***"This is an Accepted Manuscript of an article published by Taylor & Francis in International Journal of Pavement Engineering on 05 August 2019, available online: http://www.tandfonline.com/doi/10.1080/10298436.2019.1646912"***

# Development of Network-Level Pavement Deterioration Curves using the Linear Empirical Bayes Approach

Antonio Pantuso, Ph.D., [a]*, Gerardo W. Flintsch, Ph.D., P.E.[b,c], Samer W. Katicha, Ph.D.[b], and Giuseppe Loprencipe, Ph.D.[a]

*[a] Department of Civil, Constructional and Environmental Engineering, DICEA, Università degli Studi di Roma "La Sapienza", Roma, Italy; [b] Center for Sustainable Transportation Infrastructure, Virginia Tech Transportation Institute, Blacksburg, VA, USA; [c] The Charles Via, Jr. Department of Civil and Environmental Engineering, Virginia Polytechnic Institute and State University, Blacksburg, VA, USA*

*corresponding author, Department of Civil, Constructional and Environmental Engineering, DICEA, Università degli Studi di Roma "La Sapienza" via Eudossiana, 18, Rome 00184, Italy, Phone: +39 0644585124, Fax: +39 0644585121, Email: antonio.pantuso@uniroma1.it

## Abstract

Modelling the pavement deterioration process is essential for a successful pavement management system (PMS). The pavement deterioration process is highly influenced by uncertainties related to data acquisition and condition assessment. This paper presents a novel approach for predicting a pavement deterioration index. The model builds on a negative binomial (NB) regression used to predict pavement deterioration as a function of the pavement age. Network-level pavement condition models were developed for interstate, primary, and secondary pavement road families and were compared with traditional non-linear regression models. The linear empirical Bayesian (LEB) approach was then used to improve the predictions by combining the deterioration estimated by the fitted model and the observed/measured condition recorded in the PMS. The proposed approach can improve the mean square error prediction of the next-year pavement condition by 33%, 36% and 41% for Interstate, Primary, and Secondary roads, respectively, compared with the measured pavement condition without further modelling of the pavement deterioration.

Keywords: pavement management system; negative binomial; linear empirical Bayes; pavement deterioration models.

## Introduction

One of the primary objectives of a road agency is to maintain and manage the entire network at a high level of service. To achieve this goal, managers need accurate pavement deterioration prediction models. Road agencies collect a large amount of pavement condition data from inspections performed on their network. Because of uncertainties in the data collection methods, interpretation of distresses, and the inherent variability of individual sections, the pavement's recorded condition can have a high variance. Deterioration models can provide a good representation of the overall condition of the network, but are bad at representing the performance of individual pavement sections. This paper proposes a linear empirical Bayesian (LEB) approach to combine the pavement condition estimated by network-level pavement deterioration models for *families of pavements*, with the most recent observed/measured condition recorded from the inspection of the network. This approach results in a better estimate of the pavement condition compared to the family deterioration model or the measured condition. Compared to the measured condition, that is, with not using a pavement deterioration technique, the LEB approach is estimated to reduce the mean square error (MSE) of predicting the future condition (next year's condition) by as much as 33%, 36%, and 41% for the Interstate, Primary, and Secondary network families, respectively. A more accurate estimate of the future condition addresses the critical need of predicting the future pavement condition to support network-level decision-making.

## Objective

The objective of this paper is to present a novel deterioration modelling approach that is illustrated by predicting the pavement Critical Condition Index (CCI) used by the Virginia Department of Transportation (VDOT). The proposed model builds on the negative binomial (NB) regression that has been used to model pavement deterioration as a function of pavement age (Katicha et al.

2016a). The pavement condition is calculated using the LEB approach, which combines the deterioration estimated by the fitted model and the observed/measured condition recorded from the inspections for each section.

The proposed technique is also compared with the default pavement deterioration models currently in use by VDOT.

## Background

Starting in the 1980s, road agencies in the United States began developing deterministic regression models to predict network-level pavement condition (George et al. 1989). The U.S. Army Corps of Engineers proposed grouping data by sections with similar characteristics into pavement families (i.e., homogeneous sections) to develop pavement deterioration models (Nunez and Shahin 1986, Shahin et al. 1987, PAVER 2014). According to this approach, a pavement family should belong to the same maintenance system according to functional classification and have the same pavement type surface, similar traffic levels, and similar repair and maintenance records. It is assumed that sections belonging to the same pavement family have the same pattern of deterioration over time and this pattern is considered to be representative of the overall performance of that pavement family.

The pavement families approach is currently used and implemented in pavement management system (PMS) software by many U.S. departments of transportation (DOTs) such as TxDOT (Stampley et al. 1995), Indiana DOT (Gulen et al. 2001), Louisiana DOT (Khattak et al. 2009), PennDOT (Wolters and A 2010), Delaware DOT (Mills et al. 2012), Colorado DOT (Saha et al. 2017), and North Carolina DOT (Chen and Mastin 2016).

The use of pavement family modelling techniques allows agencies to incorporate simple default models capable of predicting the average behavior/pattern of a pavement family into the agency PMS. However, pavement deterioration and performance are highly variable due to many factors that introduce heterogeneity, bias, and uncertainty in the pavement condition. Furthermore, there often are important factors that affect pavement deterioration that are not included in the model. These sources of uncertainty and high variability appear as "noise" that could affect expected pavement deterioration rates and patterns that could be inconsistent with engineering judgment (Pierce et al. 2013). In fact, due to the high level of heterogeneity bias in pavement prediction (Prozzi and Madanat 2004, Chen and Mastin 2016), the pavement condition of specific sections will differ from the average model-predicted pavement condition.

Recently, researchers have developed statistical models for transportation asset management with the capability of incorporating uncertainties into deterioration forecasting (Lethanh et al. 2014, Chang and Ramirez-Flores 2015).

Particularly, deterioration models used in PMSs are limited because predictions do not accurately reflect actual pavement conditions, creating a gap between pavement network models and the specific pavement sections (Giummarra et al. 2007). Some authors have worked on the development of mathematical methods used in other disciplines, such as the empirical Bayesian (EB) approach (Zellner 1996), to combine existing knowledge (prior) and information obtained from recent observations. Some researchers have been working to apply this approach to pavement deterioration modelling based on updated pavement deterioration curves using pavement performance monitored in subsequent years (Han and Do 2015). Bayesian inference is used in Markov models which are widely applied to the deterioration process of civil infrastructures to periodically update Markovian Transition Probability Matrices (TPM) between deterioration discrete condition states (CS) as new inspection data become available. Tabatabaee and Ziyadi (2013) used Bayesian inference to accommodate uncertainties between expert-derived TPM and measurement error from inspections using the Minnesota DOT MnROAD test facility. Hong and Prozzi (2006) studied the Bayesian approach using a Markov Chain Monte Carlo simulation for the bridge deck deterioration process to update model parameters as data become available. Prozzi and Madanat (2004) developed pavement performance models using the EB approach based on the Pavement Serviceability Index (PSI). Abaza (2016) proposed a technique based on the "back-calculation" of the discrete Markov model using two consecutive cycles of pavement distress assessment. Other researchers used the Bayesian approach to analyse average pavement indicators and identify homogenous sections (Cafiso and Di Graziano 2012, de León Izeppi et al. 2011).

More recently, Katicha et al. (2016b) proposed a pavement model for the Virginia interstate network that includes the structural condition of the pavement as a variable considered in the deterioration modelling technique. This model assumes that the deterioration process can be expressed as an NB regression and uses the LEB approach to better estimate the future pavement condition.

Although previous work has used similar models to evaluate the pavement deterioration process, this work contributes to the current state of the knowledge on the topic by providing a methodology for modelling pavement deterioration using the NB regression model and the LEB approach at the network level to obtain a better estimate of future pavement deterioration. The LEB is used to combine the average pavement condition estimated by network-level NB pavement deterioration models for families of pavements (Interstate, Primary, and Secondary) using VDOT PMS data, with the most recent observed/measured condition recorded from the inspection of the network. This methodology has the main advantage that it can be applied by DOTs that evaluate pavement condition in terms of a composite index based on the Pavement Condition Index (PCI) and can be included in the network-level decision-making process to improve the prediction of future pavement deterioration. This methodology uses the LEB approach as an adjustment of the Bayesian approach, introducing a parameter α to effectively account for the variance of observed condition data in the deterioration model by minimizing the MSE between the estimated pavement deterioration and the actual deterioration. The effective use of the LEB approach requires an effort by the road agency to calibrate the Bayesian estimator to their local conditions. Therefore, a methodology for the calibration of the LEB estimator using the leave-one-out cross-validation process has been provided.

***Development of pavement deterioration curves***

VDOT summarizes pavement condition data using a CCI pavement rating, a numerical composite index ranging between 100 (perfect condition) and 0 (failed condition). This index, similar to the widely used PCI, reflects the road surface's condition as a function of the type, level of severity, and quantity of distress (Shahin (2005), McGhee (2002), and Loprencipe and Pantuso (2017)). The CCI is determined as the minimum of two rating indexes: the Load-Related Distress (LDR) and Non-Load Related Distress (NDR) (McGhee 2004).

Virginia's roadway system is classified into three categories: interstate, primary, and secondary. The system is surveyed annually for interstate and primary roads and every five years for secondary roads by automated data collection techniques using digital images and automated crack detection and rutting measuring methodologies (VDOT 2012). VDOT uses a comprehensive PMS to support decisions regarding the preservation and renewal of the networks.

The current network-level pavement performance models used by VDOT give an adequate prediction of the average pavement condition for each pavement family as a function of age, but do not consider the deviation of specific pavement sections from the average family performance. Thus, the pavement deterioration process is processed using two types of models: specific-site, project-level models based on the construction and maintenance historical records, and default network-level models when the data for specific sections is not available. The default models are nonlinear regression models for interstate and primary roads developed by Stantec Inc. (2007) to estimate the CCI as a function of the pavement age, where the pavement age is defined as the number of years since construction or last significant maintenance treatment. Katicha et al. (2016a) also included the pavement structural condition along with pavement age to develop deterioration for the interstate network and applied an LEB to combine the model estimate with the measured condition to obtain a better prediction of pavement future condition.

## Methods

### *Overview*

This study used pavement data available in the VDOT PMS for flexible pavement sections in interstate, primary, and secondary networks for three maintenance districts (Richmond, Hampton Roads, and Northern Virginia). The data included pavement condition data and repair and maintenance records from 2007 to 2016. The pavement condition rating table (pavement condition in terms of CCI) and the structure data table (including the year and type of last maintenance treatment performed on each 0.1-mile section) were merged using geographical location. The location is based on VDOT's Linear Referencing System (LRS), which uses milepost position along the routes. The matching procedure took into account that possibility that the begin and end point of the condition data and repair and maintenance treatments and even the inspection sections may not always match. The matching produced a list of values (age, CCI) for a total of 61,501 pavement sections (0.1-mile road segments) and 343,589 observations of different pavement age, number of years between CCI inspections, and year of last significant treatment record that reset pavement age (each pavement section had between 5 and 6 years of data).

### *Analysis of CCI data: Data cleansing and filtering process considerations*

The use of suitable filtering and cleansing procedures for the pre-processing of data analysis is a key aspect to developing a pavement deterioration model that can accurately explain the observed data. The following considerations have been found relevant in this study:

(1) **Pavement condition data at old ages is often inherently biased**: One of the main objectives of highway agencies is to provide road users with a high level of service. When the CCI of a section falls below a minimum value, the section typically receives a corrective maintenance (CM), restorative maintenance (RM), or reconstruction (RC) treatment, which resets its life. Thus, there are not many sections with low CCI condition. This can lead to a biased

deterioration model because condition data for old pavements may not represent the average pavement condition of all typical sections, but rather the behavior of the best-performing pavements in the family. To account for this, the model should be fitted to a range of pavement ages for which this effect is not too pronounced.

(2) **Some treatments are not recorded**: Analysis of the CCI data revealed that there are many instances of sections with significant differences in condition between two consecutive years. More importantly, there are instances where the CCI difference between two consecutive years is highly positive, which is unusual for the process of deterioration of pavements. Some of these are thought to be due to maintenance treatments that are not recorded in the database**.**

Figure 1, Figure 2, and Figure 3 present box plots of pavement condition CCI as a function of pavement age for the three systems studied. Pavement CCI data grouped by pavement age is depicted using a box plot where the median value of CCI condition is represented by a solid line, and the box represents the data between the 25th and 75th CCI percentiles. The figure shows that the distribution of pavement condition has a high variability within each pavement family. This is due to the grouping of various pavement sections in the family, in addition to the inherent pavement data condition variability (related to the collection techniques, use of a composite index, and localization of the distress within the sections).

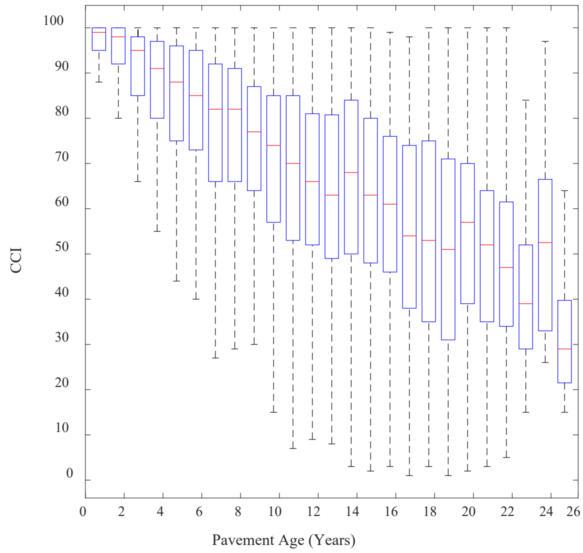



**Figure 1. Pavement CCI as function of pavement age of Interstate network.**

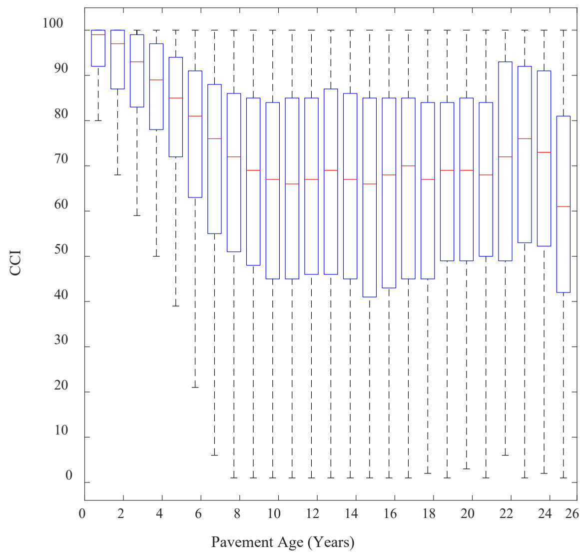



**Figure 2. Pavement CCI as function of pavement age of Primary network.**

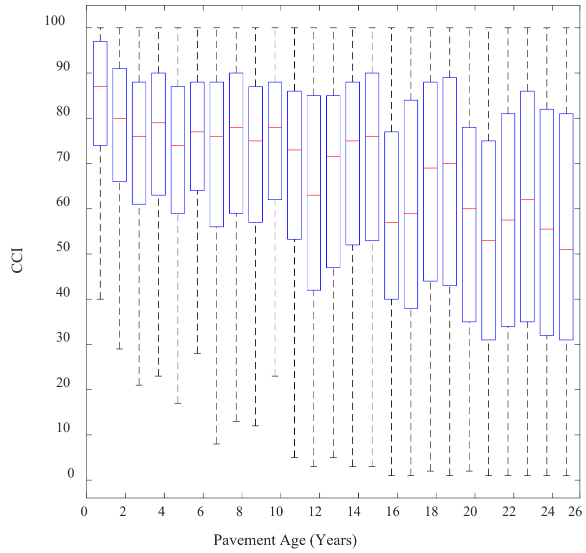



**Figure 3. Pavement CCI as function of pavement age of Secondary network.**

### *The Empirical distribution of pavement condition: Negative binomial (Poisson-Gamma model)*

The measured pavement deterioration (DI) can be defined according to equation (**1**) as the complementary to the measured CCI.

$$DI = 100 - CCI \tag{1}$$

The distributions of DI for each pavement family were fitted with different theoretical statistical distributions: a normal distribution, a Poisson distribution, and an NB distribution. Figure 4, Figure 5, and Figure 6 illustrate the distribution fittings and the empirical distribution for the Interstate, Primary, and Secondary systems, respectively.

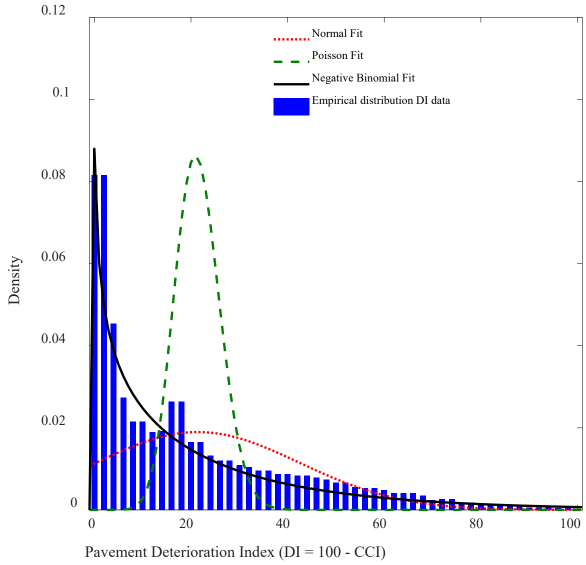



**Figure 4. Distribution of pavement deterioration (DI) of the Interstate network.**

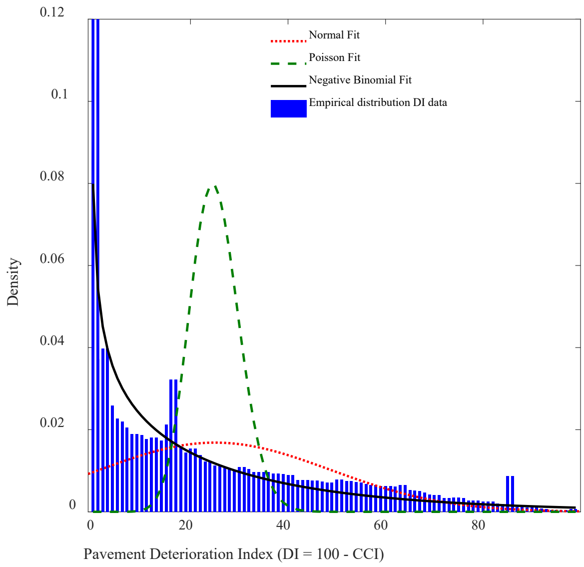



**Figure 5. Distribution of pavement deterioration (DI) of the Primary network.**

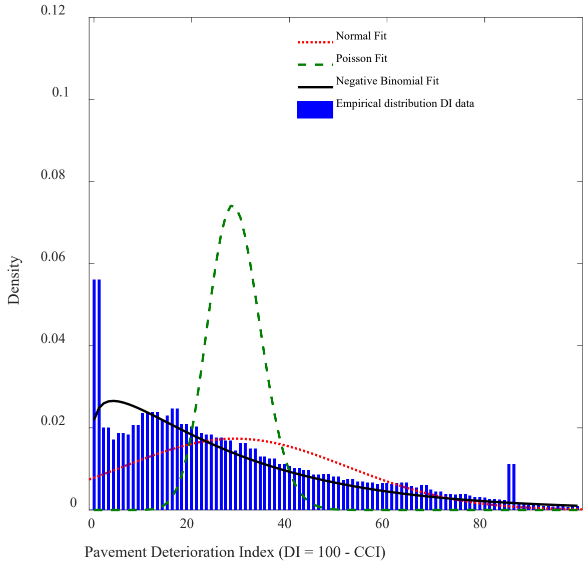



**Figure 6. Distribution of pavement deterioration (DI) of the Secondary network.**

This analysis showed that the Poisson distribution (with equal mean and variance by definition) provides a poor representation of pavement data; however, the Poisson distribution combined with the Gamma distribution which arises to the NB distribution provides a good representation of the pavement data, with each pavement family outperforming the normal distribution. The NB distribution (Poisson-Gamma model) is a discrete distribution that is more flexible than the Poisson distribution (for which the mean is equal to the variance) that allows the variance to be greater than the mean. Statistical analysis of the empirical distribution of the pavement deterioration data showed that the NB distribution provides a better representation of observed data compared with a normal distribution (Ercisli 2015).

The empirical distribution of the pavement condition over time has been studied for each pavement family. Figure 7, Figure 8, and Figure 9 explain the observed pavement condition (CCI) with descriptive histograms and the empirical distribution cummulative distribution function (CDF) and NB distribution fitting for each pavement family. These figures help to better understand CCI boundaries and the distribution of the CCI along pavement age.

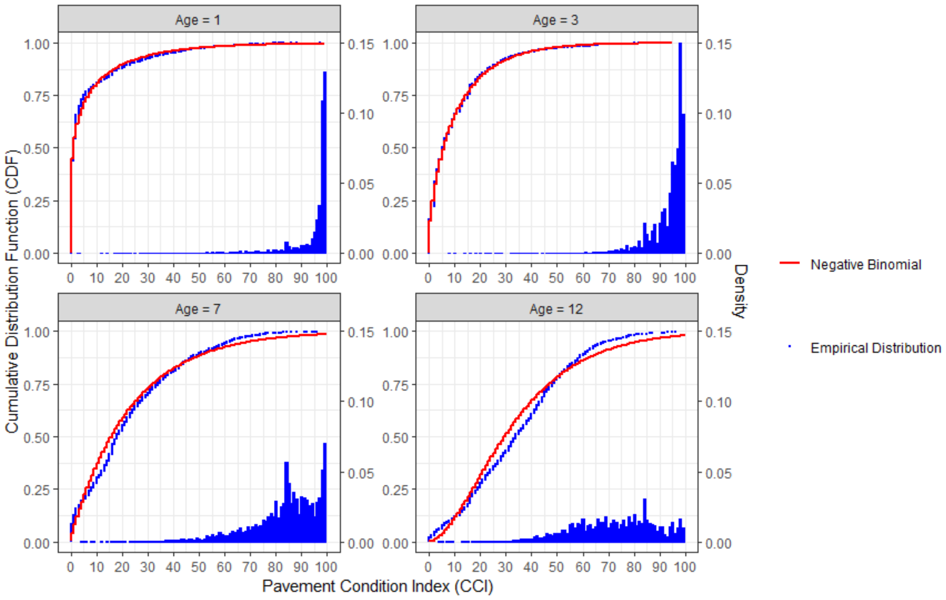



**Figure 7. Empirical distribution of pavement CCI and NB distribution fit along pavement age for the Interstate system.**

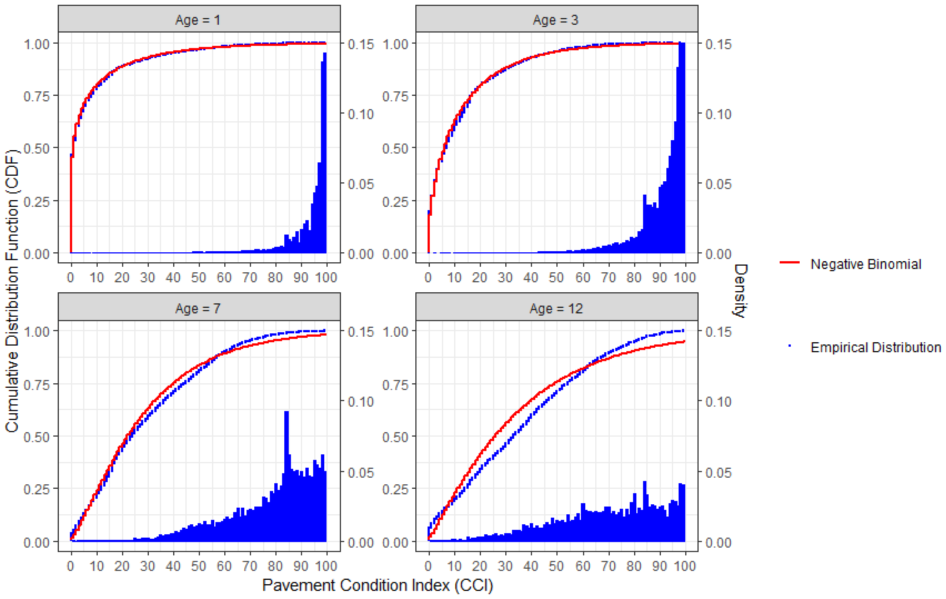



**Figure 8. Empirical distribution of pavement CCI and NB distribution fit along pavement age for the Primary system.**

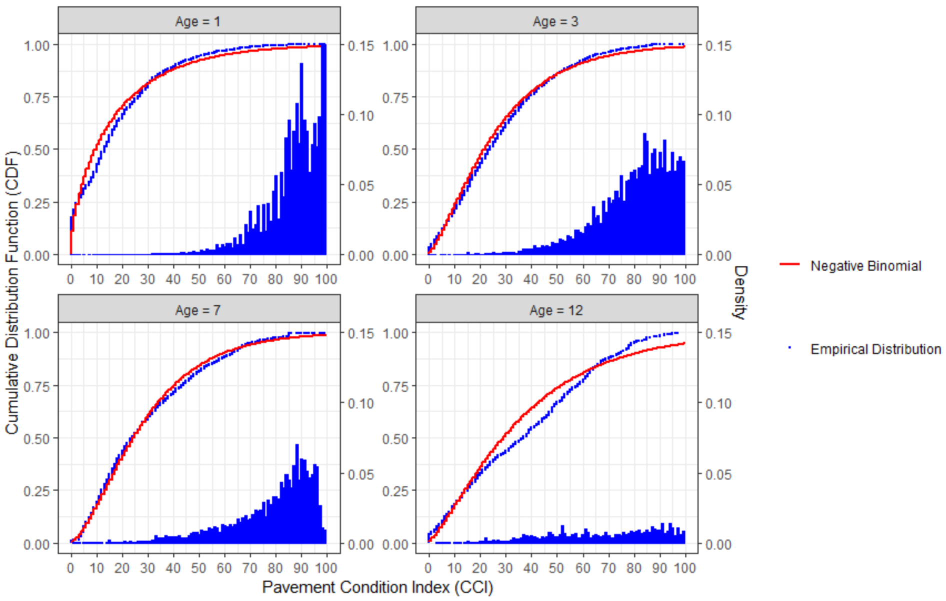



**Figure 9. Empirical distribution of pavement CCI and NB distribution fit along pavement age for the Secondary system.**

Table 1 provides descriptive CCI statistics for different pavement ages (age=1, age =3, age=7, and age =12). The NB distribution fitting parameters have been provided. The $\chi^2$ test was used for testing the goodness of fit for the NB distribution of the empirical distribution of DI (complementary of CCI).

***Table 1.*** *Descriptive Statistics and $\chi^2$ test of NB Distribution Fitting*

| **Family** | | **NB size** | **NB μ** | **χ2 test** | **DF** | ***p*-value** |
|---|---|---|---|---|---|---|
| **Interstate** | age=1 | 0.234 | 6.862 | 395.561 | 82 | 1.37E-42 |
| | age=3 | 0.661 | 10.454 | 396.362 | 76 | 7.46E-45 |
| | age=7 | 1.014 | 22.903 | 526.278 | 88 | 1.21E-63 |
| | age=12 | 2.231 | 34.809 | 526.278 | 88 | 1.21E-63 |
| **Primary** | age=1 | 0.234 | 6.862 | 395.561 | 82 | 1.37E-42 |
| | age=3 | 0.661 | 10.454 | 396.362 | 76 | 7.46E-45 |
| | age=7 | 1.014 | 22.903 | 526.278 | 88 | 1.21E-63 |
| | age=12 | 2.231 | 34.809 | 373.335 | 89 | 1.72E-36 |
| **Secondary** | age=1 | 0.234 | 6.862 | 395.561 | 82 | 1.37E-42 |
| | age=3 | 0.661 | 10.454 | 396.362 | 76 | 7.46E-45 |

| | | | | | |
|---|---|---|---|---|---|
| age=7 | 1.014 | 22.903 | 526.278 | 88 | 1.21E-63 |
| age=12 | 2.231 | 34.809 | 373.335 | 89 | 1.72E-36 |

***Data cleansing and filtering process***

As highlighted in Figure 1, Figure 2, and Figure 3 the general CCI trends are reasonable, as the average CCI decreases as the age increases, but the average CCI reaches an almost constant level for the age of 14, 11, and 8 for the Interstate, Primary and Secondary pavement families, respectively. After these ages, the average CCI for each family is inconsistent with the engineering expectation because there is a positive trend of the average of CCI for some years. This could be due to missing treatments in the database and/or the bias previously discussed because bad pavements are treated at younger ages. Considering old sections with very good condition for the development of the pavement model can lead to a biased model that would not be representative of the average performance of all pavement sections. For this reason, the regression models were fitted to sections with ages less than 14, 11, and 8 years for the Interstate, Primary, and Secondary network pavement families, respectively. For the Interstate system, the model was fitted for 29,971 observations in 3,601 sections, 91% of the data. Thus, for the primary system, only the sections with age greater than 11 years were filtered out (leaving 170,810 observations in 22,316 sections, or 74% of the data).

For the Secondary network, there seems to be even more uncertainty about repair and maintenance treatments recorded in the database. This can significantly influence the general behaviour of pavement models because it introduces bias, which shows in Figure 3. The average pavement condition for the Secondary network family after the age of 4 years is approximately constant and the average CCI increases as pavement age increases, opposite to the usual deterioration process. Specifically, 16,773 measurements (21% of the data set) have CCI values greater than 80 points for pavement ages greater than 4 years. Since such good performance is unusual, this set of measurements has been excluded from the data set used to fit the model for the

Secondary roads pavement family (the model was fitted for 38,354 observations in 26,167 sections, or 48% of the data). Table 2 summarizes the sample sizes considered to fit the deterioration models for each family.

***Table 2.*** *Sample Size of Pavement Families Included in the Study*

| Pavement family model | No. sections | No. observations | No. sections model | No. observations model |
|---|---|---|---|---|
| Interstate Network | 3,601 | 32,685 | 3,601 | 29,971 |
| Primary Network | 22,316 | 231,197 | 18,272 | 170,810 |
| Secondary Network | 35,584 | 79,707 | 26,167 | 38,354 |
| Total | 61,501 | 343,589 | 48,040 | 239,135 |

***Negative binomial model development and VDOT pavement deterioration models***

Because the NB model provides a good representation of the data, NB regression, which is a type of Generalized Linear Model, was used to determine the deterioration model for each type of pavement family.

The deterioration model relates the pavement deterioration as a function of pavement age as shown in equation (**2**). NB regression is performed to determine the parameters $\beta_0$, $\beta_1$, $\beta_2$, and an overdispersion parameter $\phi$, which is related to the model variance as shown in Equation (**3**).

$$DI_{NB_{model}} = \exp(\beta_0 + \beta_1 \cdot age + \beta_2 \cdot \log(age)) = age^{\beta_2} \cdot \exp(\beta_0 + \beta_1 \cdot age) \quad (2)$$

$$Var(DI) = DI_{NB_{model}}(1 + \phi DI_{NB_{model}}) \quad (3)$$

Current VDOT default network-level pavement deterioration models are used whenever there are no section-specific data available to predict the pavement condition in terms of CCI for a given pavement age. Thus, as a baseline, three VDOT default models—Power model Eq. (**4**), Sigmoidal model Eq. (**5**), and the VDOT exponential model Eq. (**6**) developed by Stantec Inc. (2007)—were fitted to the filtered and cleaned data for the Interstate, Primary, and Secondary networks:

$$CCI_{P\_model} = a \cdot age^b + c \quad (4)$$

$$CCI_{S\_model} = 100 - d \cdot \exp(\frac{-e}{age}) \quad (5)$$

$$CCI_{VDOT\_model} = 100 - exp\left(f + g \cdot h^{\left(\frac{1}{age}\right)}\right) \quad (6)$$

where the regression parameters were obtained using ordinary least-squares fitting. The VDOT default deterioration models fitted for the pavement condition data from the PMS were compared with the proposed NB deterioration model.

***Empirical Bayesian approach***

The NB distribution arises as a compound mixture of Poisson distributions where the mixing distribution of the Poisson rate is the Gamma distribution (Poisson-Gamma model) and can be expressed parametrically as in Eq.(**7**) (Hilbe 2011).

$$f_{NB}(x;r,p) = \int_0^\infty f_p(x;\lambda) \cdot f_G(\lambda;r,p) \cdot d\lambda = \int_0^\infty \left(\frac{\lambda^x}{x!} \cdot \exp(-\lambda)\right) \cdot \left(\frac{\lambda^{r-1}}{\left(\frac{p}{1-p}\right)^r \cdot \Gamma(r)} \cdot \exp\left(-\lambda \cdot \frac{(1-p)}{p}\right)\right) \cdot d\lambda \quad (7)$$

The Poisson-Gamma model gives rises to a Bayesian model with the Gamma distribution prior (conjugate prior of the Poisson distribution). The Gamma distribution prior is parametrized by two parameters which are related to the mean and overdispersion, according to Eqs. (**8**), (**9**), and (**10**).

$$DI_{NB_{model}} = \frac{p \cdot r}{1-p} \quad (8)$$

$$Var(DI) = \frac{p \cdot r}{(1-p)^2} \quad (9)$$

$$\phi = \frac{Var(DI) - DI_{NBmodel}}{DI_{NBmodel}} = \frac{\frac{p \cdot r}{(1-p)^2} - \frac{p \cdot r}{1-p}}{\frac{p^2 \cdot r^2}{(1-p)^2}} = \frac{\frac{p \cdot r - p \cdot r \cdot (1-p)}{(1-p)^2}}{\frac{p^2 \cdot r^2}{(1-p)^2}} = \frac{1}{r} \quad (10)$$

In the EB approach, these are $DI_{NB\_model}$ and $\phi$, determined from the data using the NB regression. Once the Gamma distribution parameters are evaluated, the posterior distribution is obtained as expressed in Eq. (**11**).

$$f_{G\_Posterior}\left(\lambda; \frac{1}{\phi} + DI, \frac{\phi \cdot DI_{NB\_model}}{\phi \cdot DI_{NB_{model}} + 1}\right) = \frac{\frac{1}{\lambda \phi}^{+DI-1}}{\left(\frac{1}{\phi \cdot DI_{NB_{model}}}\right)^{\frac{1}{\phi}+DI} \cdot \Gamma\left(\frac{1}{\phi}+DI\right)} \cdot exp\left(-\lambda \cdot \left(\frac{\phi \cdot DI_{NB_{model}}+1}{\phi \cdot DI_{NB_{model}}}\right)\right) \quad (11)$$

Consequently, a point estimate (the posterior mean) of the pavement deterioration, $DI_{\mathrm{EB}}$, can be calculated as a weighted average of the pavement condition predicted from the prior distribution ($DI_{\mathrm{NB\_Model}}$) and the observed mean pavement deterioration ($DI$), obtained as follows in Eq. (**12**):

$$DI_{EB} = \left(\frac{1}{\phi} + DI\right) \cdot \left(\frac{\phi \cdot DI_{NB_{model}}}{\phi \cdot DI_{NB_{model}} + 1}\right) = \left(\frac{DI_{NB\_model} + \phi \cdot DI_{NB_{model}} \cdot DI}{\phi \cdot DI_{NB_{model}} + 1}\right) \quad (12)$$

The posterior pavement deterioration estimated ($DI_{\mathrm{EB}}$) is calculated by rewriting Eq. (**12**) in Eq. (**13**) as follows:

$$DI_{EB} = \frac{1}{\phi \cdot DI_{NB_{model}} + 1} \cdot DI_{NB_{model}} + \left(1 - \frac{1}{\phi \cdot DI_{NB_{model}} + 1}\right) \cdot DI \quad (13)$$

In practical terms, Bayes' formula combines the average behavior of all pavement sections (expressed by the prior, $DI_{NB\_model}$) with recent observations of specific sections (observed $DI$) to obtain a better estimate of the pavement condition $DI_{EB}$ (complementary of $CCI_{EB}$, which is the posterior).

For example, from the latest pavement inspections, 10 pavement observations of different pavement ages (data retrieved from PMS after applying the matching procedure explained before) were picked randomly. For each of these observations, EB was applied to get a better estimate of the pavement condition for that specific section (see Figure 10). Questionable observed data unusual for the process of deterioration of pavements (pavements with high deterioration at early ages or little deterioration at late ages) were corrected to a better estimate of the deterioration prediction closer to the mean performance of pavement sections (represented by the proposed deterioration model).

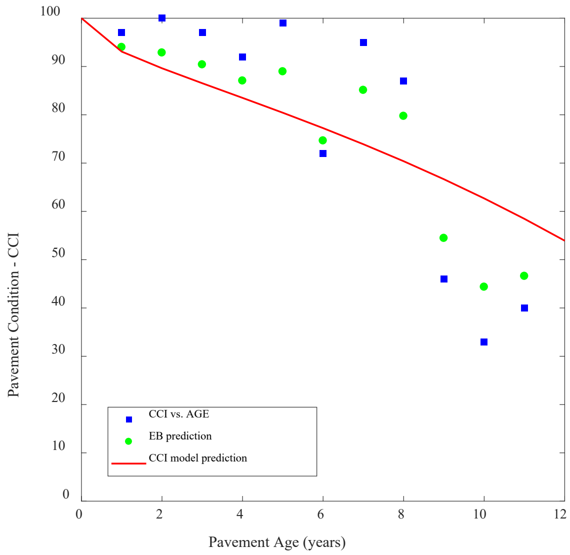



**Figure 10. Example of application of the EB approach.**

According to Efron and Morris (1973), the improvement of the LEB approach is such that the MSE is reduced by the factor shown in Equation (**14**):

$$\frac{\sigma_S^2}{Var(DI)} = \frac{\sigma_S^2}{\sigma_S^2 + \sigma_{error}^2} \tag{14}$$

where $\sigma^2_s$ is the variance of the data and Var(*DI*) is the total variance of the observed data.

According to the assumption of the Poisson-Gamma model, Var(*DI*) can be decomposed by the variance of the pavement sections' performance, $\sigma^2_s$ estimated by the Gamma distribution, and the variance of the error estimated from the pavement data, estimated by the Poisson distribution, $\sigma^2_{Poisson}$. However, the variance of the estimated error $\sigma^2_{error}$ in the measured pavement deterioration is higher than what is predicted by the Poisson distribution (see Figure 11), thus the error of the measured/observed deterioration is underestimated, leading to a suboptimal use of the EB estimator.

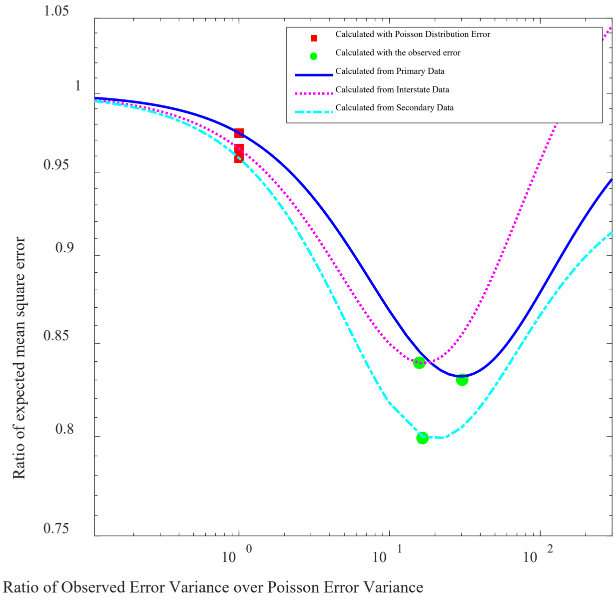



**Figure 11. Error in estimating the pavement condition with the EB method as a function of estimated error variance.**

Figure 11 illustrates the error in estimating the pavement condition with the EB method as a function of estimated error variance. When the measurement error is assumed to be practically zero, the EB estimate is equal to the observed pavement condition and the error is equal to 1. As the estimate of the measurement error gets closer to the observed measurement error variance, the MSE of estimating the pavement condition using the EB method decreases. On the other hand, when the estimate of the measurement error variance is greater than the observed measurement error of the variance, the error of EB starts to increase. The best estimate is obtained when the measurement error variance is correctly estimated.

Therefore, an adjustment has been made in the model to minimize the MSE between the estimated pavement deterioration and the true deterioration, $\sigma^2_{error} = \alpha \cdot \sigma^2_{Poisson}$, and the EB approach becomes an LEB. The adjustment can be done by modifying Eq. (**13**) as shown in Eq. (**15**), where $\phi_c = \phi/\alpha$.

$$DI_{EB} = \left(\frac{1}{\phi_c \cdot DI_{NB_{model}} + 1}\right) \cdot DI_{NB_{model}} + \left(1 - \frac{1}{\phi_c \cdot DI_{NB_{model}} + 1}\right) \cdot DI \qquad (15)$$

Alternatively, by substituting Eqs. (**10**) and (**14**) in Eq. (**13**), the LEB estimator is calculated using Eq. (**16**), which can be used for any distribution and without the knowledge of the appropriate form.

$$DI_{EB} = \left(1 - \frac{\sigma_s^2}{\sigma_s^2+\sigma_{error}^2}\right) \cdot DI_{model} + \left(\frac{\sigma_s^2}{\sigma_s^2+\sigma_{error}^2}\right) \cdot DI \qquad (16)$$

where $\sigma^2_{error}=\alpha\ \sigma^2_{Poisson}$; $\sigma^2_{Poisson}$ is the variance predicted by the Poisson distribution (the mean value of the data that fit the model), and $\sigma^2_{error}$ is the variance estimated from the data using the difference sequence method Katicha et al., (2016b, 2016a).

## Results

### *Deterioration model development*

For each pavement family, the parameters of the NB pavement deterioration model Eq. (**3**) were estimated using the NB regression on the filtered data measurements. In addition, the filtered data were fitted with the VDOT default deterioration models. Table 3 summarizes the estimated parameters obtained using the various models and the coefficient of determination.

***Table 3.*** *Goodness of Fit Pavement Deterioration Models for the Defined Pavement Families*

| Pavement family model | Estimated Parameters | $R^2$ |
|---|---|---|
| *Interstate Network* | | |
| Power model | a=-5.150; b=0.756; c=100 | 0.292 |
| Sigmoidal model | d=56.030; e=5.956 | 0.275 |
| VDOT model | f=-16.070; g=17.710; h=0.961 | 0.291 |
| NB | $\beta_0$=1.658; $\beta_1$=0.018; $\beta_2$=0.654; $\phi$=1.125; $\alpha$ = 15.86 | 0.289 |
| *Primary Network* | | |
| Power model | a=-0.09; b=2.27; c=95.29 | 0.227 |
| Sigmoidal model | d=46.360; e=7.856 | 0.184 |
| VDOT model | f=1.384; g=0.115; h=0.313 | 0.226 |
| NB | $\beta_0$=1.353; $\beta_1$=0.228; $\beta_2$=-0.250; $\phi$=1.322; $\alpha$ = 30.50 | 0.227 |
| *Secondary Network* | | |
| Power model | a=-21.9; b=0.412; c=105.000 | 0.190 |
| Sigmoidal model | d=46.230; e=1.132 | 0.184 |
| VDOT model | f=-16.040; g=18.880; h=0.975 | 0.190 |
| NB | $\beta_0$=2.837; $\beta_1$=0.001; $\beta_2$=0.487; $\phi$=0.709; $\alpha$ = 16.64 | 0.190 |

The four models are represented graphically for the Interstate family (Figure 12), Primary family (Figure 13), and Secondary family (Figure 14). In general, all the models adequately fit the average value of the pavement deterioration. However, due to the high variability of the data, the fitted models will not fit well with individual pavement sections.

The parameter ϕ presented for the NB model in Table 3 is the overdispersion parameter as defined previously. The $R^2$ was calculated manually for the NB regression using the residual sum of squares (SSE) of the model and the total variance of the data used to fit the models.

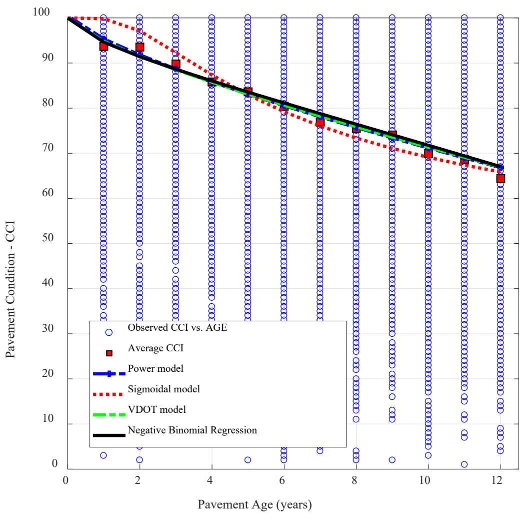



**Figure 12. Comparison between CCI deterioration curves using VDOT default models and NB regression model for the Interstate system.**

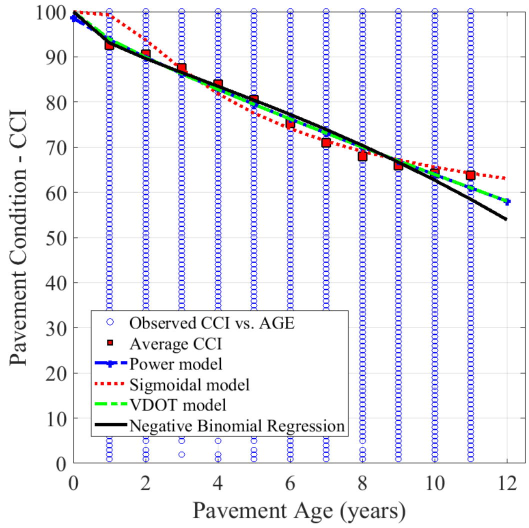



**Figure 13. Comparison between CCI deterioration curves using VDOT default models and NB regression model for the Primary system.**

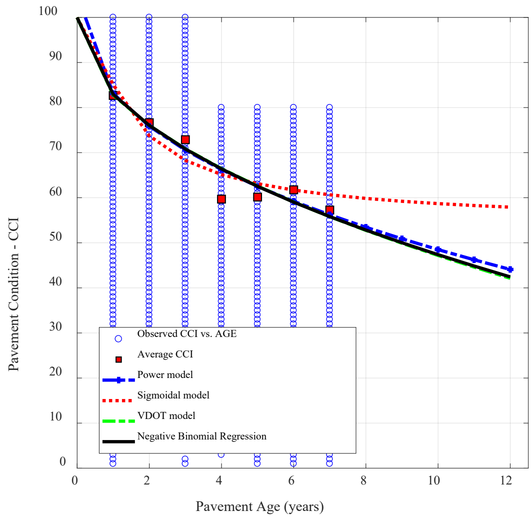



**Figure 14. Comparison between CCI deterioration curves using VDOT default models and NB regression model for the Secondary system.**

### *Validation of modelling procedure*

As mentioned previously, the EB approach can provide a better estimation of pavement condition compared with the estimation of the CCI model prediction or the CCI value recorded in the database. The best way to validate the LEB approach would be to estimate the "true" value that is being calculated by the model and verify that the chosen procedure gives a better estimation of the true value (next year's condition) compared with that obtained from estimation of pavement condition (actual condition) without the use of any modelling procedure. Therefore, a leave-one-out cross-validation procedure was implemented and five different methods were compared using the MSE as the evaluation criterion. The procedure followed including the following steps:

(1) Process the data to get the measurements (observations) $O$ on each section $S$ and determine the inspection year for each measurement.

(2) Remove from pavement section $S_i$ all the measurements obtained after year $Y_i$.

(3) Fit the model to the remaining data in the data set.

(4) Evaluate the different methods to determine estimation of the removed measurements and calculate the MSE.

Five different approaches were compared in terms of their ability to predict the pavement deterioration (illustrated in Figure 15):

(1) **Method 1**. Predict the future pavement condition with the most recent observations on the section $S$i in the PMS. This choice implies not modelling the pavement condition and assumes that the pavement does not deteriorate from last recorded inspection:

$$CCI_{i+1}^{M1} = CCI_i \tag{17}$$

(2) **Method 2**. Predict the pavement condition using the fitted family model using the NB model, Eq. (**2**):

$$CCI_{i+1}^{M2} = CCI_{NB_{model,i+1}} \tag{18}$$

(3) **Method 3**. Predict the pavement condition using the LEB approach using Eqs. (**15**) and (**16**):

$$CCI_{i+1}^{M3} = CCI_{EB\_i} \tag{19}$$

(4) **Method 4**. Predict the condition from the most recent observations of section $S_i$ and add the deterioration of section $S_i$ obtained as follows:

$$CCI_{i+1}^{M4} = CCI_i + \left(CCI_{NB_{model,i+1}} - CCI_{NB_{model,i}}\right) \tag{20}$$

(5) **Method 5**. Predict the future condition using the LEB approach adding the deterioration of the section $S_i$ obtained from the model:

$$CCI_{i+1}^{M5} = CCI_{EB\_i} + \left(CCI_{NB_{model,i+1}} - CCI_{NB_{model,i}}\right) \tag{21}$$

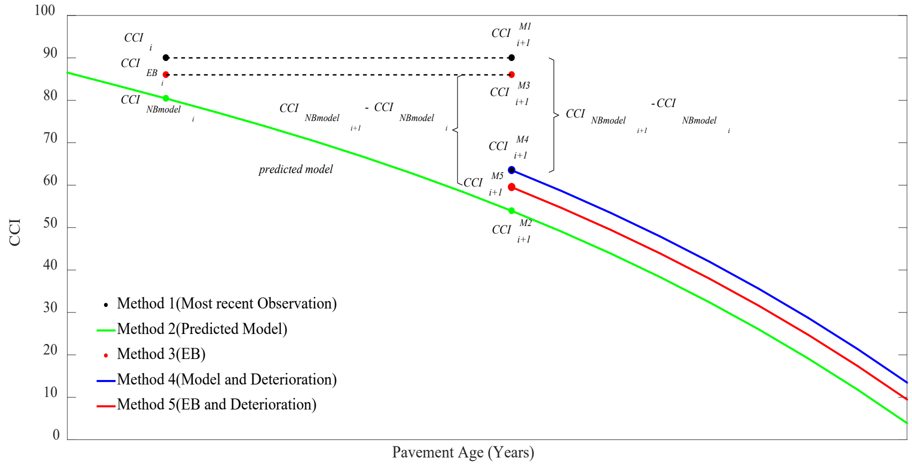



**Figure 15. Illustration of the approaches to calculate the estimate of future pavement condition.**

Each pavement condition measurement recorded in the database is excluded from the model, and the value is estimated by fitting the model for the remaining measurements in the database, and the obtained estimate of the "true" observation is compared with the observed measurement that it is recorded in the database.

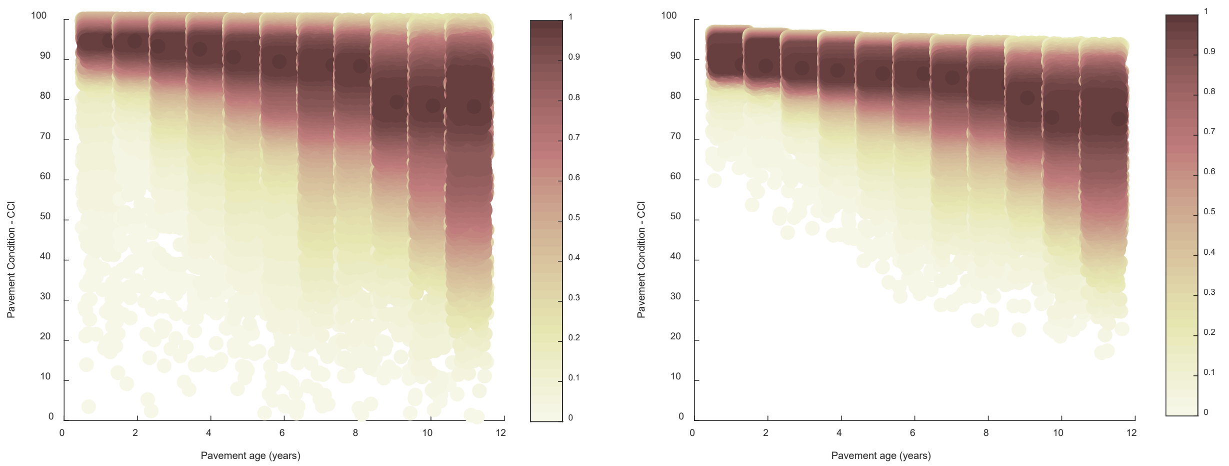
***Table 4.*** *MSE of Prediction of the Estimate of Pavement Condition using Different Approaches*

| VDOT System Family | MSE Prediction | MSE Ratio with respect to Method 1 | Improvement Prediction Ratio with respect to Method 1 |
|---|---|---|---|
| Interstate Network Family | | | |
| Method 1 | 292.17 | 1.000 | - |
| Method 2 | 325.68 | 1.115 | -12 % |
| Method 3 | 247.69 | 0.873 | 13 % |
| Method 4 | 228.60 | 0.782 | 22 % |
| Method 5 | 188.83 | 0.666 | 33 % |
| Primary Network Family | | | |
| Method 1 | 94.99 | 1.000 | - |
| Method 2 | 113.69 | 1.197 | -20 % |
| Method 3 | 87.89 | 0.925 | 8 % |
| Method 4 | 67.05 | 0.706 | 29 % |
| Method 5 | 60.91 | 0.641 | 36 % |
| Secondary Network Family | | | |
| Method 1 | 499.79 | 1.000 | - |
| Method 2 | 474.17 | 0.949 | 5 % |
| Method 3 | 459.33 | 0.919 | 8 % |
| Method 4 | 461.32 | 0.923 | 8 % |
| Method 5 | 295.94 | 0.592 | 41 % |

The results for the five different approaches (Figure 15) are summarized in Table 4, including the MSE prediction of each method. Method 1 represents the results of not considering any modelling technique; that is, the future pavement condition is considered the same as the most recent observation recorded in the database, and thus is used as the baseline for comparison with the other methods. Method 2 represents the results of predicting pavement condition using the NB regression model, the cross-validation process showed that for the tested data sets (affected by high variance such as Interstate family and Primary family) is ineffective to predict future pavement condition. The ratio of prediction error compared with the Method 1 was -12% for the Interstate family and -20% for the Primary family. However, Method 2 is more effective for the Secondary family (+5% ratio of prediction error compared with Method 1) because is affected by less variance of specific pavement section deterioration prediction compared with the other pavement families.

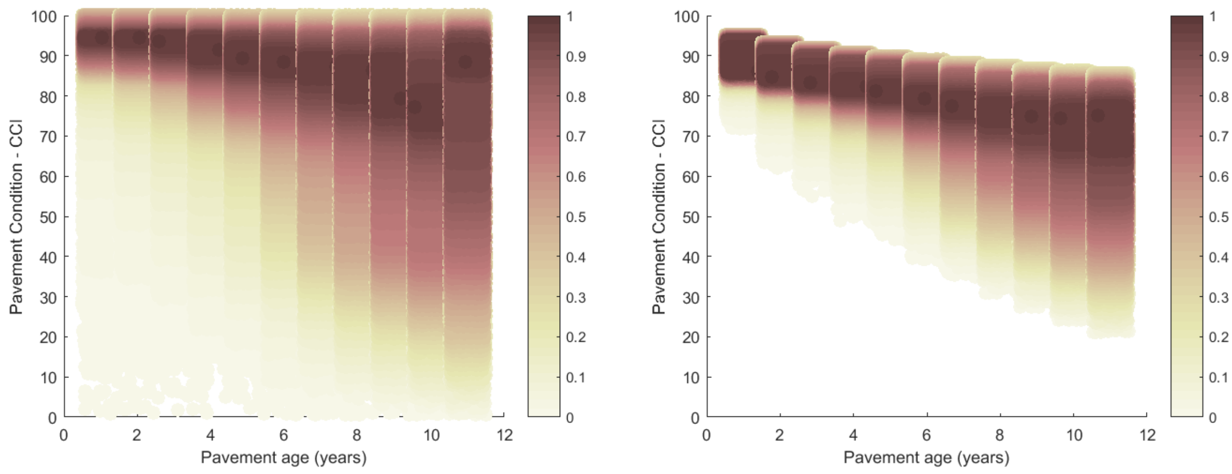
The EB approach used in Method 5, which predicts the future pavement condition as the combination of the EB of the individual pavement section and the expected deterioration predicted by the model, has the best prediction ability. It gives the lowest prediction error with an improvement of 33% for the Interstate family, 36% for the Primary family, and 41% for the Secondary family compared with not considering any pavement deterioration technique (Method 1).

Colour maps were created to visually illustrate the effectiveness of the proposed modelling technique for the Interstate (Figure 16), Primary (Figure 17), and Secondary (Figure 18) pavement families. Pavement condition data in terms of VDOT CCI are represented as a function of pavement age. Density is the estimated average density across the observed data points using a color scale: darker colours represent a greater density of observations and brighter colours represent a lower density of observations. The same scale of colours is used in Figure 16, Figure 17, and Figure 18 (a) and (b). Visual inspection indicates that the application of the linear EB estimator is effective because more reasonable deterioration trends were observed in the colour map after applying the modelling technique. The LEB estimator corrects unreasonable pavement condition predictions for specific pavement sections (pavement sections with little deterioration at late ages and/or highly deteriorated sections at early ages), getting closer to the mean deterioration of the pavement family. A significant reduction of the variance of the pavement condition data is thus obtained.



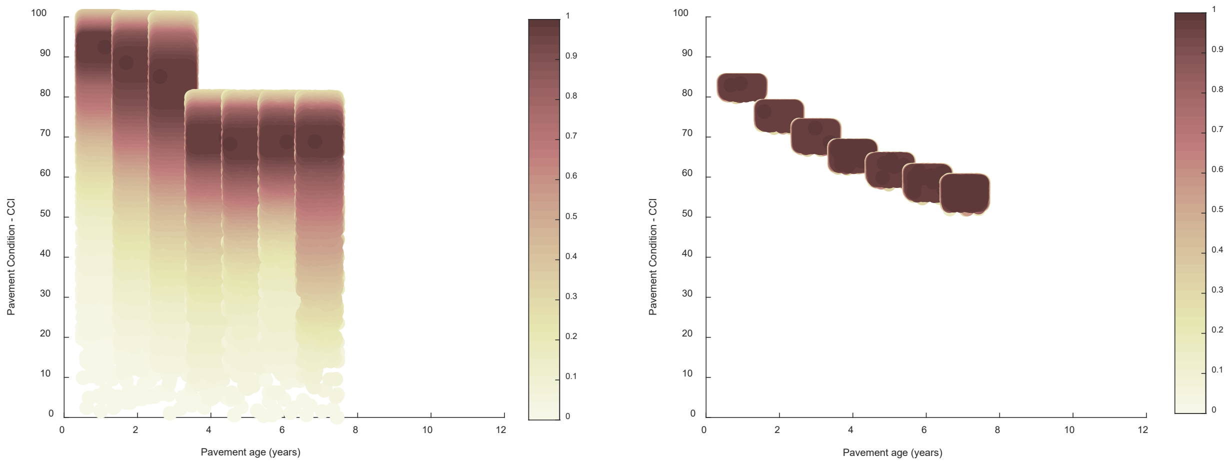

(a) (b)

**Figure 16. Colour map plot of the Interstate network pavement family; (a) Raw observed pavement condition data; (b) Pavement condition modelled using the LEB approach.**




(a) (b)

**Figure 17. Colour map plot of the Primary network pavement family; (a) Raw observed pavement condition data; (b) Pavement condition modelled using the LEB approach.**




(a) (b)

**Figure 18. Colour map plot of the Secondary network pavement family; (a) Raw observed pavement condition data; (b) Pavement condition modelled using the LEB approach.**

**Discussion**

The results presented in this paper show that the presented deterioration modelling approach can improve the prediction of future pavement condition in terms of CCI for VDOT for the pavement families studied (Interstate, Primary, and Secondary).

The empirical distribution of the data is better represented by the NB distribution than the normal distribution). The resulting model using this distribution provides a similar coefficient of determination with respect to the current default models but has the advantage of allowing the pavement condition to be estimated using the LEB approach as a weighted combination of the average condition provided by the model and the observed pavement condition. Furthermore, the definition of the NB distribution as a Poisson-Gamma model allows the use of the EB approach to improve the overall performance of the model by calculating the best estimate of the pavement condition based on the average condition predicted by the model and the last measurement recorded in the network for each specific section. This approach can be useful for pavement network-level predictions by combining the prediction of the fitted model with the observations of the pavement condition.

The main advantage of this method is that it allows better predictions of next year's pavement condition if compared with the prediction of the family pavement deterioration regression model, improving consistency between the network- and project-level deterioration curves.

However, there are still inherent errors in the PMS collected data (missing construction record data, uncertainties in the data collection process) that are necessary to deal with. The cross-validation process proves that estimating the pavement condition from the new deterioration regression model would cause significant errors in the prediction of the deterioration in pavement sections (Method 2). Nevertheless, the model is still needed to account for the mean deterioration of pavement sections and to obtain the LEB estimator.

The use of the LEB approach turns out to be an effective method to estimate the future pavement condition (next year) by subtracting the modelled pavement deterioration from the LEB estimator (Method 5). Method 5 improves the MSE prediction of the future (next year) pavement condition between the observed and predicted future condition by 33% for the Interstate family, by 36% for the Primary family, and by 41% for the Secondary family.

## Conclusions

This paper presents a methodology for modelling the pavement deterioration process at the network level using the LEB. The methodology was tested for historical pavement data for three roadway systems from the VDOT PMS (Interstate, Primary, and Secondary). Based on the analysis performed, some key observations can be summarized as follows:

(1) The deterioration process can be expressed using the NB regression model.

(2) The proposed LEB modelling approach can account for the variance of pavement condition data. It allows the effectiveness of the EB estimator to be maximized to obtain a deterioration model that can get the estimate of the measurement error as the observed measurement error.

(3) The LEB approach can improve the prediction of future pavement condition in terms of CCI for the VDOT for the pavement families studied.

Although, the CCI was used as the performance indicator to analyse the pavement deterioration model at the network level, the methodology proposed is not limited to this indicator; it can be applied to pavement condition data from other road agencies collected in terms of a composite index, such as the PCI. Nevertheless, the effectiveness of the application of this methodology would be limited by the quality of fit between empirical distribution of the pavement data and the NB distribution.

The proposed LEB shows how to effectively adjust the traditional EB approach to maximize the effectiveness of the Bayesian estimator. The calibration methodology presented in the paper tested the LEB for VDOT Interstate, Primary, and Secondary systems with pavement historical data from 2006 to 2017, providing a suitable validation of the approach. Regardless, the validation process is specific to the historical data available, and the application of this approach by other agencies would be suboptimal.  However, in the future, with more performance data available, a better validation of the approach can be achieved.

## Notation List

*The following symbols are used in this paper:*

*CCI* = Critical Condition Index
*DI* = Deterioration Index
$\Gamma$ = Gamma function
f = probability mass function
p = probability of success, negative binomial distribution.
r = number of failures, negative binomial distribution.
$\alpha$ = correction parameter linear empirical Bayes estimator
$\beta$ = parameters of the negative binomial model
$\theta$ = scale of Gamma distribution
$\lambda$ = expected value of Poisson distribution
$\mu$ = mean value of the negative binomial distribution
size = dispersion parameter of negative binomial distribution
$\chi^2$ = Chi-Squared distribution
DF = Degrees of freedom of chi-squared distrubution
*DI* = mean value of the distribution of observed pavement data
$DI_{EB}$ = mean value of the posterior distribution of the observed pavement data
$DI_{NB\_Model}$ = mean value of the prior distribution of observed pavement data
MSE = mean squared error
$\sigma^2_{error}$ = variance of the estimated error
Var(*DI*) = total variance of the observed pavement data
$\sigma^2_{Poisson}$ = variance of Poisson distribution
$\sigma^2_s$ = variance of observed pavement data
$\phi$ = overdispersion parameter of negative binomial regression
$\phi_c$ = corrected overdispersion parameter of negative binomial regression

The authors gratefully acknowledge the Virginia Department of Transportation (VDOT) for providing access to the data used in this research.

## References

Abaza, K. A. 2016. Back-calculation of transition probabilities for Markovian-based pavement performance prediction models. *International Journal of Pavement Engineering,* 17(3), 253-264.

Cafiso, S. and Di Graziano, A. 2012. Definition of Homogenous Sections in Road Pavement Measurements. *Siiv-5th International Congress - Sustainability of Road Infrastructures 2012,* 53, 1069-1078.

Chang, C. M. and Ramirez-Flores, R. A., 2015. *Development of probability-based pavement performance curves for pavement management systems.*

Chen, D. and Mastin, N. 2016. Sigmoidal Models for Predicting Pavement Performance Conditions. *Journal of Performance of Constructed Facilities,* 30(4), 04015078.

de León Izeppi, E., *et al.*, Continuous friction measurement equipment (CFME) data processing and analysis software. ed. *Transportation Research Board Annual Meeting, Washington DC, USA*, 2011.

Efron, B. and Morris, C. 1973. Stein's Estimation Rule and its Competitors—An Empirical Bayes Approach. *Journal of the American Statistical Association,* 68(341), 117-130.

Ercisli, S., 2015. *Development of Enhanced Pavement Deterioration Curves.* (Masters). Virginia Polytechnic Institute and State University.

George, K. P., Rajagopal, A. S. and Lim, L. K. 1989. Models for predicting pavement deterioration. *Transportation Research Record,* (1215), 1-7.

Giummarra, G., *et al.* 2007. Establishing deterioration models for local roads in Australia. *Transportation Research Record,* (1989), 270-276.

Gulen, S., *et al.* 2001. Development of improved pavement performance prediction models for the indiana pavement management system. *Joint Transportation Research Program*, 69.
Han, D. and Do, M. 2015. Life Cycle Cost Analysis on pavement inspection intervals considering maintenance work delay. *KSCE Journal of Civil Engineering,* 19(6), 1716-1726.
Hilbe, J. M., 2011. Negative binomial regression: modeling. *Negative Binomial Regression.* 2 ed.: Cambridge University Press, 221–283.
Hong, F. and Prozzi, J. A. 2006. Estimation of Pavement Performance Deterioration Using Bayesian Approach. *Journal of Infrastructure Systems,* 12(2), 77-86.
Inc., S. C. and Inc., H. W. L., 2007. *Development of performance prediction models for Virginia Department of Transportation pavement management system.* Richmond,VA: Virginia Dept. of Transportation.
Katicha, S. W., *et al.*, 2016a. *Development of Enhanced Pavement Deterioration Curves.* Virginia Transportation Research Council Charlottesville, VA United States: Virginia, VTRC 17-R7.
Katicha, S. W., Flintsch, G. W. and Bryce, J. M. 2016b. Modeling Pavement Condition and Deterioration: Linear Empirical Bayes Approach. (17-04619).
Khattak, M. J., Baladi, G. Y. and Sun, X., 2009. *Development of index based pavement performance models for pavement management system (PMS) of LADOTD.*
Lethanh, N., Kaito, K. and Kobayashi, K. 2014. Infrastructure deterioration prediction with a Poisson hidden Markov model on time series data. *Journal of Infrastructure Systems,* 21(3), 04014051.
Loprencipe, G. and Pantuso, A. 2017. A Specified Procedure for Distress Identification and Assessment for Urban Road Surfaces Based on PCI. *Coatings,* 7(5), 65.
McGhee, K. H., 2002. *Development And Implementation Of Pavement Condition Indices For The Virginia Department Of Transportation, Phase I Flexible Pavements.* 1401 E. Broad St. Richmond, VA 23219, USA: Virginia Department of Transportation (VDOT).
McGhee, K. H., 2004. *Automated Pavement Distress Collection Techniques.* Transportation Research Board.
Mills, L. N. O., Attoh-Okine, N. O. and McNeil, S. 2012. Developing Pavement Performance Models for Delaware. *Transportation Research Record,* 2304(2304), 97-103.
Nunez, M. M. and Shahin, M. Y. 1986. Pavement condition data analysis and modeling. *Transportation Research Record,* (1070), 125-132.
PAVER, 2014. User Manual - PAVERTM Version 7.0. Champaign, Illinois: US Army Corps of Engineers Engineer Research and Development Center (ERDC) Construction Engineering Research Laboratory (CERL).
Pierce, L. M., McGovern, G. and Zimmerman, K. A., 2013. *Practical guide for quality management of pavement condition data collection.* 1200 New Jersey Avenue, SE Washington, DC USA 20590: Federal Highway Administration.
Prozzi, J. A. and Madanat, S. M. 2004. Development of Pavement Performance Models by Combining Experimental and Field Data. *Journal of Infrastructure Systems,* 10(1), 9-22.
Saha, P., Ksaibati, K. and Atadero, R. 2017. Developing Pavement Distress Deterioration Models for Pavement Management System Using Markovian Probabilistic Process. *Advances in Civil Engineering,* 2017.
Shahin, M. Y., 2005. Pavement Management for Airports, Roads, and Parking Lots. 2nd ed. New York, NY, USA: Springer Publishing Company, 15-37.
Shahin, M. Y., *et al.* 1987. New techniques for modeling pavement deterioration. *Transportation Research Record,* (1123), 40-46.
Stampley, B. E., *et al.* 1995. Pavement management information system concepts, equations, and analysis models.
Tabatabaee, N. and Ziyadi, M. 2013. Bayesian Approach to Updating Markov-Based Models for Predicting Pavement Performance. *Transportation Research Record,* (2366), 34-42.
VDOT, 2012. A Guide to Evaluating Pavement Distress Through the Use of Video Images. Version 2.6 ed. 1401 E. Broad Street, Richmond, VA 23219, USA: Virginia Department of Transportation, Maintenance Division.
Wolters, A. S. and A, Z. K., 2010. *Research of Current Practices in Pavement Performance Modeling.* Pennsylvania Department of Transportation, FHWA-PA-2010-007-080307.
Zellner, A., 1996. *An Introduction to Bayesian Inference in Econometrics.* John Wiley \& Sons Inc.

Figure 14. Comparison between CCI deterioration curves using VDOT default models and NB regression model for the Secondary system.

Figure 15. Illustration of the approaches to calculate the estimate of future pavement condition.

Figure 16. Colour map plot of the Interstate network pavement family; (a) Raw observed pavement condition data; (b) Pavement condition modelled using the LEB approach.

Figure 17. Colour map plot of the Primary network pavement family; (a) Raw observed pavement condition data; (b) Pavement condition modelled using the LEB approach.

Figure 18. Colour map plot of the Secondary network pavement family; (a) Raw observed pavement condition data; (b) Pavement condition modelled using the LEB approach.