break down the population, creating ad-hoc groups. The formulation of specific regression models per group (no-pooled model) may improve the prediction whenever there are sufficient observations per group, which is not always the case. Rather, the recognition of multiple-levels (or hierarchies) allows the experimenter to improve the calibration of the regression coefficients. It is equivalent to accepting the existence of variable parameters controlling the intercept, shape and rate of deterioration of multiple models.

According to Gelman and Hill (2007) multilevel regression models recognize such nature by considering that the data is structured within different levels. Bayesian multilevel regression models go beyond by returning probabilistic distributions of the parameters instead of assuming fixed values (Bishop 2006). Bayesian multilevel models not only produce a more efficient inference of the regression parameters, but they also enhance the overall prediction by borrowing information across the groups to improve predictions for those clusters with few data (Spiegelhalter et al. 2009), Figure 1 illustrates four cases of groups with different levels of data and how partially pooled regression lines perform better as compared to complete and no-pooled models. Only observations for the no-pooled model are shown.

The Bayesian model proposed here assumes model coefficients to be stochastic variables and estimates a posterior probabilistic distribution from a combination of expert criteria and the observed data. Any Bayesian Regression modeling requires prior information for the stochastic nodes (variables). A functional form is used to represent the mean, $E\{*\}$ accompanied by a variance term, $VAR\{*\}$. Finally, both elements are embedded in a probabilistic distribution (normal in this paper) and a few thousand iterations are performed employing a MCMC simulation, which is divided in two steps: the first part (known as the “burn-in”) consisted of a few thousand iterations conducted until the convergence of the model. The second part consisted of some extra few thousand more iterations and its results produce the posterior distributions of the parameters and the predictive model for pavement deterioration.

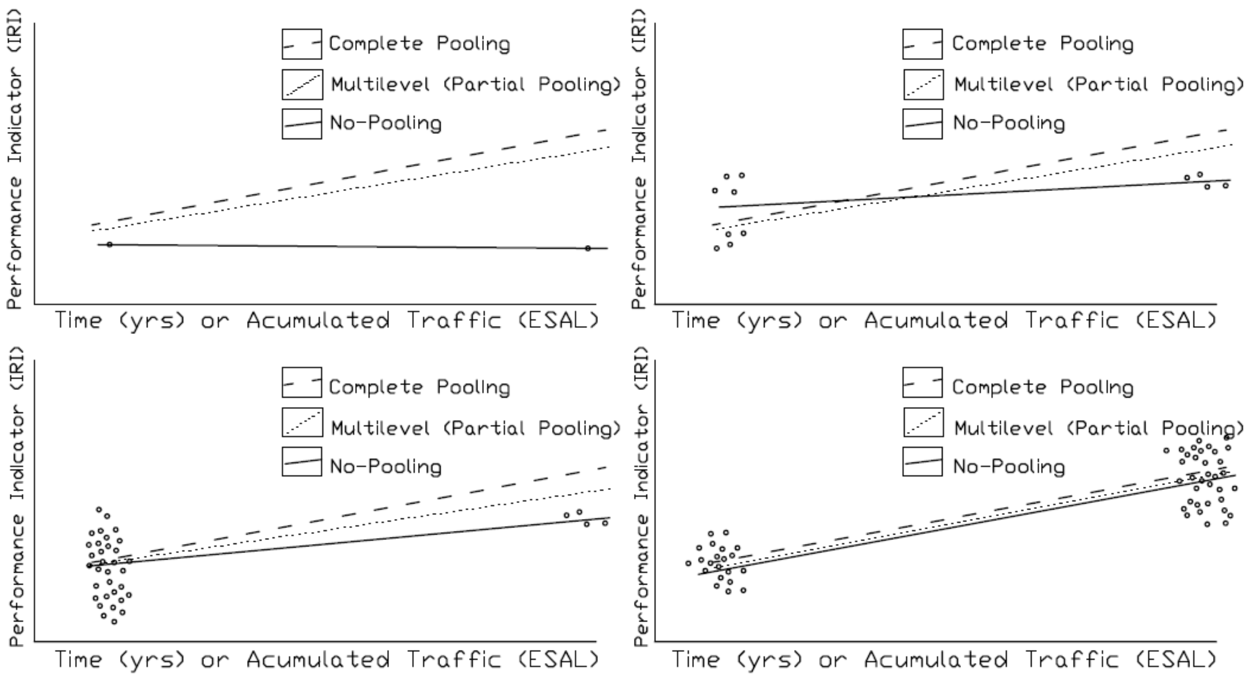


Figure 1. **Comparison of performance on groups with different availability of data**

In summary, the use of Multi-level Bayesian regression modeling has several advantages: (1) it provides a probabilistic estimation of expected responses (condition) at any point in time; (2) it is capable of estimating parameters from observed data (hence the ability to re-calibrate the model to local conditions); (3) it is possible to incorporate expert criteria; (4) it weighs the expert opinions, knowledge and reasonable expectations with observed data to produce a better prediction; and (5) it borrows strength across groups in order to improve predictions on those with few observations.

DATA ANALYSIS

The case Study Data Set

This paper uses a case study to demonstrate the power of the multi-level Bayesian regression method in estimating deterioration model parameters and the associated uncertainty. Time progression of pavement roughness (IRI) of a provincial network is used as an example. The case study is based on the arterial and collector road network of the New Brunswick Department of Transportation (NBDOT). The entire road network comprises 6580 km of asphalt concrete (AC) pavement and 9664 km of chip seal roads grouped into four classes: national, arterial, collector highways, and local numbered and named roads.

The first challenge is always that of integration of all available relevant data from different sources, and collected over different years and systems. The available data for the case study consisted of linearly referenced International Roughness Index (IRI) data, point-data of Dynaflect deflections and traffic volumes in terms of average annual daily traffic of trucks (AADTT). Composition of truck traffic required for computing equivalent single axle loads (*ESAL*) was available in Excel files with values given for some control points along major corridors. Traffic data was transformed into *ESAL* employing the load equivalency factors proposed by RTAC (1986) as exemplified by TAC (1997). Manual assignment of the traffic load (*ESAL*) across the network was required in this case. Environmental data coming from weather stations was incorporated to the closest road segment using a proximity criterion. The international roughness index (IRI) data was available in plain text format for every 160 m segments. Deflection data employed to build an "*AREA*" deflection basin parameter was available every 200 meters. A linearly referenced base map was created with segments of pavement every 160 meters. The final map contained all segments with IRI, *AREA* (strength parameter), *ESAL* (traffic loading) and environment factors assigned.

The deflection basin parameter "*AREA*" (Equation 2) was used instead of the modified structural number since the thickness of the pavements was unknown. The "*AREA*" parameter provides a measure of load-bearing capacity or strength of the pavement. Studies have shown that deflection basin parameters (such as *AREA*) have high correlation with pavement strength, strains in pavement layers and observed deterioration (Kim 1998). The surface deflection is highly dependent on the environment conditions such as temperature and moisture in the base and subgrade. The normalized *AREA* deflection is positively correlated to the strength of the pavement without accounting for the subgrade. It is also known that the deflection at the centre of the load, $D_{0,}$ is negatively correlated to the subgrade bearing capacity (Foxworthy and Darter, 1989).

$$AREA = 6 \cdot \left[1 + 2 \cdot \left(\frac{D_1}{D_0} \right) + 2 \cdot \left(\frac{D_2}{D_0} \right) + \left(\frac{D_3}{D_0} \right) \right] \qquad [2]$$

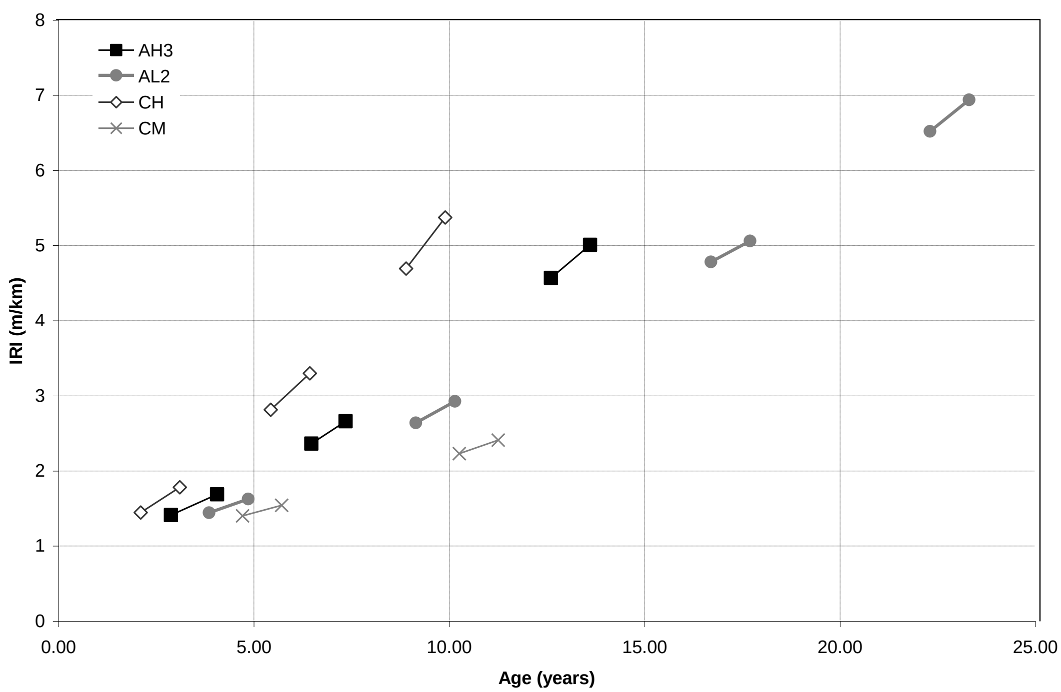
where; D_0, D_1, D_2, D_3 are Dynaflect deflection readings at zero offset, first, second and third geophones, respectively.

Other data such as surface type, functional class and last year of rehabilitation was available in indexed tables. Manual integration of these data with the condition database was used to establish ages for the final selection of road segments for analysis. The final dataset contained 3790 segments of 160 m (1/10 of a mile). Table 1 summarizes the available data and its range.

Table 1. Summary of available information from 1991 to 1996

	ESAL/ year	***IRI (m/km)***	***Area basin***	***Surface***	***Functional class***	***Year***	***Moisture Index***	***Freeze thaw days/year***
From	82,839	0.66	11	Chipseal	Arterial	1967	60	11.58
To	1,213,990	11.68	36	Asphalt	Collector	2006	100	19.2

The selected sections showed deteriorating trends varying in time length according to several factors influencing the deterioration such as: surface type, environmental region and traffic intensity. Therefore the data was broken into 18 datasets responding to different combinations of surface type, environment region and traffic intensity. Figure 2 show initial deterioration trends for some of those combinations of the causal factors. These trends were obtained by fitting observed mean changes on condition from 1994 to 1995 as suggested by Amador and Mrawira (2009).



KEY: *(A) Asphalt, (C) Chipseal road, (H) High traffic intensity, (M) Moderate traffic intensity, (L) Low traffic intensity, (3) (2) and (1) Environmental regions for more than 17 days/year, between 14 and 17 and less than 14 days/year of freeze-thaw cycles, respectively.*

Figure 2. **Initial deterioration models for asphalt and chipseal roads at different traffic intensity and environmental regions**

Dealing with Missing data

Missing data is a common issue in research and practice of engineering. Creating a database of predictors and responses for deterioration modeling does not escape this reality. Entire datasets with missing predictors are normally dropped off the analysis and hence the final model is build upon a reduced number of observations.

One possible solution for not wasting partially available information is by estimating the missing predictors. For the particular case of time-series, missing data can be estimated from the available data points by establishing trends and using them to fill in missing data. However this requires manual processing. This approach has been criticized from theoretical and practical points of view. Theoretically, filling in a missing value with any value (the mean of the trend) misrepresents the spread and reduces the standard error of the overall prediction. From the practical point of view, there are no guarantees that the subset of sections employed to generate the missing value share the same key characteristics of the original group and this can bias the prediction.

In some cases it is possible to infer missing data from the available data. Another alternative for using partial records is by randomly generating the missing predictor from a probabilistic distribution that itself maintains the characteristics of the observed data without biasing it. Such approach provides integrity, not only in the predicted response, but also in the variability. In this respect, the WINBUGS software suite (Spiegelhalter et al. 2009) is able to address the issue of missing values by assuming that they come from a probabilistic distribution. WINBUGS employs this technique to carry out the Bayesian computations for finding the posterior distributions of those parameters of interest. This method is fast and allows the modeller to overcome the challenge of missing data that would otherwise be neglected. Extrapolation of values can be done by mixing both approaches (i.e. estimation from other available data accompanied by random generation) such approach is exemplified in the case study.

The absence of "*AREA*" (i.e. surrogate for pavement strength) would have signified an important reduction of the sample size. Hence a procedure to handle this issue was developed for this particular case. The extrapolation of missing values was done through a semi-random approach. This approach filled in missing data with the adjusted closest observation since a temporal – spatial point of view. The adjustment was randomly generated from a normal distribution of the differences between two consecutive observations in order to account for the deterioration in condition.

The *AREA* deflection basin parameter data was carefully reviewed to identify and remove those sections that may have received structural improvement. If the *AREA* parameter increased from one year to the next, the data was retained; otherwise it was dropped from subsequent analysis. The assumption is that any section that exhibited significant strength improvement, must have received treatment. For example, in the event of missing values during 1995 a value was generated to fit the missing by assuming that the value are normally distributed with a mean of 0.769 and standard deviation of 0.707. The mean and standard deviation values were obtained from the available data before filling in the missing values.

There were sections of the database (17%) for which no values of the "*AREA*" parameter were available and therefore a full random generation of their values was required in WINBUGS. The procedure consisted of assuming that the missing values come from a probabilistic distribution such that the distribution will not affect the final variability of the model. The assumption of a normal distribution is able to align the observed standard deviation and hence will not affect the standard error of the estimated model.

Environmental Effects on Pavement Roughness

Natural Resources Canada (1995) divides the province in several zones corresponding to values of moisture index of 60, 80 and 100. Thornthwaite (1955) established a climate system for classifying regions according to the annual amount of moisture and energy balance. Thornthwaite moisture index is the result of a soil water deficit containing precipitation and evapotranspiration, requiring the knowledge of temperature, rainfall and vegetative coverage for the region under study. According to Thornthwaite (1955) for a region varying from humid to arid the moisture index goes from 100 to -100 (low moisture index corresponds to drier, less humid climates).

Pavements were grouped based on the moisture index and assigned a value of the environmental coefficient *m* as defined by Paterson and Attoh-Okine (1992) and Watanatada et al. (1987). However, values of such coefficient for temperature freezing - humid zones were not available. Hence, an extrapolation from those values presented in Watanatada et al. (1987) was used to find the required values. This extrapolation produced values of the coefficient *m* of 0.07, 0.074 and 0.08 for moisture indices 60, 80 and 100, respectively. These values were used as prior mean in the Bayesian model.

Another environmental criterion was that of freeze-thaw cycles. The freeze heave effect is known to accelerate the deterioration of pavements and ground structures (Khan, 2008). The maximum daily temperature data from Environment Canada from 1971 to 2000 was used to produce a count of days when temperature changed from positive to negative from one day to the next. This was used as a measure of potential of freeze heave effect on pavements. Data from forty nine stations in New Brunswick was spatially merged with the road network employing a proximity criterion. Three environmental regions were created for the province. Those having the most freeze-thaw cycles and the highest moisture index (100%) are expected to suffer a higher degree of deterioration if other causal factors (i.e. same intensity of traffic loading and similar structural strength) remain constant.

Selection of Model Functional Form

The goal of this present work is to demonstrate how multi-level Bayesian regression can be used to estimate a mechanistic deterioration model and characterize the uncertainty. A case study to demonstrate such goal was built on an existing mechanistic formulation. The simplified roughness model developed by Paterson and Attoh-Okine (1992) was adopted (Equation 3). The parametric equation is part of the original formulation of the Highway Design and Management (HDM-III) Model developed by Watanatada et al. (1987). It relates the change in roughness to three causal factors namely, the environment, the surface defects and the structural deformation. Traffic loading and the time since last rehabilitation are also causal factors.

$$RI_t = e^{mt}[IRI_0 + a(1 + SNC)^{-5} \cdot NE_t] \qquad [3]$$

where, RI_t = roughness at pavement aged t [international roughness index, in m/km]; IRI_0 = initial roughness [m/km, IRI]; NE_t = cumulative ESALs at age t (million *ESAL*/lane); t = pavement age since last rehabilitation or construction (years); m = environmental coefficient (Watanatada et al. 1987); SNC = structural number modified for subgrade strength; IRI_0 and a coefficients were obtained from fitting the model to observed data.

The term e^{mt} introduces the deterioration of the pavement due to environmental factors. It was recommended to fix the power coefficient for the pavement strength (Paterson and Attoh-Okine 1992).

For the case study of the New Brunswick network, available data consisted of pavement deflections, traffic loading and environmental data (in the form of moisture index and freeze-thaw cycles).

Information on layer thicknesses was unavailable and therefore it was impossible to compute modified structural numbers. It was argued that the modified structural number reflects the ability of the pavement to withstand loading, hence deflections readings (either Dynaflec or FWD) are correlated to the structural number. The modified structural number coefficient (*SNC*) was replaced by the "*AREA*" deflection basin parameter. This change in model specification (by replacing *SNC* by *AREA*) is expected to affect the alpha coefficient in the model.

The selection of the formulation for relating roughness with available causal factors was also limited by availability of surface distress data (Paterson & Attoh-Okine 1992) and other damage indicators. Due to the absence of such data the simpler relationship (Equation 3) was adapted for this case study. However, that functional form was changed since initial IRI data was not available. The initial IRI was assumed to be a stochastic variable, β to be estimated by the model (Equation 4).

$$RI_t = e^{m \cdot age}\left[\beta + \alpha\left(AREA\right)^{-5} \cdot NE_t\right] \qquad [4]$$

The final calibration of the coefficients was left to be estimated from the observed causal factors and levels of roughness (m/km of IRI). Precise information from the treatment tracking system in NBDOT allowed the incorporation of age since last rehabilitation or reconstruction; hence the effect of the term $e^{m \cdot age}$ was fully taken into consideration.

Equation 4 was used as the mean of the stochastic response IRI for predicting the performance progression across time. This equation was embedded in a normal distribution for a multilevel Bayesian regression model in order to obtain probabilistic distributions for each coefficient. The multilevel Bayesian model carries computations across groups and return probabilistic distributions of every parameter.

Key Analysis Steps for a Multilevel Bayesian Model

A Bayesian model must contain (1) mean of the response; (2) precision; (3) estimate of priors; (4) probabilistic distributions of the parameters and some initial guess or starting point for their values; and (5) the observed data. In this case, another term was required: (6) probabilistic distribution for the missing "*AREA*" values. The assumption of probabilistic distributions for every parameter per class was also required to satisfy the multilevel modeling.

Two separate multilevel models were analyzed: (a) one for estimating the probabilistic performance deterioration based on environmental regions and, (b) another one based on surface type. Although models for other condition indicators are possible, IRI was chosen since treatment allocation decisions commonly rely on surface condition. Previous studies have also correlated IRI with riding comfort index (RCI) (TAC 1997), and with pavement condition index (PCI) (Mrawira et al. 2008).

IRI as presented in Equation 3 (or 4) disregards the levels (or structures –nested or not) in the data pertaining to traffic intensity, environmental zones or surface type. Therefore multi-level models considering environmental regions and surface type were developed for IRI as a function of several factors: accumulated traffic (million of *ESAL*), pavement age (years), environmental coefficient *m*, pavement strength area, and pavement age since last rehabilitation. The model presented by Equation 4 was used as the expectation of the predicted response IRI. Effects for the multilevel structure were studied to determine deterioration models per environment zones and surface type.

The model functional form contains 2 parameters with clear interpretation: β (beta) represents the initial IRI value and α (alpha) the rate of deterioration on the main component given by traffic intensity and pavement strength. Both were introduced as stochastic variables of normal distributions with priors defined in terms of mean and precision. The priors were selected as: β (beta) with a normal distribution with mean of 1.4 and precision of 16, which theoretically allow it to fluctuate from 1.25 to 1.65 and α (alpha) with a normal distribution with mean of 168 and precision of 0.0001 which allow it to fluctuate from 168 to 368. However WinBUGS works in such a way that it may cross those “boundaries” depending on the evidence (observations of casual factors), providing a very reliable mechanism for calibrating mechanistic expressions to local factors, even if the “initial priors” of the modeler are biased.

RESULTS AND DISCUSSION

Results of the IRI probabilistic performance model for New Brunswick are divided in three: (1) multilevel performance prediction for environment, (2) two-level performance prediction model per surface type and (3) nested environment into surface type.

Initial opinion (prior) may suggest that environment do not play a significant role – at least for asphalt roads with high traffic intensity. The initial deterioration trend fitting (Figure 3) based on observed mean changes on IRI condition from 1994 to 1995 gave a misleading model.

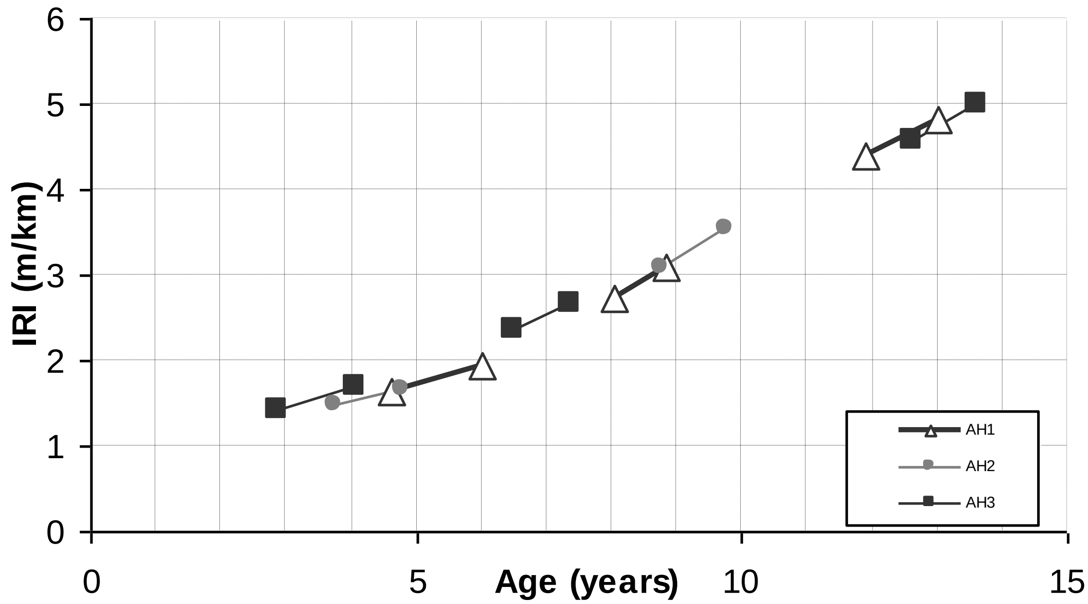


Figure 3. **Initial deterioration model for asphalt roads (A) with high (H) traffic: no significant differences found among the environmental groups (1, 2 and 3).**

The result of carrying out a multilevel Bayesian regression produced a more reliable deterioration model.

Posterior values for the stochastic parameters show that poor initial priors for the as-built IRI near 1.4 were wrong. Even with imprecise estimates of prior values of IRI, the model corrected them to values of 0.9 for the environment region 3 (higher freeze-thaw cycles per year), and 1.19 for the environment region 1 (lowest freeze-thaw cycles per year). The model showed closed values of alpha irrespective of the environmental region (Table 2).

Table 2. **Posterior Probabilistic distributions of deterioration rate (alpha) and initial value of IRI (Beta)**

Node	Mean	Std. Dev.	MC error	2.5%	Median	97.5%
Alpha[1]	264.6	100.9	2.219	67.43	268.2	455.0

Alpha[2]	253.0	101.3	4.148	53.9	254.2	450.6
Alpha[3]	267.9	101.6	2.258	71.37	265.1	465.7
Beta[1]	1.190	0.01849	4.059E-4	1.155	1.190	1.226
Beta[2]	1.073	0.00741	1.522E-4	1.059	1.073	1.087
Beta[3]	0.900	0.01529	3.907E-4	0.870	0.900	0.929

Results from the two-level Bayesian model for each surface type show that chipseal roads (index = 2) not only deteriorate faster but have an initial as-built IRI higher than asphalt roads. Table 3 presents a summary of those results.

Table 3. Summary Statistics for Parameters α and β

Parameter	Mean	Std. Dev.	MCMC error	2.50%	Median	97.50%
Alpha[1]	262.7	99.34	1.575	67.89	262.3	459.1
Alpha[2]	266.1	99.84	0.848	71.22	265.1	461.8
Beta[1]	1.064	0.00969	8.09E-05	1.045	1.064	1.083
Beta[2]	1.170	0.05002	4.03E-04	1.072	1.170	1.268

KEY: *(1) Asphalt; (2) Chipseal road; Std. Dev.= standard deviation.*

A three-level Bayesian model was developed to account for both: environmental effects and surface type. Results from this model showed that it is not possible to distinguish differences in the deterioration rate once environmental zones are nested into surface type (Table 4). After alpha reaches a value close to 266 disregarding of the environment - surface type group. This drawback of the model can be explained by the fact that there were very few points for chipseal roads and the majority of them were for the environmental zone 2. Hence the multilevel model borrowed information from the asphalt model (as explained in Figure 1). In terms of as-built IRI quality, it is possible to conclude that one can observe higher initial deterioration on chipseal roads.

Table 4. Statistics for Posterior Distributions of Parameters α and β

Parameter	Mean	Std. Dev.	MCMC Error	2.50%	Median	97.50%
Alpha[1,1]	268.90	99.83	0.995	68.69	268.90	463.60
Alpha[1,2]	251.30	103.90	1.898	40.98	254.70	454.20

Alpha[1,3]	265.80	101.50	0.993	66.66	264.90	467.00
Alpha[2,1]	265.90	99.55	0.913	71.59	265.50	461.60
Alpha[2,2]	265.90	100.50	1.028	67.64	265.10	464.10
Alpha[2,3]	266.30	99.87	0.961	69.33	266.20	459.80
Beta[1,1]	1.194	0.02034	2.03E-04	1.155	1.194	1.234
Beta[1,2]	1.070	0.00815	8.40E-05	1.054	1.07	1.086
Beta[1,3]	0.885	0.01682	1.70E-04	0.852	0.885	0.917
Beta[2,1]	1.333	0.16150	0.001543	1.018	1.332	1.650
Beta[2,2]	1.155	0.03694	4.00E-04	1.084	1.155	1.228
Beta[2,3]	0.952	0.09285	9.56E-04	0.770	0.951	1.134

KEY INDEXING [a,b]: a=surface type (1=asphalt, 2=chipseal), b=environment zone (3,2,1 = high, moderate and low variation in freeze-thaw cycles)

SENSITIVITY ANALYSIS OF MODEL PARAMETERS

A numerical experiment was conducted in order to determine the influence of the model predictors in the response. Ten thousand set of inputs were randomly generated for predictors in order to cover the entire space of all possible combinations of practical levels of the casual factors. The mechanistic expression from Equation 4 was transformed to normalize the casual factors on the response (IRI). Practical levels for four model parameters were established based on observed values on the dataset. The coefficient *m* ranged from 0.07 to 0.08, the parameter age was observed to range from 1 to 26 years, the accumulated number of *ESAL* in millions (expressed as the variable NEM) range from 0.13 to 17, and the *AREA* basin from 23 to 36. Figure 4 show a scatter plot of the one of the six possible pairs of predictors using a random sample of only 1000 set of inputs, the rest of scatter plots were very similar to this one. As shown in the figure the experimental space was adequately covered.

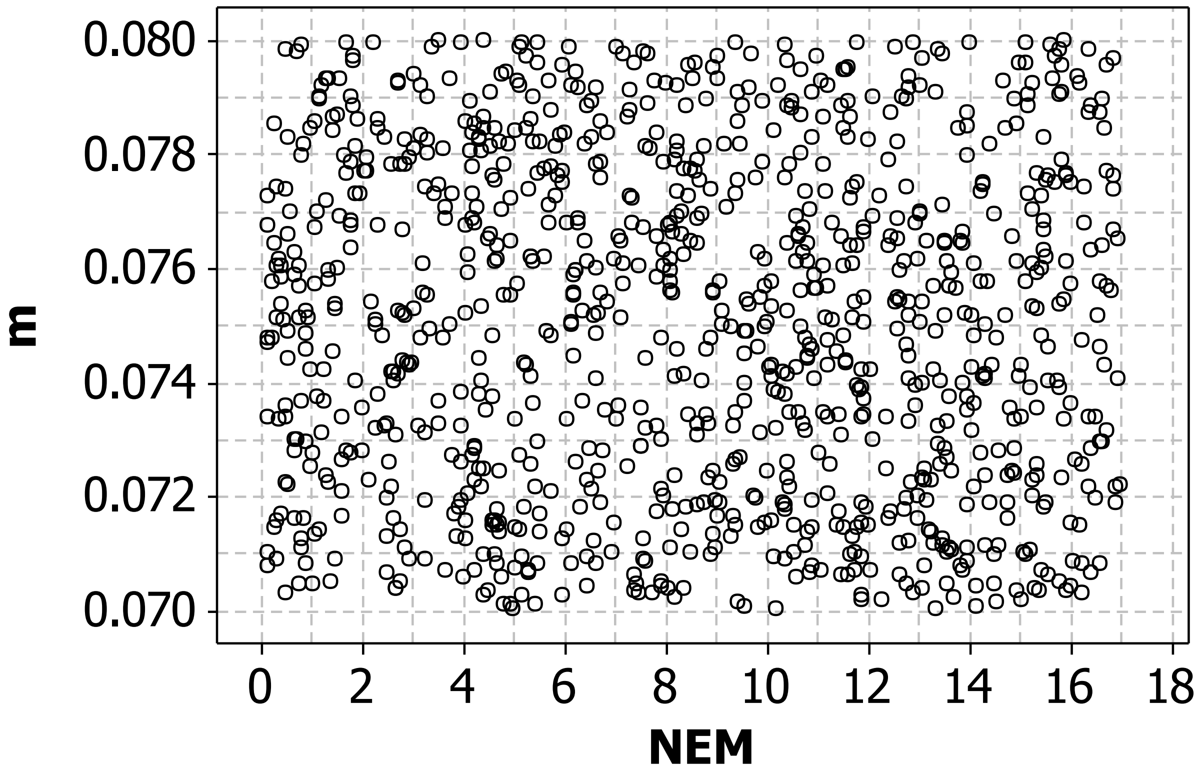


Figure 4. **Sample scatter plot for one pairs of predictors.**

The input sets of predictors were used to compute the mean (deterministic) values of the response (IRI). The generated data set (predictors and response) was analysed by a first order (linear) regression model of the form presented below as suggested by Mrawira et al. (1999):

$$Y = \beta_1 \frac{x_1}{b_1 - a_1} + \beta_2 \frac{x_2}{b_2 - a_2} + \cdots + \beta_4 \frac{x_4}{b_4 - a_4} \qquad [5]$$

where $Y = IRI$ (i.e., response variable of interest); $x_j = j^{th}$ input variable with range $[a_j, b_j]$ for the four causal factors above mentioned and, $\beta_1, \ldots, \beta_4$ are coefficients to be estimated by the least square approach.

According to Mrawira et al. (1999) the coefficients β_j provide an estimate of factor sensitivity for x_j, because they reflect the impact on IRI of changing every causal factor from its minimum to its maximum value. Table 5 summarizes the estimated regression coefficients sorted by magnitude. All coefficients β_j are larger enough to be significant at a 1% significance level (*p* value less than 0.01), but the one with the largest impact on the response is the environmental coefficient *m*. The age of the road, the area and, the accumulated traffic –all– showed a similar order of magnitude in the sensitivity of the response.

Table 5. Fitted Normalized Coefficients of the Predictors of the IRI Model

Input Factor	Estimated Coefficient β	Standard Error	*t*-statistic	*p* value
m	38588	2546	15.14	0.000
Age	236.46	3.291	71.85	0.000
NEM	114.62	4.912	23.34	0.000
AREA	-193.75	6.144	-31.53	0.000

The largest sensitivity of the model to the factor *m* can be explained in the context of the New Brunswick road network as the effect due to the environment (Table 3). Low volume traffic roads in New Brunswick seem to deteriorate due to the environment, because of the freeze-thaw cycles. This hypothesis is also supported by the preliminary results from initial deterioration models of Figures 2 and 3 where environment seems to have a lesser contribution on deterioration for the high traffic intensity whereas the impact increase on roads with lower traffic. The intensity of traffic loading in New Brunswick ranges from 100,000 to 1.0 million *ESAL / year*. In comparison to other regions for example, route 20 in Quebec, the traffic loading can be in excess of 5 to 10 million ESALs / year.

PRACTICAL IMPLEMENTATION OF THE PERFORMANCE MODEL

Results from WINBUGS such as those presented in Table 2 are useful for generating two types of deterioration models: (1) a deterministic model for the mean response accompanied by variability boundaries for the 95% confidence interval and (2) a transition probability matrix (TPM).

For a deterministic model one need to select the mean along with the 2.5% and 97.5% values of every coefficient in the mechanistic equation and produce a plot of performance deterioration with three lines representing the mean along with the variability envelop. For pavement families stated above, the boundaries are obtained for the extreme cases of predictors affecting the performance, for each pavement family and range of traffic loading. Figure 5 shows an example of a deterministic deterioration model for asphalt roads in environmental zone 3 with annual traffic load of 0.5 – 1.0 million *ESAL* per year.

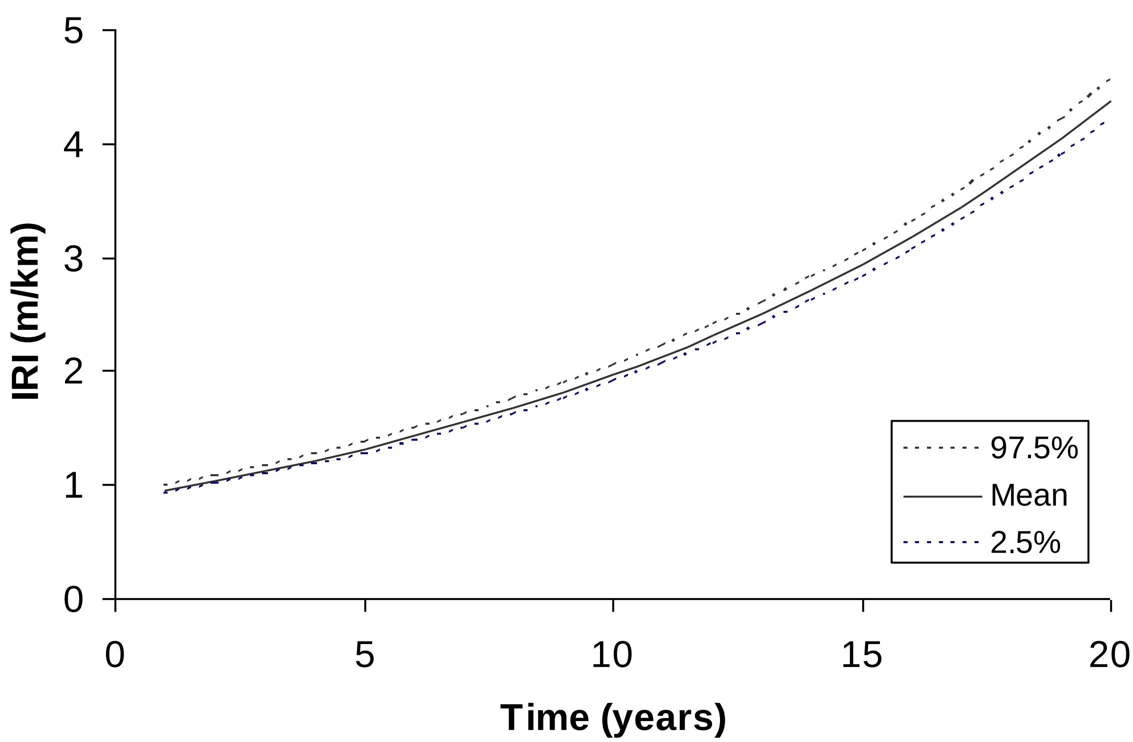


Figure 5. **Deterministic deterioration model with variability for 95% confidence interval**

To obtain a transition probabilities matrix one need to generate a probabilistic distribution of the response every year considering how the variability of the coefficients impact the variation of the expected response. A normal distribution of the mean of IRI and standard deviation was used. This distribution was discretized by a histogram of frequencies of the response (IRI). The frequency of every bar (interval) in the histogram was normalized to obtain percentages. Finally the set of normalized values per year were grouped and their percentages assigned to cells of the TPM. This concept is illustrated in Figure 6. Every row in the TPM comes from a group range of IRI per year. Table 6 show the TPM for the same pavement family during the first 10 years.

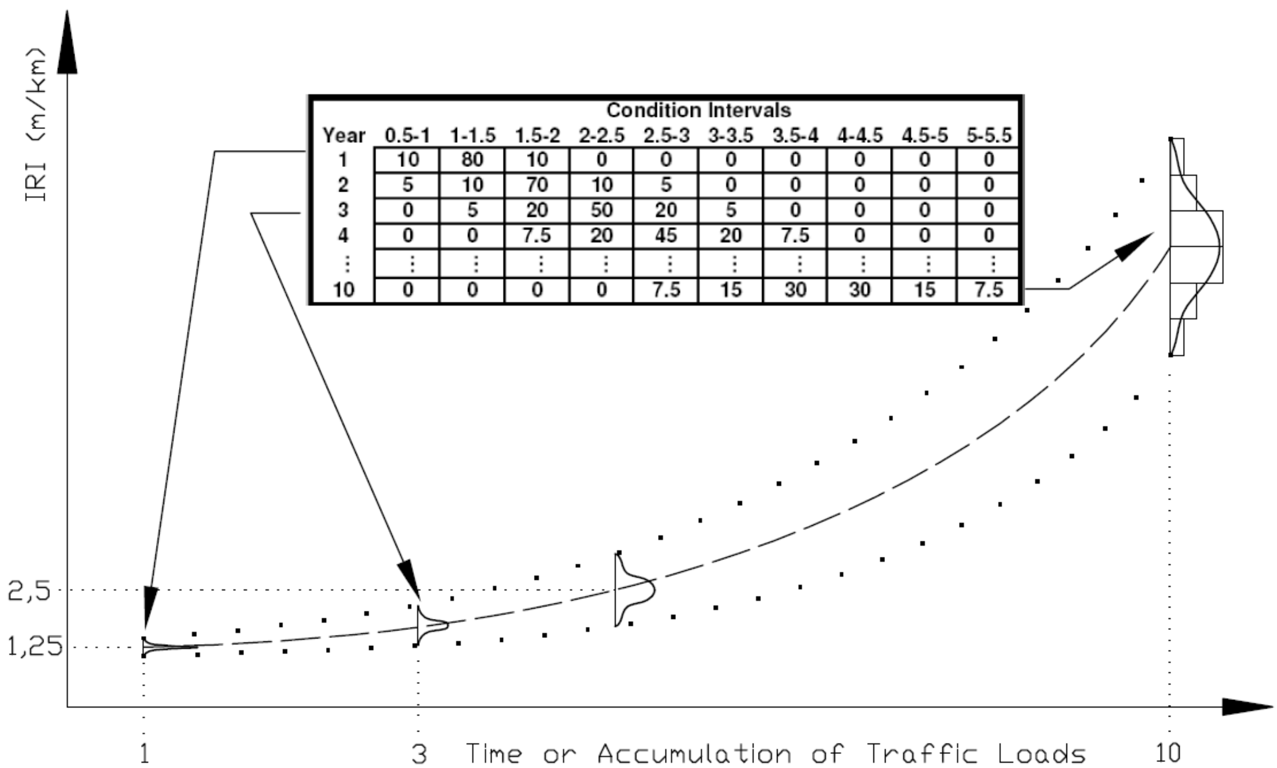


Figure 6. **Procedure to express a deterministic performance model in form of a TPM.**

Table 6. Transition Probabilities Matrix for the Deterministic Model of Figure 9

Year	Condition Intervals											
	0.9 - 1	1-1.1	1.1-1.2	1.2-1.3	1.3-1.4	1.4-1.5	1.5-1.6	1.6-1.7	1.7-1.8	1.8-1.9	1.9-2	2-2.1
1	**100**	0	0	0	0	0	0	0	0	0	0	0
2	**2**	**98**	0	0	0	0	0	0	0	0	0	0
3	0	**6**	**93**	**1**	0	0	0	0	0	0	0	0
4	0	0	**16**	**84**	0	0	0	0	0	0	0	0
5	0	0	0	**21**	**79**	0	0	0	0	0	0	0
6	0	0	0	0	**11.4**	**88.3**	**0.3**	0	0	0	0	0
7	0	0	0	0	0	**5**	**90**	**5**	0	0	0	0
8	0	0	0	0	0	0	**0.4**	**74.4**	**25.2**	0	0	0
9	0	0	0	0	0	0	0	**0**	**28**	**71**	**1**	**0**
10	0	0	0	0	0	0	0	0	0	**2.5**	**77.5**	**20**

It should be noted that in a multilevel Bayesian model one has coefficients for the model depending on how many pavement families or homogeneous groups were created. For example, there are six deterioration models by environmental region and type of surface from the three-level Bayesian model whose results are presented on Table 4.

CONCLUSIONS AND RECOMMENDATIONS

Multi-level Bayesian modeling has demonstrated capable of calibrating mechanistic model parameters from local observed data while capturing reliability by estimating a desired confidence interval of the predictions. The proposed modeling technique is capable of incorporating expert criteria in cases of insufficient data.

Multi-level Bayesian proposed a more logical treatment of the data because it avoids the traditional creation of pavement families and the dilemma of justifying the criteria for partitioning the dataset, preventing biased results. However, multi-level Bayesian modeling poses a limitation in terms of physical interpretability of calibrated coefficients from groups from few observations.

Using a mixed approach for the estimation of the pavement strength parameter "area" proved to address limitations of traditional methods producing randomize values capable of preserving the integrity of the original dataset (i.e. mean, skewness and standard deviation) to avoid biased results.

Pavement strength originally intended to be represented by the modified structural number can be replaced by the deflection basin "*AREA*" parameter, which facilitates the implementation of the proposed model. Environment plays a significant role in the deterioration of roads with lower volumes of traffic. This preponderance vanishes when traffic intensity increases and become more relevant.

It was observed in the three-level Bayesian model that chipseal pavements for environment zones 1 and 3 were scarce in observations and hence their results look like those of the environmental zone 2. Meanwhile it was possible to affirm that chipseal roads deteriorate faster and have higher initial as-built IRI; however, the model could not distinguish in terms of deterioration rate for chipseal roads at different environmental zones.

It was possible to confirm that the influence of priors in posterior distributions of the parameters weakens as the amount of available observation increases. This was the case with the prior assumption that the average as-built IRI is about 1.4 irrespective of environmental zone. After running the multilevel model, the corrected posteriors showed that initial IRI values actually ranged from 0.9 to 1.19 depending on the environmental region.

Practical implementation of performance modeling based on the results of a multi-level Bayesian modeling can be done by: (1) a modified classical deterministic approach or (2) a Markov Chain. The modification in the deterministic prediction consisted of the incorporation of measures of variability in the predicted response. Markov chain has the additional advantage of explicitly showing the probability of occurrence of every level of the response.

Estimated values of the coefficient alpha are particular to the modified model presented in this paper reflecting the replacement of the modified structural number with the *AREA* parameter.

REFERENCES

Albert, J. 2007. Bayesian Computation with R. Springer-Verlag, New York, N.Y.

Amador, L., Mrawira, D. 2009. Roads Performance Modeling and Management System from Two Condition Data Points: Case Study of Costa Rica. Journal of Transportation Engineering, ASCE, December 2009: pp999-1007

Andrieu, C., Freitas, N., Doucet, A. and Jordan, M. (2002). An Introduction to MCMC for Machine Learning (ML) Volume 50 (1-2): pp5-43

Bishop, C.M. (2006). Pattern Recognition and Machine Learning. Spriensger Science + Business Media, LLC. New York, NY.

Butt, A., Shahin, M.Y., Feighan, K.J. and Carpenter, S.H. (1997). Pavement Performance Prediction Model Using the Markov Process. In *Transportation Research Record: Journal of the Transportation Research Board, No.* 1123, TRB, National Research Council, Washington, D.C., pp. 12- 19.

Canadian Strategic Highway Research Program (C-SHRP). (1997). Transportation Association of Canada. C-SHRP Bayesian Modelling: A User's Guide. Canadian Strategic Highway Research Program, Ottawa.

Foxworthy, P.T. and Darter, M.I. (1989). "ILLI-SLAB and FWD Deflection Basin for Characterization of Rigid Pavements". Nondestructive Testing of Pavements and Backcalculation of Moduli, ASTM STP 1026, A.J. Bush III and G.Y. Baladi, Editors, American Society for Testing Materials, Philadelphia. Pp. 368-386.

Freitas, N. (1999). Bayesian Methods for Neural Networks. Trinity College, University of Cambridge. PhD Thesis [online]. Available from http://www.cs.ubc.ca/~nando/publications.html [cited June 3, 2009]

Gamerman, D., Lopes H. (2006). Markov Chain Monte Carlo; Stochastic Simulation for Bayesian Inference. 2nd Edition Chapman and Hall/CRC, Taylor Francis Group. Boca Raton, FL.

Gelman, A., Hill, J. (2007). Data Analysis Using Regression and Multilevel/Hierarchical Models. 1st Edition, 5th reprinting. Cambridge University Press. New York, NY.

George, K.P., Rajagopal, A.S. and Lim, L.K. (1989). Models for Predicting Pavement Deterioration. In Transportation Research Record: Journal of the Transportation Research Board, No. 1215, TRB, National Research Council, Washington, D.C., 1989, pp. 1 – 7.

Haas, R., Hudson, W.R., Zaniewski, J. (1994). "Modern Pavement Management". Krieger Publishing Company. Malabar, Florida.

Hong, F., Prozzi, J.,A. (2006). Estimation of Pavement Performance Deterioration Using Bayesian Approach. Journal of Infrastructure Systems, 12: 77- 86

Khan, A.A., (2008). Feasibility of using lightweight aggregate to mitigate frost damage in flexible pavements. PhD. Thesis. University of New Brunswick. Canada

Kim, Y.R. (1998). Assessing Pavement Layer Condition Using Deflection Data. National Cooperative Highway Research Program Project 10-48, Transportation Research Board, Washington D.C.

Mrawira, D., MacNaughton, J., and Amador, L. (2008). *"Pushing the Frontier of Asset Management: using trade-off analysis across asset-types to support agency's budget exercise".* Proceedings of the 7th International Conference on Managing Public Assets. Calgary Alberta.

Mrawira, D., Welch, W.J., Schonlau, M., and Haas, R. (1999). Sensitivity Analysis of Computer Models: World Bank HDM-III Model. Journal of Transportation Engineering. Vol. 125, No.5, September/October 1999. pp 421 – 428.

Natural Resources Canada. (1995). The National Atlas of Canada 1978-1995: Thornthwaites Classification Moisture Regions [online]. Available from http://atlas.nrcan.gc.ca/site/english/maps/archives/5thedition/environment/climate/mcr4096 [cited June 23, 2009]

Paterson, W.D.O., and Attoh-Okine, B. (1992). Summary of Models of Pavement Road Deterioration Based on HDM-III. In *Transportation Research Record: Journal of the Transportation Research Board*, *No.* 1344, TRB, National Research Council, Washington, D.C., 1992, pp. 99 – 105.

Road Transportation Association of Canada. (1986). “Vehicle Weights and Dimensions Study”, Technical Committee Steering Report, RTAC, Ottawa, December 1986.

Spiegelhalter D.J., Thomas, A., Best, N.G.,Gilks, W., and Lunn, D. (2002). BUGS: Bayesian inference using Gibbs sampling. MRC Biostatistics Unit, Cambridge, England [online]. Available from www.mrc-bsu.cam.ac.uk/bugs/ [cited February 25 2009]

Transportation Association of Canada (TAC). (1997). “Pavement Design and Management Guide”, Transportation Association of Canada, Ottawa, Canada.

Watanatada, T., Harral, C.G., Paterson, W.D.O., Dhareshwar, A.M., Bhandari, A., and Tsunokawa, K. (1987). The Highway Design and Maintenance Standards Model. Volume 1: Description of the HDM-III Model. The International Bank for Reconstruction / The World Bank. The John Hopkins University Press, Baltimore, Maryland, U.S.A.