Development of Enhanced Pavement Deterioration Curves

Safak Ercisli

Thesis submitted to the faculty of the Virginia Polytechnic Institute and State University in partial fulfillment of the requirements for the degree of

Master of Science
In
Civil Engineering

Gerardo Flintsch
Samer Katicha
Brian Diefenderfer

August 3, 2015
Blacksburg, Virginia

Keywords: pavement deterioration curve, deterioration model, performance prediction, regression, empirical Bayesian

Development of Enhanced Pavement Deterioration Curves

Safak Ercisli

ABSTRACT

Modeling pavement deterioration and predicting the pavement performance is crucial for optimum pavement network management. Currently only a few models exist that incorporate the structural capacity of the pavements into deterioration modeling. This thesis develops pavement deterioration models that take into account, along with the age of the pavement, the pavement structural condition expressed in terms of the Modified Structural Index (MSI). The research found MSI to be a significant input parameter that affects the rate of deterioration of a pavement section by using the Akaike Information Criterion (AIC). The AIC method suggests that a model that includes the MSI is at least 10^{21} times more likely to be closer to the true model than a model that does not include the MSI. The developed models display the average deterioration of pavement sections for specific ages and MSI values.

Virginia Department of Transportation (VDOT) annually collects pavement condition data on road sections with various lengths. Due to the nature of data collection practices, many biased measurements or influential outliers exist in this data. Upon the investigation of data quality and characteristics, the models were built based on filtered and cleansed data. Following the regression models, an empirical Bayesian approach was employed to reduce the variance between observed and predicted conditions and to deliver a more accurate prediction model.

ACKNOWLEDGEMENTS

I would like thank the Virginia Tech and the Virginia Tech Transportation Institute Faculty, including but not limited to, firstly my advisor Dr. Gerardo Flintsch, and Senior Research Associate Dr. Samer Katicha for all of their support and guidance. I also would like to acknowledge all the faculty, staff, and fellow students of Transportation and Infrastructure Engineering program, in the Virginia Tech Charles E. Via, Jr. Department of Civil & Environmental Engineering, for their continuous support and encouragement. In addition to Virginia Tech, I also want to thank Virginia Department of Transportation, especially to Dr. Brian Diefenderfer, for all their help and support for my research.

I also would like to thank all my friends in Blacksburg, who became my family here. This journey would have not been the same without all the amazing people that I have met along the way.

Finally, I would like to thank my family, my parents Nadiye and Cevat, my sister Hasret, and my grandmother Gulizar. Despite the greatest distances, I have always felt your support. And the greatest distances, we have always travelled them. Thank you for always being there.

Table of Contents

List of Tables

List of Figures

CHAPTER 1. INTRODUCTION

1.1 Background

Accurate pavement deterioration prediction is key for efficiently and effectively managing the allocated budget for keeping an agency's road network operated at an "optimal" serviceability level. Therefore, it is important to have accurate pavement condition prediction models.

The deterioration of pavement over time can be defined by comparing and evaluating the changes in the pavement condition, and by examining its history of serviceability. The pavement condition is commonly classified as functional or structural condition (Park et al., 2007). Functional condition measures the service provided to the road users. Functional condition indicators are generally oriented to service, user perception, safety and sufficiency. Examples include roughness, surface deterioration, friction, and macro and micro-texture of the pavement. On the other hand, the structural condition is generally not directly perceived by the users but rather indicate the physical condition and structural load-carrying capacity of the pavement (Park et al., 2007).

Although they reveal two different properties of the pavement, the two types of indicators are interrelated. Research has shown that the structural condition of the pavement directly affects the functional performance (Bryce et al., 2013). A poor structural pavement condition results in greater negative rate of change in the functional condition.

Pavement indices, such as IRI (International Roughness Index), PCR (Pavement Condition Rating), PCI (Pavement Condition Index) and CCI (Critical Condition Index) are commonly used to provide information on the overall pavement condition. However, none of them explicitly takes the structural capacity into account. The International Roughness Index (IRI) summarizes the roughness qualities of the pavement as a longitudinal profile that impact vehicle response (Sayers, 1986). The composite pavement indices, such as PCR, PCI and CCI, reflect the overall condition of the pavement by considering the composite effects of different distress types, their severity and extent.

In this study, the functional condition of the pavement is captured in terms of CCI, which is currently used by the Virginia Department of Transportation (VDOT) and uses a scale of 0-100. The CCI is calculated as the lower of the two indices: LDR (Load Related Distress Rating) and NDR (Non-load Related Distress Rating). LDR and NDR are used only for asphalt-surfaced flexible pavements. CCI value of 100 represents a pavement with no visible distress, while a pavement with a CCI value of zero indicates a pavement in heavily distressed condition. In general, pavement sections with a CCI value below 60 (poor and very poor) are considered 'deficient' and should be further evaluated for maintenance and rehabilitation actions (VDOT, 2006). **TABLE 1** summarizes the scale used by VDOT to rate pavements based on CCI.

TABLE 1 Pavement Condition Definitions. Virginia Department of Transportation, C. O., Asset Management Staff (2006). "State of Pavement - Interstate and Primary Flexible Pavements." URL: http://virginiadot.org/info/resources/2006_Condition_Report.pdf 2015.

Index Scale (CCI)	Pavement Condition	Likelihood of Corrective Action
90 and Above	Excellent	Very Unlikely
70-89	Good	Unlikely
60-69	Fair	Possibly
50-59	Poor	Likely
49 and Below	Very Poor	Very Likely

Previous work has shown that good functional condition of pavement does not necessarily indicate good structural condition (Zaghloul et al., 1998) and that there is a very weak correlation between the surface condition and the structural condition of the pavement (Flora 2009; Bryce et al. 2013). This is thought to be at least partially due to the fact that maintenance practices tend to enhance the functional parameters of the road, while the structural capacity of the pavement remains virtually unchanged. However, Bryce at al. (2013) posed that the structural condition has a significant effect on the rate of deterioration of the pavement and, the CCI change over time was affected by the pavement Modified Structural Index (MSI) values. The MSI is a structural index developed for use in network-level pavement evaluation. The pavement sections with low MSI values (in poor structural condition) deteriorated more rapidly than the pavement sections with higher MSI values (in adequate structural condition).

1.2 Problem Statement

Since the accuracy in the pavement condition prediction, with respect to the actual condition, directly influences pavement management decisions, it is important to model the performance of the pavement as accurately as possible. Recent studies have shown that pavements in poor structural condition tend to have a faster functional deterioration rate compared to pavements with adequate structural condition, when the same treatment is applied to both pavements (Zaghloul et al., 1998; Flora 2009; Bryce et al. 2013). The current deterioration models used by VDOT at the network-level were developed based on expert opinion and windshield pavement condition evaluations (Stantec, 2007). These models are maintenance category specific but do not take into account the structural capacity of the pavement as a factor. However, project-level decisions often include structural capacity as a key parameter. Incorporating structural capacity as a factor that affects pavement deterioration may help make network-level recommendations match as close as possible the project-level decisions. Lack of consistency at these levels may result in sub-optimal resource allocations and financial inefficiencies.

1.3 Objective

The objective of this thesis is to develop a new set of deterioration models for the Virginia Department of Transportation (VDOT) Pavement Management System (PMS). These models incorporate the impact of the structural condition into the deterioration rate for flexible pavement sections of VDOT. This development may help to minimize the difference between network-level and project-level treatment decisions. The new models will help VDOT achieve more accurate network-level pavement performance predictions and resource allocations.

1.4 Scope

This study is a continuation of a previously completed research study by Bryce et al. (2013) which introduced a network-level structural capacity index for VDOT. This index, MSI, can be used to link pavement deterioration rate to its structural condition.

The inclusion of the MSI of the pavement section is expected to allow better predictions of the pavement condition. This will help determine the type and timing of treatment or rehabilitation that will be required, resulting in more reliable network-level programs.

The thesis starts with a literature review about pavement deterioration models (Chapter 2) followed by an investigation of the quality of the data, identification of the data characteristics, and selection of appropriate data cleansing and filtering procedures (Chapter 3). The condition data extracted from the VDOT Pavement Management System database only includes flexible pavements on the Interstate system, as it was the only network with structural capacity data available. Chapter 4 presents the various model types, forms, and modeling approaches considered, compares the various models considered, and identifies the models deemed most appropriate for modeling the deterioration of Virginia's Interstate pavements. Chapter 5 presents the findings of the study, concludes the analysis and derives recommendations both for implementation and for future research.

1.5 Methodology

Due to many factors affecting the pavement performance, it requires a complex study to predict the future performance. The development of enhanced performance prediction models in this thesis starts with an extensive literature review. Secondly, the data to be used in this study, obtained from VDOT, is analyzed and then filtered from outliers and noise to avoid feeding the models with biased information. The knowledge regarding existing deterioration curves and applicable modeling techniques, obtained in the literature review phase, is used to create several regression models with the filtered data. The models are evaluated based on the defined model selection criteria. To improve the prediction accuracy of the overall pavement network, application of Empirical Bayesian method was also considered by completing a detailed investigation of Empirical Bayesian approach and determining a prior distribution for the data. Lastly, the Empirical Bayesian method is used to develop a model with improved predictions. The improvements done by each model are compared to find the optimal modeling solution.

1.6 Significance

Pavement performance prediction is important for managing pavement assets. Haas et al. (1994) proposed that the incorporation of a structural index into the network-level PMS database and into the condition prediction models can be helpful to determine average network structural conditions, predict deterioration behaviors, evaluate future structural inadequacies, plan for future work program, and assess future funding requirements. Implementation of enhanced deterioration curves into the pavement management system will improve the overall performance of the system by providing predictions that are more accurate, leading to higher-efficiency maintenance and rehabilitation activity programs.

The approach followed better represents the actual observed data as the negative binomial distribution better fit the measured data. The approach also accounts for error in the recorded condition as well as natural performance variation and difference between different pavement sections.

CHAPTER 2. LITERATURE REVIEW

This chapter begins with a review of existing literature about pavement performance, pavement deterioration, and the factors that affect pavement performance. Then it briefly explains the functional and structural performance of pavement, plus their indicators. It continues with a summary of the currently in-use deterioration models and delivers a discussion on the applicable modeling approaches to predict pavement deterioration.

2.1 Pavement Performance

The pavement performance should be defined differently depending on if it is used at the project, network, or strategic level (Lytton, 1987). At the project level, the distress data, decrease in serviceability index, skid resistance and damage done by traffic can define the performance. At the network level, performance is usually defined by the overall condition of the network, the condition and trends of individual projects, and the level of performance that is provided by each functional type and class of road. At the strategic level, the focus is on the overall performance of the pavement network for each geographical subdivision rather than the conditions and trends of the individual projects. This strategic level is mostly concerned with policy and economics, including cost and fund allocation, concerning equity in taxation and in delivering adequate service to citizens. The overall performance of a network is an important indicator to show the needs for funding and the effects of pavement performance on user costs.

The major purpose of monitoring pavement performance is to objectively examine and determine the current condition of pavements, as well as its historical trends, to evaluate this information for developing a management action plan. An action plan includes all planning and decision-making steps related to the maintenance, rehabilitation, construction, and reconstruction of pavements. Deterioration prediction models enhance the capability of a pavement management system since the accuracy of the predictions is the key that drives support for decision-making. All the data provided through monitoring and inventory activities are used to model the performance. Then the alternative activities are analyzed, designed, planned, compared, ranked and optimized. As a result, achieving an optimal budget and fund allocation becomes easier. **FIGURE 1** illustrates these concepts in an example of systematic framework.

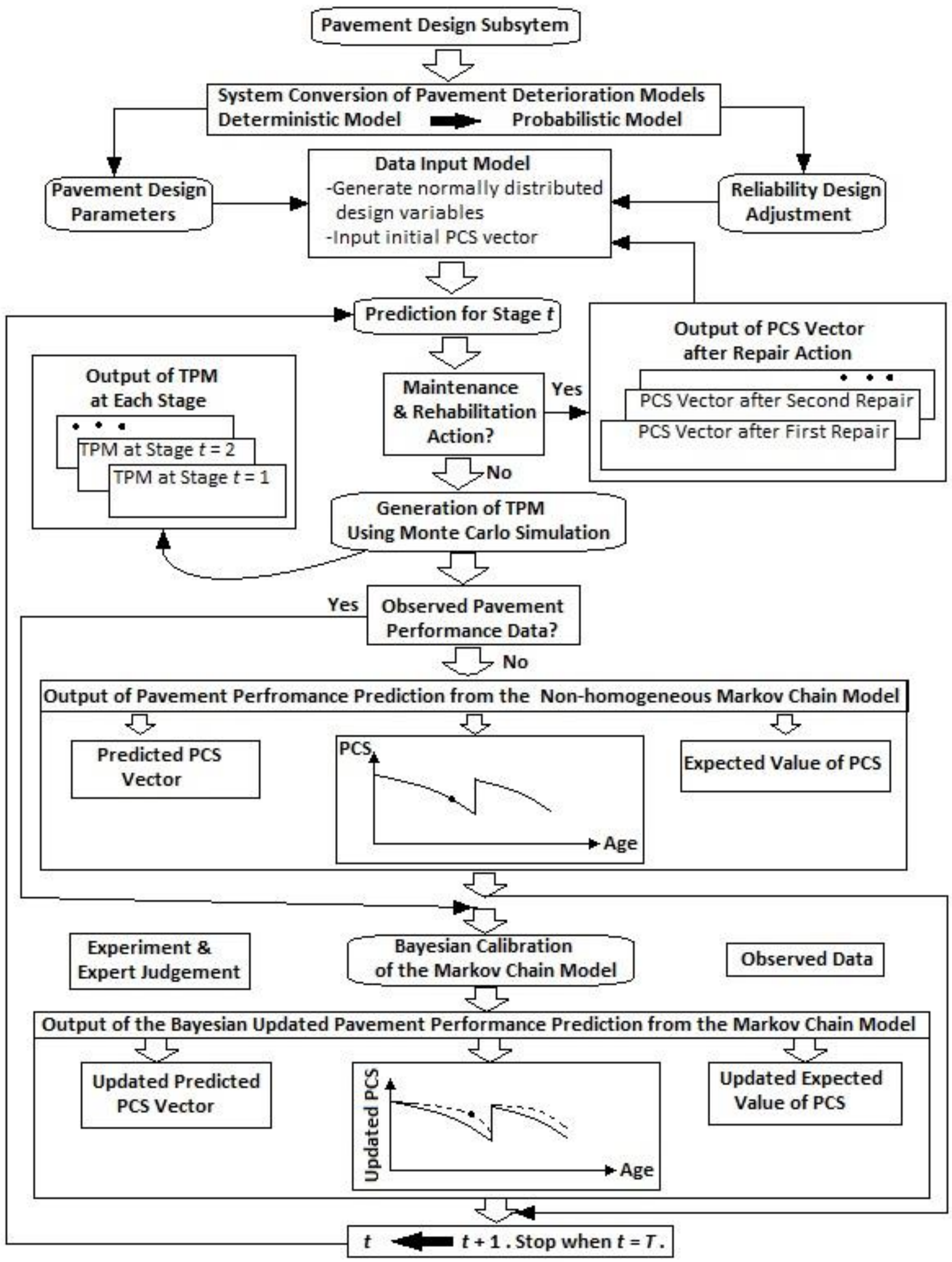


FIGURE 1 Framework for Pavement Performance Prediction. Li, N., Haas, R., and Xie, W.-C. (1997). "Development of a new asphalt pavement performance prediction model." Canadian Journal of Civil Engineering, 24(4), 547-559. Used with permission of Canadian Science Publishing, 2015.

2.2 Pavement Performance Factors

Pavement performance is defined as the ability of a pavement to satisfactorily serve traffic over time (AASHTO, 2003). **FIGURE 2** displays numerous factors that have been used in predicting pavement performance. Factors that are generally included in the pavement performance prediction models can be grouped into several categories such as traffic loading associated factors (loading variables), material properties and composition variables, environmental associated factors and construction and maintenance variables. There are also other independent factors, such as geometric features, that cannot be grouped in these categories (Huang, 1993).

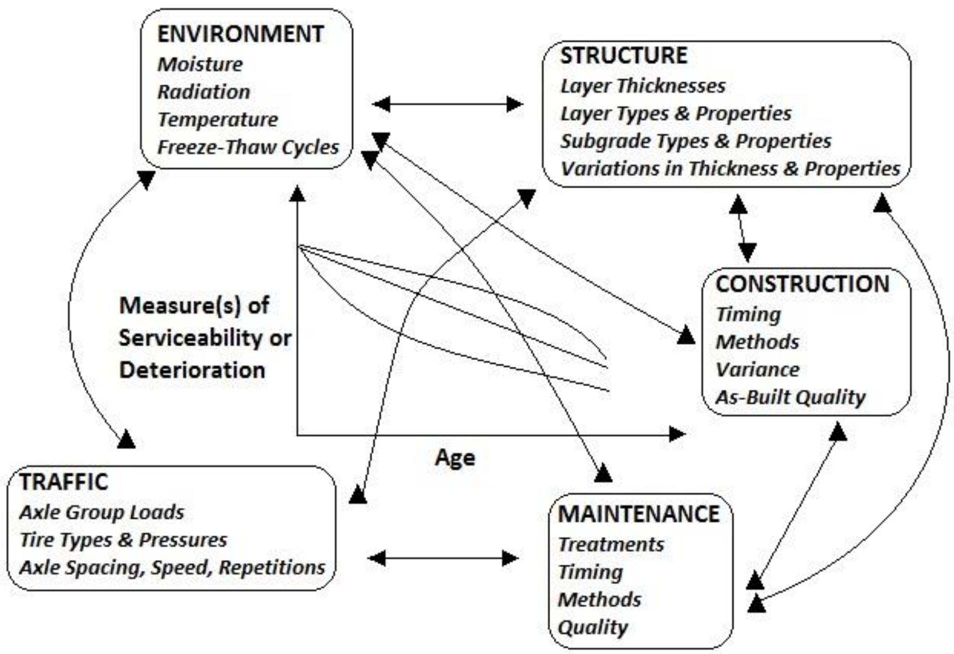


FIGURE 2 Factors and interactions that can affect pavement. Tighe, S., Haas, R., and Ponniah, J. (2003). "Life-cycle cost analysis of mitigating reflective cracking." Transportation Research Record: Journal of the Transportation Research Board(1823), 73-79, Used under fair use, 2015.

Traffic-loading associated factors have the most impact on pavement performance. These factors include traffic volume in the form of Annual Average Daily Traffic (AADT), axle load, Equivalent Single Axle Loads (ESALs), wheel load dynamic effects, truck axle types, tire pressure, design loads and overloading effects, load application time and mechanism, loading area shape and configuration, and position of the wheel in transverse section (Rada, 2013). Factors due to material properties and composition variables are the main engineering properties of the materials such as bearing capacity or strength of the material, mechanical properties in terms of durability, stability, flexibility, impermeability etc., skid resistance of the pavement, elastic and resilience modulus, viscosity, dilatancy, stiffness and Poisson ratio (Fwa, 2005). Environmental factors may show great differences depending on the geographic location. Some of the main environmental factors that affect pavement performance are temperature, moisture and humidity, precipitation, ground water characteristics, solar radiation, suction, and freeze and thaw effect (Fwa, 2005).

Construction variables are the factors that are based on the design and construction phase of the pavement and associated variability. These include layer thicknesses and variability along the section, compaction and residual stress induced by compaction during the construction, variability in gradation, asphalt and moisture content changes, size, amount, shape and location of construction joints, quality of the initial construction work. Factors such as construction quality variations are usually defined on the qualitative level, and are difficult to include in a mathematical model to predict pavement performance. Others such as layer thickness and strength can be aggregated to a single number (if an empirical design procedure is followed), which is called Structural Number (SN), to represent the overall structural strength of the multiple flexible pavement layers (AASHTO, 1993). Maintenance activities also affect the pavement performance.

Other factors that were not grouped above are usually the geometric features of the roadway such as longitudinal and transverse slopes and curvature, drainage system, maintenance activities and surface characteristics.

2.3 Pavement Deterioration Indicators

Both functional and structural performance are important to track the deterioration of pavements. Different indicators are used to characterize the pavement structural or functional condition, and composite indices provide a measure of general performance. **FIGURE 3** summarizes the different pavement condition data collected by state agencies, depending on the management level.

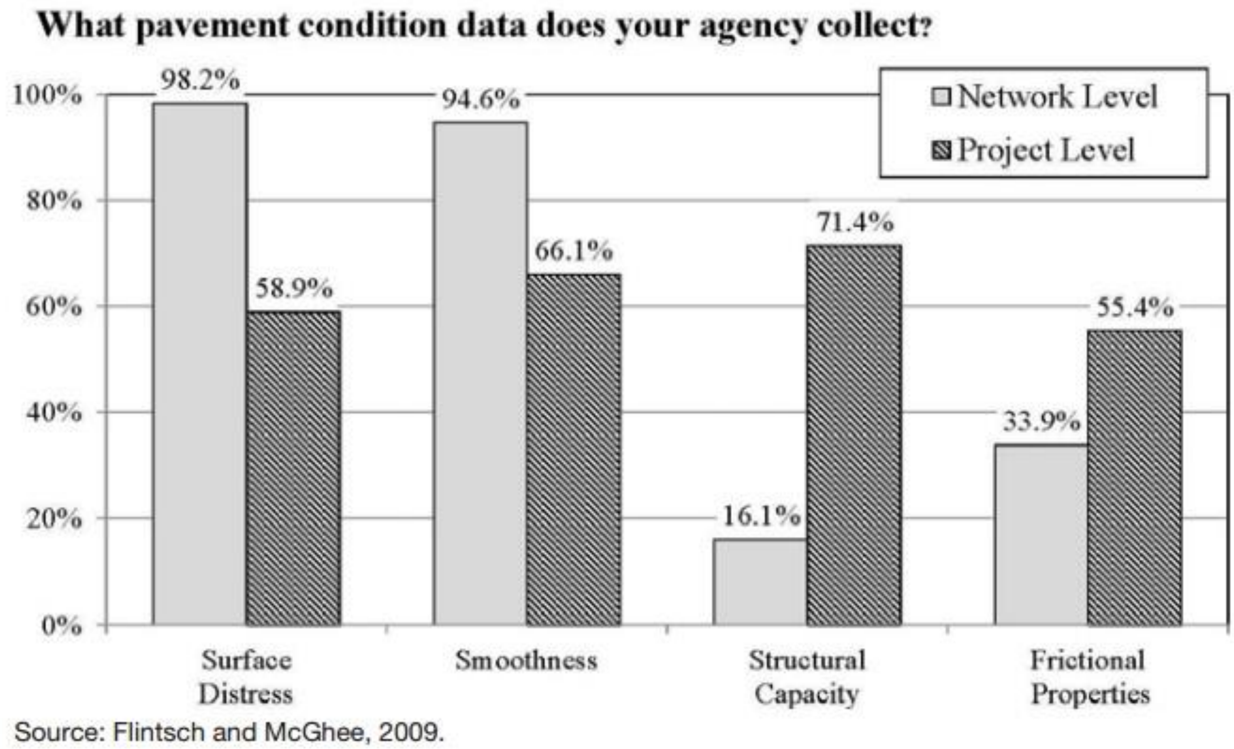


FIGURE 3 Summary of current practices for pavement data collection in North America. Flintsch, G. W., and McGhee, K. K. (2009). Quality management of pavement condition data collection, Transportation Research Board, Used under fair use, 2015.

2.4 Functional Condition Indicators on Flexible Pavements

Functional condition of flexible pavements can be evaluated according to one or several different criteria. With the lack of a universally used deterioration indicator, each agency may choose their indicator to define pavement deterioration for their management.

Functional condition is defined primarily with the following metrics: serviceability, distresses, roughness/riding quality, surface friction/macrotexture and microtexture, safety, and noise. Serviceability is a more general term frequently used to describe the user's satisfaction with the ride. Studies have shown that roughness contributes to most of a driver's perception about the serviceability of a pavement (Haas et al., 1994). Roughness defines the vertical changes in the pavement longitudinal profile, transverse profile, and cross slope. However, the longitudinal profile in the wheel path is the one main indicator that is used to describe the roughness or riding quality of a pavement.

Distresses, namely surface rutting and fatigue cracking, are the most commonly modeled performance indicator for flexible pavements (Prozzi, 2001). Bleeding, transversal cracking, longitudinal cracking, edge cracking and block cracking are some of the other distress types.

Distresses are typically measured by type, frequency, and extent. The distresses are then weighted according to its importance to the pavement condition for calculating a summary index. **TABLE 2** summarizes the different types of distresses considered in Long Term Pavement Performance (LTPP) program on asphalt concrete pavement surfaces.

TABLE 2 Asphalt Concrete Pavement Surface Distresses Considered in LTPP. Federal Highway Administration, Miller, S., Bellinger Y. (2014). "Distress Identification Manual for the Long-Term Pavement Performance Program (Fifth Revised Edition)" Report FHWA-HRT-13-092, Used under fair use, 2015.

Distress Categories/Type	
Cracking	Surface Deformation
Fatigue Cracking	Rutting
Block Cracking	Shoving
Edge Cracking	Surface Defects
Wheel Path Longitudinal Cracking	Bleeding
Non-Wheel Path Longitudinal Cracking	Polished Aggregate
Reflection Cracking at Joints	Raveling
Transverse Cracking	Miscellaneous Distress
Patching and Potholes	Water Bleeding and Pumping
Patch/ Patch Deterioration	Lane-to-Shoulder Dropoff
Potholes	

In the literature, there are many indices developed using the abovementioned indicators to reflect the overall or partial functional condition of the pavements in a summarized format. The first approach to define the functional performance was Present Serviceability Index (PSI), which was developed based on the AASHO road test (AASHO, 1962). It is defined on a scale from 0 to 5 with 0 being very poor condition and 5 being excellent. Later on, Hajek et al. (1986) presented the Pavement Condition Index (PCI), which indicates both roughness and pavement surface distresses for asphalt concrete surfaces on a scale from 0 to 100. The PCI includes further detail in data measurement and calculation compared to PSI, and delivers more accurate information about pavement condition on a wider scale. Another composite index covering both distress and roughness is the Pavement Condition Rating (PCR), a model based on point deduction for observed condition defects, and it was defined on a scale of 0 to 100. (George et al., 1989). Besides the composite indices, there are others focused on single aspects of functional condition. The International Roughness Index (IRI) is based only on the roughness of the pavement to describe the riding quality (Sayers, 1986).

The Critical Condition Index (CCI), which is used by Virginia Department of Transportation, is an index reflecting the surface-observable distresses by taking different distress types, their severity and extent into account. CCI, on a scale of 0 to 100, is determined as the lower of two rating values, Load Related Distress Rating (LDR) and Non-Load Related Distress Rating (NDR), which are calculated based on the pavement surface distresses. Usually, a CCI value lower than 60 indicates poor or very poor condition; the sections with such values are considered deficient and need action in terms of maintenance or rehabilitation treatments.

In addition to the indices explained above, many others have been developed that are not used as widely. Some of them are the Present Serviceability Rating (PSR), Profile Index (PI), cracking index, distress index, mean panel rating, overall pavement condition, overall pavement index (OPI), pavement condition evaluation system (PACES), pavement condition survey, pavement distress index (PDI), pavement quality index, etc. (Ammons, 2012). These indices show different characteristics in terms of parameters, scale, computation practices, level of detail, and output styles. They all represent the efforts made in the literature on better describing the pavement performance.

2.5 Determination and Expression of Structural Capacity of Flexible Pavements

Structural capacity of a pavement is another indication of its performance and generally determines its capability to deal with traffic loading and environmental factors that influence the deterioration. Structural capacity evaluation of pavement is important and is in use especially to plan maintenance and rehabilitation activities or to decide on loading restrictions in extreme climate conditions or areas. A recent study by Bryce (2013) confirmed that incorporation of structural condition in network-level pavement management decisions allow these decisions to match better with the ones made during project-level assessment than those based only on functional condition and surface distress. As the functional properties of a pavement alone do not provide enough information to describe the overall pavement condition, consideration of the structural condition becomes necessary in network-level pavement management decisions.

Measurement of structural capacity can be done through destructive and non-destructive methods. Non-destructive surveying methods, such as deriving the pavement thickness through Ground Penetrating Radar (GPR), and using high-speed deflectometers are emerging applications that can help improving data collection practices for pavement structural capacity. The improvement in taking measurements at high speeds has started to allow agencies to improve their data collection practices and PMS databases significantly (Saarenketo and Scullion, 2000).

A major indicator of pavement structural condition is the deflection. Deflection measurements are used as inputs in various pavement condition assessment tools, including structural capacity prediction models to calculate the remaining service life of pavements (Gedafa et al., 2010a). The Falling Weight Deflectometer (FWD) is currently the most prevalent device used to measure pavement deflections (Hadidi and Gucunski, 2010). The FWD device applies an impulse load to the pavement surface on a circular plate and then measures surface deflections through sensors located at the loading center and at fixed radii from the loading center (Noureldin et al., 2003). The measured deflections plot a deflection basin (or bowl) as seen in **FIGURE 4** to reflect the overall structural capacity of the pavement.

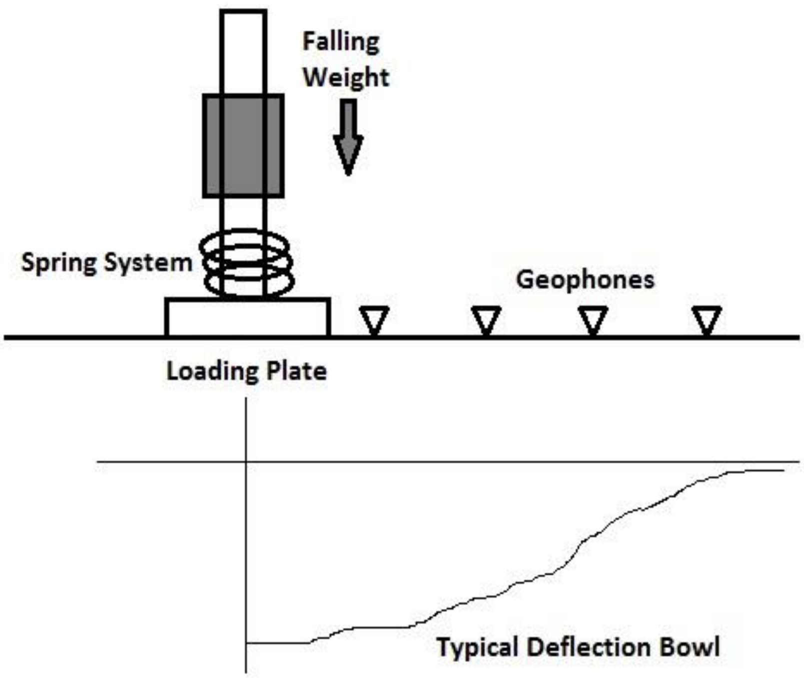


FIGURE 4 Typical deflection basin (bowl) obtained from a FWD loading. Saltan, M., and Terzi, S. (2008). "Modeling deflection basin using artificial neural networks with cross-validation technique in backcalculating flexible pavement layer moduli." Advances in Engineering Software, 39(7), 588-592, Used under fair use, 2015.

In addition, continuous deflection devices exist that measure the pavement deflection and collect data while constantly moving. The Rolling Wheel Deflectometer (RWD) and Traffic Speed Deflectometer (TSD) are the two most promising devices to deliver the information needed by highway agencies (Flintsch et al., 2013). TSD is an articulated truck that uses four Doppler lasers on a servo-hydraulic beam to measure and record the deflection velocity of a loaded pavement. TSD technology allows data collection at speeds as high as 45 mph (70 km/h). RWD measures the deflection by comparing the loaded and unloaded condition records of pavement surface, which are measured with two different sets of lasers on the same prone truck (Flintsch et al., 2013).

There are various indices to summarize pavement structural capacity. One of them is the Structural Strength Indicator (SSI). SSI uses the center deflection measurements obtained from FWD tests to develop a function based on cumulative distributions of deflections for a given pavement family. The SSI function is based on the distribution equation in Equation 1 and determined as in the form of Equation 2 (Flora, 2009).

$$SSI = 100 * \left[1 - F\left[\left(\delta_{ijk}\right)_1\right]\right] \quad (1)$$

Where $F\left[\left(\delta_{ijk}\right)_1\right]$ is the cumulative probability distribution of $\left(\delta_{ijk}\right)_1$, the deflection measured at the center sensor.

$$SSI_{jk} = \left(1 - ae^{\frac{\beta}{(\delta_1)^{\gamma}}}\right) \quad (2)$$

Where j, k denotes the pavement family, and α, β, and γ are the coefficients.

α, β, and γ are found by minimizing errors between Eq. 1 and Eq. 2

Another structural capacity indicator, the Structural Capacity Index (SCI), was developed by the Texas Department of Transportation to be used as a network level index. The inputs used in the calculation are the existing Structural Number (SN), estimated from FWD testing results and layer thicknesses, and the required structural number. SCI is the ratio of the effective SN (existing) to the design SN (required). To put simply, an SCI value greater than one signifies that the pavement is in sound structural condition for design traffic (Zhang et al., 2003).

The Kansas Department of Transportation and Kansas State University prepared a way to determine structural capacity by using a set of regression equations that estimates the remaining service life (RSL) of a pavement, with FWD data as the input and sigmoidal curves as the model (Gedafa et al., 2010). The equations can be seen below:

$$RSL = \delta + \frac{\alpha}{1 + e^{\beta - \gamma d_0}} \quad (3)$$

Where;

$$\delta = \delta_0 + \delta_1 D + \delta_2 EAL + \delta_3 ETCR + \delta_4 EFCR + \delta_5 Rut + \delta_6 SN_{eff} \quad (4)$$

$$a = a_0 + a_1 D + a_2 EAL + a_3 ETCR + a_4 EFCR + a_5 Rut + a_6 SN_{eff} \quad (5)$$

$$\beta = \beta_0 + \beta_1 D + \beta_2 EAL + \beta_3 ETCR + \beta_4 EFCR + \beta_5 Rut + \beta_6 SN_{eff} \quad (6)$$

$$\gamma = \gamma_0 + \gamma_1 D + \gamma_2 EAL + \gamma_3 ETCR + \gamma_4 EFCR + \gamma_5 Rut + \gamma_6 SN_{eff} \quad (7)$$

The factors δ_n, α_n, β_n and γ_n (for n=1,2,3,4,5& 6) are constants derived from regression models for different pavement types. EAL is the Equivalent Axle Load per day, ETCR is the Equivalent Transverse Cracks, EFCR is the Equivalent Fatigue Cracking, Rut is the rut depth in inches and SN_{eff} is the effective structural number of the pavement.

The Modified Structural Index (MSI), developed by Bryce in 2013, is a network-level structural capacity index that minimizes errors between network-level predictions and project-level work done on the flexible pavements of the Virginia Department of Transportation. The equation is based on FWD data, traffic volume and resilient modulus. It is modified from the SCI of the Texas Department of Transportation, and is finalized with three different equations depending on the road classification as the interstates, divided primary roads, un-divided primary roads,and high-volume secondary roads. The final form of the MSI for bituminous interstate pavements is as follows (Bryce et al., 2013):

$$MSI = \frac{0.4728*(D_0 - D_{1.5Hp})^{-0.4810} * Hp^{0.7581}}{0.05716*(\log(ESAL) - 2.32*\log(M_R) + 9.07605)^{2.3677}} \quad (8)$$

Where D_0 = Peak Deflection under the 9,000 lbs load
$D_{1.5Hp}$ = Deflection at 1.5 times the pavement depth
ESAL= Cumulative Equivalent 18-kip Single Axle Load
M_R = Resilient Modulus

The MSI is calculated for each location separately by including only the values of the parameters at that specific location. Thus, as an index without limiting values, the MSI delivers the absolute condition of the pavement independent from the condition relative to the other locations in the pavement network. On the other hand, lack of a certain scale prevents the MSI matching with other index scales such as condition ratings used in the pavement management system. However, a set of thresholds were developed by Bryce (2013) to solve this issue and to facilitate the incorporation of MSI into scaled condition indices.

2.6 Existing Deterioration Models

Because highway agencies control large pavement networks, it is very important to have a systematic approach to monitor these assets and keep the entire network at a serviceable condition. The systematic approach ensures accurate performance modeling, resource optimization and strategic planning. In this framework, it is crucial to have the capability to predict the future condition of pavement assets.

Despite the continuous efforts that have been made in the pavement engineering field, improvements are still needed to predict pavement life more accurately (Molenaar, 2003). Aside from the sheer number of factors influencing pavement performance, the unpredictability of these factors, such as unusual weather conditions, unanticipated traffic volumes, variations in pavement structure etc., makes it difficult to reach accurate deterioration models. Most of the existing deterioration models are based on various assumptions, involve simplifications, and have several limitations.

Several classifications have been used over the years. The most common classification recognizes two basic types of pavement performance prediction models: deterministic and probabilistic. While the deterministic models give a discrete result as the prediction, probabilistic models deliver a distribution of possible results (Lytton, 1987). Later on, Tighe and Haas (2003) grouped the many deterioration models into three classes: empirical, mechanistic-empirical and subjective/experience based models.

Deterministic-based models cover a wide range of traditional models such as primary response models, structural performance models, functional performance models, damage models and load equivalence factors. Subdivisions of probabilistic models are survivor curves, Markov models and Semi-Markov models (Lytton, 1987). Pavement management level is a crucial factor to select the right model, as each of them requires a different level of detail for the inputs and a different level of effort for data collection, processing and computation. Lytton (1987) summarized the types of performance models according to which pavement management level the models are used in, as shown in TABLE 3.

Empirical models generally relate measured or predicted parameters such as deflection, traffic, etc., to the pavement age and to a deterioration indicator (e.g. loss of serviceability) by using regression analysis. Deterministic regression models and stochastic Markov chains are the most common empirical models. Models that use soft computing such as artificial neural networks, fuzzy logic, and neuro-fuzzy systems are also considered empirical models.

Mechanistic models deliver pavement responses, such as calculated stress and strain attributes that can be used as parameters for the empirical models. However, a merely mechanistic performance prediction model has not been developed yet. Mechanistic-empirical models calibrate the mechanistic factors, which are the physical causes of stresses in pavement structure, with the observed performance. Subjective or experience-based models usually estimate deterioration versus age with a combination of different variables and procedure applications, such as Markov transition matrices or Bayesian models. Pavement performance prediction models can also be classified as having continuous or discrete variables.

TABLE 3 Pavement Management Levels where Performance Models are Used. Lytton, R. L. "Concepts of pavement performance prediction and modeling." Proc., 2nd North American Conference on Managing Pavements, Used under fair use, 2015.

Levels of Pavement Management	Types of Performance Models						
	Deterministic Models				Probabilistic Models		
	Primary Response	Structural	Functional	Damage	Survivor Curves	Transition Process Models	
	Deflection, Stress, Strain, Temperature, Thermal Stress, Moisture, Energy Frozen and Unfrozen, Water Content	Distress, Pavement Condition Index	Serviceability Index, Skid Loss, Wet Weather Safety Index	Load Equivalence, Marginal Load Equivalence		Markov	Semi-Markov
National Network				✓	✓	✓	✓
State/Provincial Network		✓	✓	✓	✓	✓	✓
District Network		✓	✓	✓	✓	✓	✓
Project	✓	✓	✓	✓			

The following sections present various deterioration models currently in use. Because this study focuses on deterministic prediction models, the following research primarily covers deterministic models.

Present Serviceability Rating Model

A model generated to predict the PSR, by using various input variables, is seen below (Lee et al., 1993):

$$log_{10}(4.5 - PSR) = 1.1550 - 1.8720 * log_{10}SN + 0.3499 * log_{10}AGE + 0.3385 * log_{10}CESAL \quad (9)$$

Where SN = Structural Number
AGE = age of pavement since construction or major rehabilitation (years), and
CESAL = cumulative 18-kip ESALs applied to pavement in heaviest traffic lane (millions)

The Structural Number is an index that accounts for roadbed soil conditions, pavement layer thickness, and layer properties. Inclusion of a pavement's Structural Number allows the model to consider the pavement structural condition and properties in addition to the pavement age and the carried traffic load, while predicting the functional condition in the future.

Mississippi PCR Models

PCR, as a composite index developed for Mississippi Department of Transportation (MDOT), defines the performance in terms of roughness and distresses (George, 2000). There are different PCR model equations to predict the pavement deterioration depending on if there is an asphalt overlay, no overlay, or if the pavement is composite. Equation 10 represents the model with an asphalt overlay:

$$PCR = 90 - a[\exp(AGE^b) - 1].\log\left[\frac{ESAL}{MSN^c}\right] \quad (10)$$

Where AGE = time in years since last construction
ESAL = yearly 18-kip single axle load
MSN = modified structural number for subgrade support, and
a, b, c = regression constants

While predicting a composite index, this model also considers the structural capacity with the input of the Modified Structural Number that partially covers the inclusion of subgrade support.

Distress Maintenance Rating Model

Distress Maintenance Rating (DMR) is another composite index revealing the distress intensity and frequency on the pavement surface (Sadek et al., 1996). The non-linear equation to predict DMR in the future is as follows:

$$DMR = 100 - 5.06 * AGE^{0.48} YESAL^{1.29} DEPTH^{-0.20} \quad (11)$$

Where AGE = pavement age since last overlay (years)
YESAL = average yearly ESALs (millions), and
DEPTH = thickness of last overlay (cm)

As a model that dates back to 1996, the structural condition is not considered thoroughly in this equation. The only related independent variable is the thickness of the last overlay. Furthermore, the subgrade soil characteristic is not included in this prediction model.

2.7 Applicable Modeling Approaches

As listed and explained in the previous section, many deterioration models are currently in use that attempt to deliver the best estimations for future pavement performance. Even though there have been many studies on the subject, accurate and precise prediction of pavement life is still not possible (Molenaar, 2003). This fact indicates the need for innovative approaches to improve the predictions.

Due to the nature of the process, which involves a high level of subjectivity and uncertainty, soft computing techniques such as neural networks, fuzzy logic systems, genetic algorithms, and neurofuzzy systems are increasingly used over the traditional regression methods (Flintsch and Chen, 2004). In addition to these techniques, another improvement method for the prediction accuracy is the Bayesian regression, which is based on incorporating existing knowledge and observed data with the predictions (Zellner, 1971). On the other hand, models developed through traditional regression analysis are still widely used for pavement performance prediction and have potential to satisfy the accurate model criteria. In addition to investigating regular models such as exponential, sigmoidal and logistic function, this study also covers suitability and adoption of several growth curves commonly used outside the pavement industry.

Probabilistic models deliver the prediction in the form of a probability distribution. The application of such models requires the determination of the probability distribution of the pavement condition (Ortiz-Garcia et al., 2006). As in many studies of discrete outcomes, the sampling distribution of pavement data often results in an over dispersion than would be expected from a Poisson distribution where the variance occurs higher than the mean value. Therefore, it is necessary to use models that accommodate the over-dispersion (Byers, 2003). To this end, negative binomial distribution was also examined in this study as a part of applicable modeling approaches.

Artificial Neural Networks

Artificial Neural Networks (ANN) modeling is based on the structure of a human brain in that it can learn when fed a range of examples and can deliver valid answers accordingly from a noisy dataset (Zhang et al., 1998). They are capable of considering variations in the observed data, which is not possible by traditional regression methods. These models are commonly used in a wide range of modeling practices such as process monitoring, fall diagnosis, natural events and artificial intelligence (Dimitrova, 1996). Attoh-Okine (1995a), amongst others, applied ANN in predicting the performance of pavements.

Fuzzy Logic

Fuzzy logic was introduced by Zadeh (1965) to be used in modeling. It consists of an extended set of conventional (Boolean) logic combining fuzzy qualitative and partial linguistic variables to create truth-values such as true, false, not true, very true, quite true, not very true and not very false etc. It has been used with different applications in pavement engineering such as treatment prediction (Kaur and Pulugurta, 2008), investigation of fatigue behavior of asphalt concrete pavements (Tigdemir et al., 2002), and modeling of deflection behavior against dynamic loading in flexible pavements (Saltan et al., 2007).

Genetic Algorithms

Genetic algorithm is a deterioration modeling approach inspired by Darwin's theory of evolution (Shekharan, 2000). Chromosomes, which are called the population, represent sets of solutions and each set is comprised of genes or alleles. The alleles take numeric values to represent the parameters of pavement deterioration. After the alleles are processed through three genetic operators- reproduction, crossover and mutation- the solution from the genetic algorithm delivers the improved deterioration model.

One of the main advantages of adopting the genetic algorithm is its ability to reach global maximum-or-minimum values. Secondly, it does not require specification of derivatives, which allows the direct use of the payoff or subjective function.

Empirical Bayesian (EB) Approach

The Bayesian theory is a modeling technique for pavement management by improving predictions. It combines both objective and subjective data. This approach enables the incorporation of existing knowledge into the prediction so that previous experience can be used rather than ignored (Zellner, 1971). Models are developed by regression analysis, but each of the variables used in the model is described in terms of a probability distribution (AASHTO, 2012). On the other hand, requiring the prior distribution to be stated before including the observations is found too exacting to be realistic (Belsley, 1991).

The Empirical Bayesian method, which is based on estimating the prior distribution from the data, has not been used for pavements yet. The Empirical Bayes approach can deliver better estimates than the actual measurement even if the variance error is overestimated to some degree. Hartigan (1969) and Efron (1973) stated that the linear Bayes estimators improve the estimate of the condition versus consideration of the measurement alone, regardless of the true distribution of data or the true distribution of the error in the measurement. Application of the empirical Bayesian allows updating of the model's posterior probability results when additional observed data become available.

This method can be useful for agencies that have recently implemented a new pavement management system, or that lack reliable historical data. Additionally, it may help to overcome the influence of low-quality data in the system (AASHTO, 2012).

General Equations for Models based on Regression Analysis

Nonlinear regression models can be grouped, according to their basic behavior, into families such as exponential models, power models, sigmoid models etc.(Ratkowsky and Giles, 1989). Note that most of the existing deterioration models, mentioned in the previous section, use the forms of regressions models listed below in their general form.

Exponential Growth

The exponential growth model is used when the growth rate of a mathematical function's value is proportional to the function's current value. It is used in a wide range of fields such as biology, physics, engineering, economics, and computer science. The most common form of an exponential growth curve is as follows:

$$y = a - (a - d) * e^{-st} \tag{14}$$

Where
a = Upper asymptote
d = Lower asymptote
t = Time
m = Time of maximum growth
s = Growth rate

Sigmoidal

The sigmoidal model has been preferred very frequently in pavement deterioration modeling because of its flexibility to fit the boundary conditions and incorporate parameters. Numerous studies in many application fields have resulted in the discovery, reinvention, and adaption of nonlinear S-shaped curves, which brought its various forms in to literature. This forms include the Logistic curve, Verhulst-Pearl equation, Pearl curve, Richard's curve (Generalized Logistic), Growth curve, Gompertz curve, S-curve, S-shaped pattern, Saturation curve, Sigmoid curve, Weibull curve, Foster's curve, Bass model, and many others (Rowe et al., 2009).

Below is a simple S-curve equation:

$$y = a * \left[\frac{t^b}{c+t^b}\right] \tag{15}$$

Where a, b and c are the regression coefficients.

Logistic

Logistic function is one of the most common equations in modeling, and is used in many different fields. The disadvantage of the model is that it is capable of computing "t" over a small range of real numbers. The simple logistic function can be defined as the equation below:

$$y = \frac{1}{1+e^{-t}} \tag{16}$$

The cumulative distribution function of the continuous logistic probability distribution is the logistic function, seen as follows:

$$y = \frac{a}{1+e^{-\frac{t-m}{s}}} + d \tag{17}$$

Where
a = Upper asymptote
d = Lower asymptote
t = Time
m = Time of maximum growth
s = Growth rate

Weibull

Weibull (1951) described a non-symmetric sigmoidal model as a continuous statistical probability distribution, which is widely used in modeling survival rates. The cumulative distribution function for Weibull to be used in modeling can be described as follows:

$$y = a - (a-d) * e^{(-(st)^m)} \tag{18}$$

Where
a = Upper asymptote
d = Lower asymptote
t = Time
m = Parameter that controls the x-ordinate for the point of inflection
s = Growth rate

It should be noted that, when the parameter "m" equals one, the Weibull equation is basically an exponential growth curve.

Gompertz

Gompertz (1825) suggested a sigmoid function as a type of mathematical model for time series, where the growth is slowest at the start and end of a time period. The equation is widely used in biology and medicine to define aging, or spreading of cancer cells, or in demographics to describe population in a confined space, birth rates etc (Rowe et al., 2008). The basic equation of a Gompertz curve is as seen below:

$$y = ae^{-be^{-ct}} + d \qquad (19)$$

Where a = Upper asymptote
d = Lower asymptote
t = Time
b,c = Positive coefficients (c sets the growth rate)

Richards

Richards (1959) developed a flexible sigmoid function, which is also known as the generalized logistic curve. It is commonly used for growth modelling and easily fits various S-shaped curves (Mubareki and Sallam, 2014). The general representation of a Richards' curve is as seen below:

$$y = d + \frac{a-d}{(1+\lambda e^{-st})^{1/m}} \qquad (20)$$

Where a = Upper asymptote
d = Lower asymptote
t = Time
m = Sets asymptote near which maximum growth
s = Growth rate
λ = Related to initial y value

Negative Binomial

Negative binomial is an extension of the Poisson series that allows the expected σ^2 to be different than the mean, μ, the parameter of the Poisson distribution. Hence, this generalization of Poisson allows the mean and variance to be different (over-dispersion) by including a disturbance or error term (Byers et al., 2003). It is widely used in cases of over-dispersion and frequent-zero counts when linear models lack the distributional properties to adequately describe data and Poisson distribution cannot account for the over-dispersion (Poch and Mannering, 1996). The range of its applications includes driving accidents, neurologic lesions, leukocytes, healthcare utilization, and counts of rare animals (Byers et al., 2003). The probability mass function of negative binomial distribution is as follows:

$$f(k; r, p) \equiv \Pr(X = k) = \binom{k + r - 1}{k} p^k (1 - p)^r \qquad for\ k = 0,1,2, \dots \qquad (21)$$

2.8 Criteria to Select a Good Model

The general concepts that are involved in the model selection are model basics, necessary input information, mathematical configuration, goodness-of-fit statistics, consideration of maintenance and rehabilitation interventions, model limitations, and boundary conditions.

Several methods exist to assess the precision and accuracy of the regression models. The standard error of estimate, the coefficient of determination, the residual analysis, correlation coefficient, and F-test are some of the commonly used regression diagnostics (Draper and Smith 1981; Smith and Rose, 1995). However, coefficient of determination (R2) is not always as reliable of a parameter to measure the goodness-of-fit for non-linear analysis as it is for linear regression analysis (Tran and Hall, 2005). The reasonableness, appearance of forecast models, ease of use and number of coefficients in the model are the qualitative criteria used in this study to consider the goodness of a fit.

Good database records of constructions, maintenance and rehabilitation activities and measurements is one factor that can easily increase the reliability of a model. Additionally, inclusion of more significant variables, calibration with observations and improved model understanding are factors that can increase the accuracy of pavement performance prediction models.

2.9 Summary of the Literature Review

Pavement performance, which consequently shows the success or failure of any pavement asset, is defined differently depending on the level of decisions in which it is used. While project-level decisions require more detailed numeric results like skid resistance or distress data, at higher levels, namely strategic- or network-level, the performance is generally stated by overall performance indicators.

Pavement performance monitoring and prediction allows determination of the current network condition as well as development of long and short-term management plans, including all the maintenance, rehabilitation, construction and reconstruction activities. Hence, budget and resource allocation is optimized.

Many factors affecting the performance of pavement can be grouped into four categories: traffic loading associated factors, material properties and composition variables, environmental factors, and construction variables. There are also independent factors that influence the performance, such as geometric features and maintenance activities.

Functional and structural indicators exist that reflect the deterioration of pavement. The scaled expression of the deterioration can also be based on functional and/or structural condition. Therefore, many various indices to evaluate and monitor pavement deterioration have been developed using only functional or structural condition indicators, or by combining both.

Structural capacity of pavement has proved to have an effect on deterioration rate. The measurement can be done with several techniques; however, the most common methods are Falling Weight Deflectometer, Rolling Wheel Deflectometer and Traffic Speed Deflectometer.

Deterioration models are crucial in pavement management since accurate performance prediction enables agencies to optimize their budget with better planning and scheduling. To this end, great effort has been put into developing deterioration models in the pavement engineering field. Still, there is need for improvement in terms of accuracy and practicality. The number of deterioration models that include the structural condition is somewhat limited. On the other hand, several modeling techniques have recently been used to improve the accuracy and to facilitate the computation (e.g. Empirical Bayesian, Artificial Neural Networks, Fuzzy Logic systems and genetic algorithms). Additionally, various modeling approaches and developed equations have shown potential to be applicable in pavement deterioration modeling.

To decide on the quality and competence of a model, many factors should be considered. Input requirements, mathematical configuration, data collection needs, reasonableness, appearance of forecast model, ease of use, goodness-of-fit statistics, and ability to satisfy boundary conditions and key points of the deterioration behavior are some of the criteria to evaluate the ability of a model.

CHAPTER 3. DATA COLLECTION AND PROCESSING

This chapter examines the available data including the collection practices, the data characteristics and quality, the strength and limitations of the dataset and its possible impacts on the model to be developed. This examination is followed by data processing upon the results of these analyses. Data processing comprises of the selection and application of suitable data reduction, cleansing and filtering procedures.

3.1 Data Collection

The pavement surface condition data used in the project was extracted from VDOT's Pavement Management System (PMS) database. The average CCI for each year of measurement along with the age of the pavement section at the measurement year was obtained for interstate pavement sections where structural condition information was available (flexible pavement sections only). The average MSI value for the pavement section was computed for each section following the methodology proposed by (Bryce et al., 2013). A constant structural index value was assigned to each PMS section for the entire analysis period, 2007 to 2012.

The database used for this thesis was developed by combining the MSI, CCI, LDR, NDR data with pavement age and section information. Pavement condition and maintenance history along with FWD data were available for most of the interstate network (2,185.8 miles) of flexible pavements under the management of VDOT.

Data points with an MSI value of 3 or higher were removed since such a high value is not realistic for any pavement (Bryce et al., 2013). This filtering resulted in the removal of just one section from the database, which included 4 measurements. The final compiled database included 3,465 data points, which were grouped into 933 PMS sections of varying lengths. The average length for a section is 2.34 miles; however, the range varies from 0.05 miles to 11.88 miles, with a standard deviation of 2.07 miles. The minimum, maximum, mean values, and the standard deviation for each variable are summarized in TABLE 4.

TABLE 4 Summary Pavement Data

	Min	Max	Mean	Median	Std. Dev.
MSI	0.40	2.88	1.16	1.07	0.39
Age	1	26	7.6	7	4.2
CCI	5	100	76.42	82	19.03

The distribution of the age of the sections at the year of measurement is presented in FIGURE 5. The plot shows that there is a concentration of sections within the age interval of 5 to 10 years, with 7 being the most common age. Approximately three-quarters of the sections (76%) are younger than 11 years and there are very few sections older than 20 years. Some of the resulting older sections are thought to be due to missing recorded treatments. VDOT is currently working on reviewing this information since the erroneous ages have been identified as a critical factor.

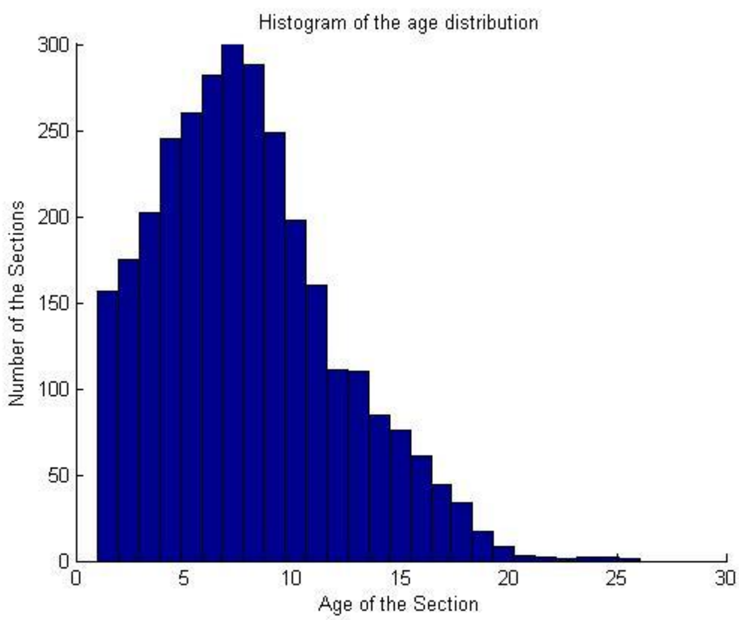


FIGURE 5 Distribution of Pavement Ages.

FIGURE 6 shows the distribution of MSI values and FIGURE 7 shows the distribution of CCI values for the dataset used. The MSI mode corresponds to values between 1.0 and 1.1 and 37.9% of the sections have an MSI value lower than 1. Only a few sections have an MSI higher than

2.0. FIGURE 7 shows that most of the CCI values are at the high end of the scale (e.g., larger than 60). This is expected as the roadway agencies try to keep their roads at an acceptable level and apply the necessary treatments when the roadway reaches the desired condition. Therefore, the number of measurements showing low values is expected to be low. More than half (52%) of the measurement showed a CCI value higher than 80 points, while only 11% had a CCI value lower than 50.

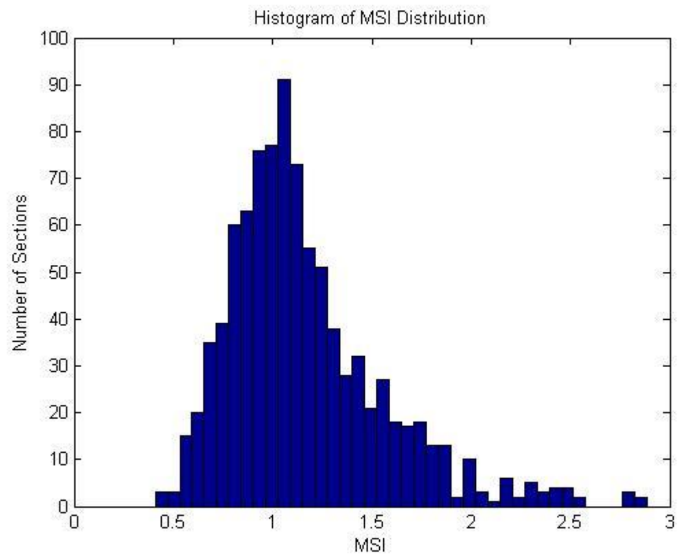


FIGURE 6 MSI Distribution.

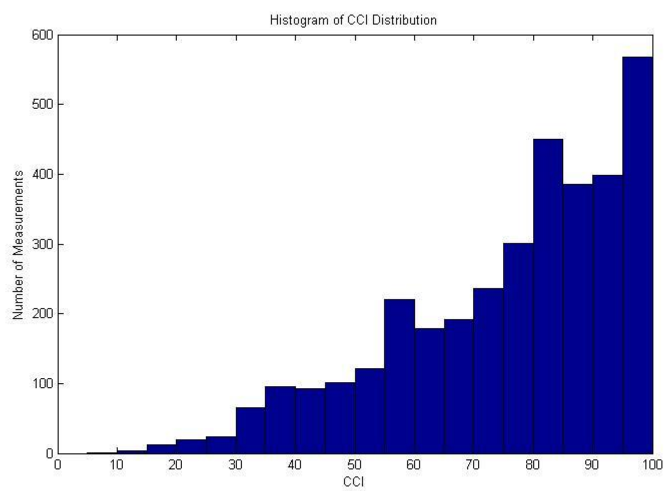


FIGURE 7 CCI Distribution.

The complement of the CCI, as defined in the Equation 22, is introduced as the Deterioration Index (DI) to allow a more comprehensive and flexible preliminary analysis. The distribution of DI, shown in FIGURE 8, shows that a vast majority of the measurements has a deterioration index lower than 50 points.

$$DI = 100 - CCI \qquad (22)$$

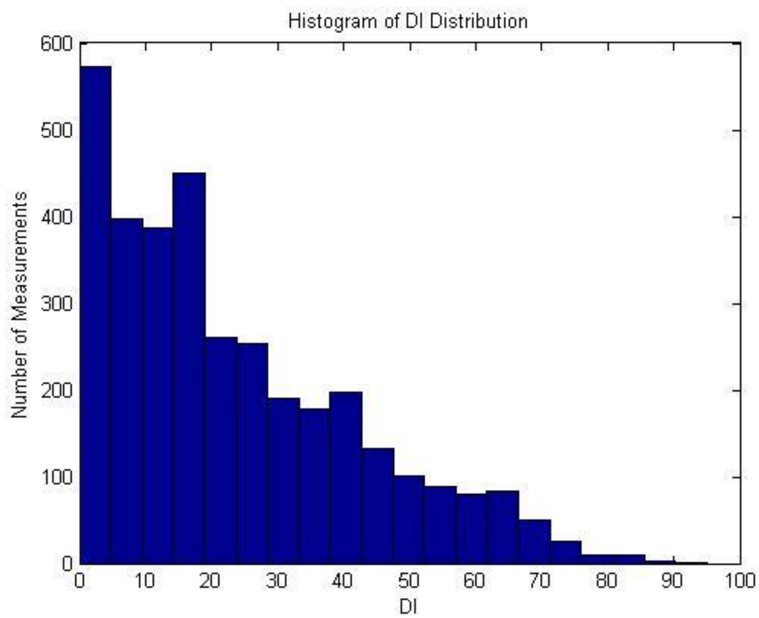


FIGURE 8 DI Distribution.

3.2 Data Filtering

The primary objective of the data analysis is to quantify the impact of the MSI value on the change of a section's functional condition over the years. To obtain the best possible models, it is important to have reliable data and remove outliers and questionable data. FIGURE 9 shows all the *CCI* values as a function of pavement age. The mean of each age group and the two plus and minus standard deviation range are also shown. Clearly, the two-standard-deviation range (which is used for the 95% confidence interval of normally distributed data [1.96 is the correct range but is often taken as 2 for simplicity]) does not give an adequate representation of the data as it extends beyond the 100 limit on the *CCI*. This is to be expected as FIGURE 7 shows that the CCI values have a very non-symmetrical distribution and are not normally distributed.

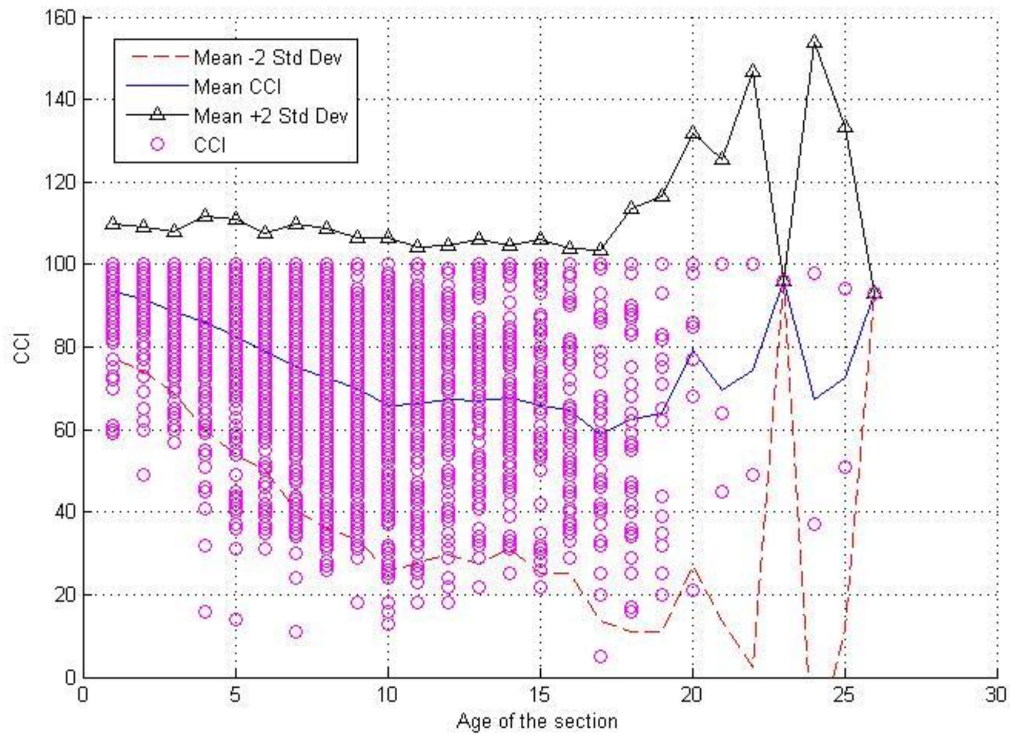


FIGURE 9 CCI vs. pavement age with mean and mean +/- 2 standard deviation lines.

From years 1 to 10, the average *CCI* decreases with increasing age as expected. However, from year 11 onwards, the average *CCI* practically stays at the same level as year 10 (after year 16, the *CCI* varies significantly due to the limited data). This contradicts engineering common sense that pavements will continue to deteriorate with time. A possible explanation in this case is that poorly performing pavements are treated before reaching year 11, and that sections that are treated are not updated in the PMS. The average *CCI* for pavements 11 years and older is a biased representation of the performance of all pavements as only good performing pavements are allowed to reach that age. To reduce the biasing effect that pavement treatment has on model estimation, the regression model was fitted to data from observations of pavement sections that had the last treatment performed less than 10 years prior to the observation, which actually covers 76% of the original data. Furthermore, it should be noted that the information about the age of the pavement is not very reliable, especially for the very old pavements, because of missing applied treatments.

Another possible reason for the very high dispersion of the data is the actual variability in the measurement and location within a section in each year's measurement. The measured spots may not coincide exactly for each year and the measured value may vary because of possible different characteristics within a section due to the construction practices, material misapplication, weather or environmental impacts, etc. To explore the influence of short sections, FIGURE 10 shows CCI differences in two consecutive measurements in every section versus the length of the section. The length of sections varies within a very large range, therefore estimating the error in measurement for each section becomes very difficult. The figure seems to indicate that the year-to-year variability becomes more stable for longer sections.

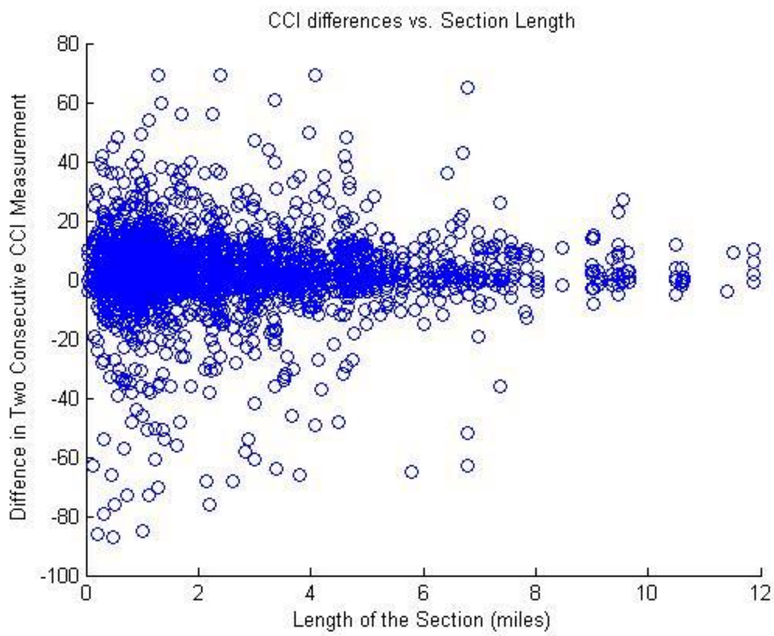


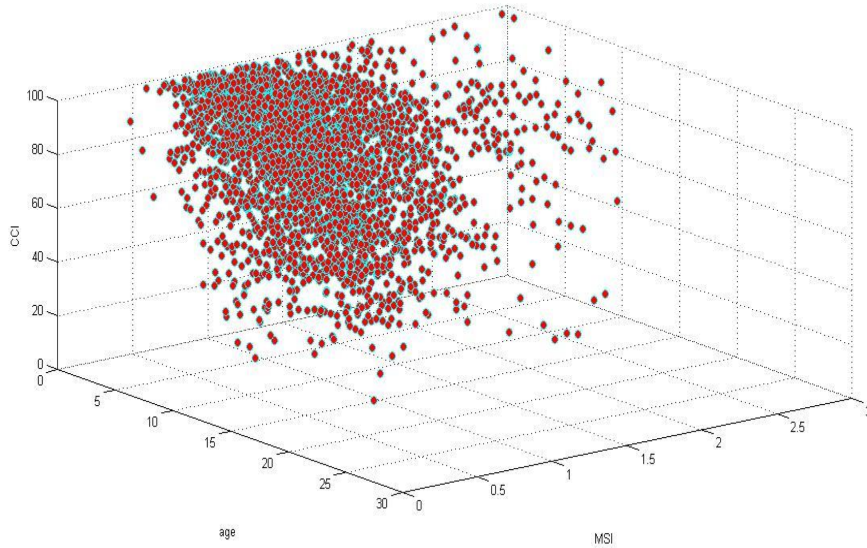
FIGURE 10 CCI differences in two consecutive measurements versus the length of the section.

3.3 Preliminary Exploratory Analysis

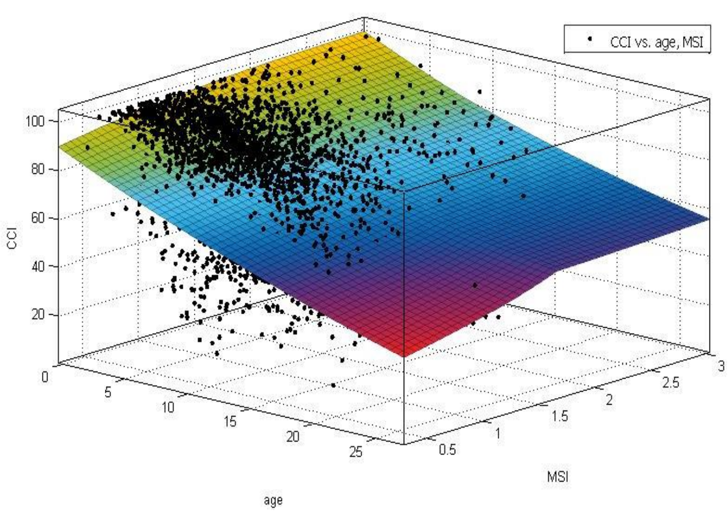
Visualization was used as a preliminary step to analyze the dataset created. Curve-fitting tools and 3D graphs available within the MATLAB software allowed clarification of the relation between the variables.

The raw data plotted in FIGURE 11 (a) depicts the high variability in the dataset. For example, it shows that sections exist with high MSI values that perform better at later ages than the sections with lower MSI values. However, it also shows a visible trend of faster deterioration for lower

MSI values. FIGURE 11 (b) shows a locally weighted smoothing linear regression with a moving average, roughly indicating that the trend mentioned above was observed. The dependence of the CCI on MSI can be observed at the two extreme planes:



(a) Raw Data.



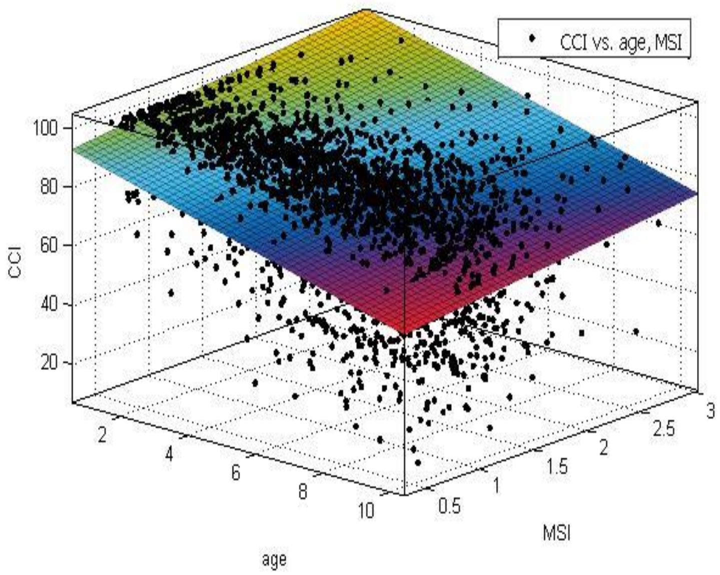
Locally weighted smoothing linear regression:
f(x,y) = lowess (linear) smoothing regression computed from p where x is normalized by mean 7.691 and std 4.302 and where y is normalized by mean 1.135 and std 0.3719
Coefficients: p = coefficient structure
Goodness of fit:
SSE: 8.532e+05
R-square: 0.2251
Adjusted R-square: 0.2243
RMSE: 16.67

.

(b) Locally weighted smoothing linear regression

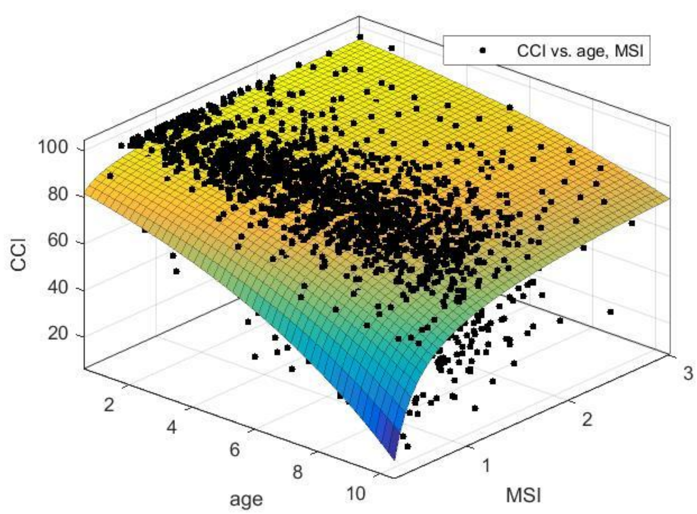
FIGURE 11 3D visualization of the complete dataset

FIGURE 12 superimposes basic one-degree linear polynomial (a) and exponential (b) curve-fitting models to the filtered data. The plots show that the sections with higher MSI values tend to deteriorate slower than the weaker sections with lower MSI.



Linear model Poly11:
f(x,y) = p00 + p10*x + p01*y
Coefficients (with 95% confidence bounds):
p00 = 92.87 (90.65, 95.1)
p10 = -3.147 (-3.368, -2.927)
p01 = 4.422 (2.888, 5.955)
Goodness of fit:
SSE: 6.171e+05
R-square: 0.2346
Adjusted R-square: 0.234
RMSE: 15.3

(a) Linear polynomial fit



Model using the exponential trend.

General model:

$$f(x, y) = 100 - a * \exp(-b * x + (y^\wedge d))$$

Coefficients (with 95% confidence bounds

$$a = 2.838 \ (2.56, 3.115)$$
$$b = -0.1567 \ (-0.1691, -0.1444)$$
$$d = -0.4552 \ (-0.5378, -0.3726)$$

Goodness of fit:
SSE: $6.094e+05$
$R - \text{square}$: 0.2442
Adjusted $R - \text{square}$: 0.2436
RMSE: 15.2

(b) Exponential fit

FIGURE 12 3-D visualization of the filtered dataset.

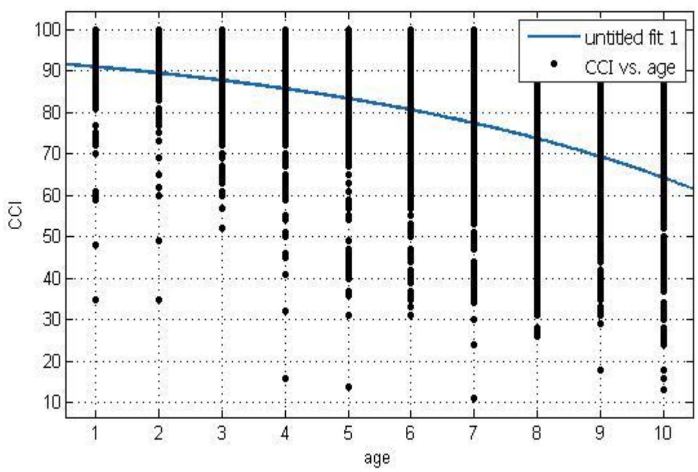
The impact of adding MSI as a variable in the model can be observed by comparing the goodness-of-fit statistics of the (b) Exponential fit

FIGURE 12 with a similar model that does not include MSI given in equation (23) and depicted in FIGURE 13. The results suggest that inclusion of MSI slightly improves the performance of the model as it increased the adjusted coefficient of determination (R^2) and produced a lower Root Mean Squared Error (RMSE) and Summed Square of Residuals (SSE). The improvement is slight especially considering the variability within the data. However, the example model shown in Equation 24 is a simple and straightforward equation derived from Equation 23 to include MSI, and is open to modifications to improve the prediction performance.

$$CCI = 100 - a * \exp(-b * AGE) \quad (23)$$

The MSI-inclusive form of the exponential pavement deterioration is given in the following equation:

$$CCI = 100 - a * \exp(-b * AGE + (MSI)^c) \quad (24)$$



General model:
f(x) = 100-a*exp(-b*x)
Coefficients (with 95% confidence bounds):
a = 7.772 (7.002, 8.542)
b = -0.153 (-0.1655, -0.1405)

Goodness-of-fit:
SSE: 6.288e+05
R-square: 0.2201
Adjusted R-square: 0.2198
RMSE: 15.44

FIGURE 13 Deterioration model with pavement age as the only parameter.

3.4 Additional Data Filtering

The data showing the pavement condition consists of highly censored data because of maintenance treatments that are often applied to pavements in the worst condition. This can lead to a biased deterioration model. Most of the sections are treated before reaching an old age once they reach an unacceptable serviceability level, even if they deteriorated in accordance with expectations. Therefore, data collected from old pavement sections do not represent the average pavement condition of all the sections, making it biased. For example, there are many old pavement sections that perform unexpectedly good. Although some of them could actually be very good performing sections despite their age, it is suspected that many of them are the result of wrongly computed pavement ages due to treatments that have not been recorded in the system. This has been partially confirmed by VDOT looking at raw pictures of the pavement sections. Due to the limited number of parameters used in the models, the impact of biased data can be severe. Hence, data filtering to select the model input becomes necessary in order to have an accurate model.

The data filtering approach of the model development phase is mainly based on bias identification and correction. If censoring of the events is not properly accounted for, the model may suffer from censoring bias (Paterson, 1987; Prozzi and Madanat, 2000).

FIGURE 14 shows the number of occurrences of differences between two consecutive measurements in one section after the pavement age was limited to 10 years. It shows that 1452 of the differences are negative, indicating a decrease in measured CCI in a year, while 857 of them are positive, showing an increase in measured CCI without any recorded construction activity. Theoretically, an increase in pavement CCI value without any recorded maintenance activity is unusual, and a significant increase is likely due to an error in measurement or a missing record of applied treatments. Even after the pavement age is limited to 10 years, it is possible to see increases as high as 70 points in the dataset while most of such increases are less than 30 points. While the mean of the decreases in CCI is 8.95 points with a standard deviation of 8.99, the mean of increases is 9.21 with a standard deviation of 13.60. The higher mean and standard deviation of CCI increases further support the hypothesis that some of these increases are unrealistically high, probably because of unrecorded maintenance activities.

To analyze the impact of such variations, both the deteriorations between two consecutive measurements and the increases were limited to a maximum value of 40 points. The mean and standard deviation of the differences are tabulated in TABLE 5.

TABLE 5 Statistics of Differences between Two Consecutive Measurements when the Difference is Limited to +/-40 Points

	Mean	*Standard Deviation*
Differences Between Two Consecutive Measurements	-2.43	9.65
Positive Differences (Increases in CCI)	6.71	7.56
Negative Differences (Deterioration)	-8.06	6.91

The limitation at 40 points maximum eliminated the biased data at some extent, indicated by the decreases in both the mean and the standard deviation of CCI increases. However, in order to eliminate the impact of the questionable data, the maximum allowed CCI increase without any recorded maintenance activities was selected to be 10 points.

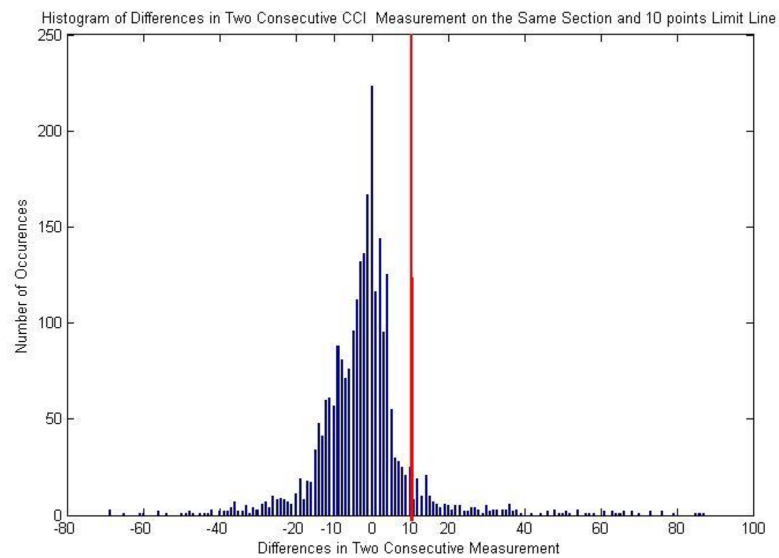


FIGURE 14 Differences in two consecutive CCI measurements within each section (a negative difference indicates a decrease in CCI).

Therefore, the data were further filtered by removing the sections that have more than a 10-point increase in CCI values between two consecutive measurements without any recorded construction work. The application of a CCI jump filter removed 188 sections (20% of total sections) and left 2043 measurements, without CCI jumps higher than 10 points and maximum age limited to 10 years, which were used to develop the model (59% of the original data).

CHAPTER 4. MODEL DEVELOPMENT

As noted in previous chapters, the purpose of this study was to develop enhanced deterioration models that incorporate the pavement structural capacity to improve the accuracy of pavement condition predictions. This chapter focuses on the model development phase. The available panel data allows the estimation of the pavement condition either as one time series regression for each section or as one cross-sectional regression at each point in time. Beside these frequently used techniques, there may be other more appropriate techniques. If the parameters of the deterioration model are assumed constant across all the sections and over the entire time interval, the estimation can be done by combining all data into a single regression, thereby pooling the data (Prozzi, 2001). This approach was adopted in this thesis for modeling CCI progression.

4.1 Modeling Approach and Key Considerations

While developing the deterioration model, which is targeted to be applicable at the network level and easy to implement into current pavement management system, there are several key points to consider. The most important characteristic is that the model should predict the deterioration observed on the pavements in-service as close as possible. In addition, the model should be practical and use only data that is available from network-level inspections. Therefore, the model was limited to two input variables: pavement section age and MSI value.

As discussed in Chapter 2, the MSI is a comprehensive structural index that considers the pavement structural number, traffic load in ESALs and subgrade modulus (Bryce et al., 2013). Furthermore, the model should satisfy the following boundary conditions:

- The initial condition index value equals 100.
- The condition index value cannot be negative.
- The slope of the deterioration curve should always be negative unless a treatment is applied, due to the irreversible nature of deterioration process.

Two different approaches were tested to find the most appropriate, promising model and then compared in terms of quantitative and qualitative criteria.

The first approach consisted of modeling the pavement condition (CCI) using regression analysis and a variety of models identified in the literature review. Data selection and filtering is crucial as they significantly determine the form and coefficients of the regression model.

The second approach modeled the pavement deterioration. The distribution of Deterioration Index (DI) was modeled using a negative binomial distribution and the predictions were then enhanced using an Empirical Bayesian approach to update the estimates and obtain an improved estimate of the individual pavement section's future condition.

4.2 Modeling Pavement Condition

Several model formulations were investigated to identify the most effective way of incorporating the MSI into the CCI prediction curves. These formulations include various equations used in applied sciences and existing pavement deterioration models: polynomial models, sigmoidal and true sigmoidal, exponential growth, simple logistic, Richards, Gompertz and Weibull equations.

4.2.1 CCI Regression Models

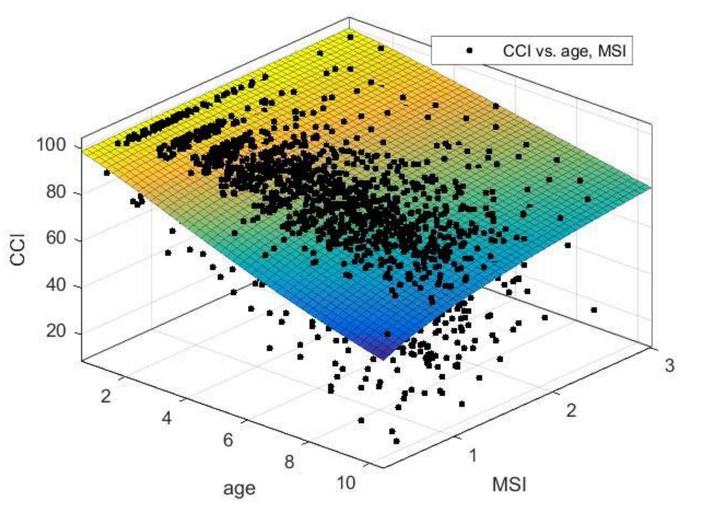
Various equations were tried for modeling CCI as a function of age and MSI. Given the known behavior of the pavement condition over time, and the nature of the current deterioration curves, some of the equations emerged as better solutions to the curve-fitting problem. The empirical equations selected as appropriate to model the deterioration of the pavement included: sigmoidal function, logistic growth curve, Gompertz model, Richards model, exponential growth curve and their derivations. The best-fit equations were determined for each of these models and superimposed to the filtered dataset in 3D curves using MATLAB. These plots allowed visualization of how well the model can explain the data to select the most suitable models and identify areas that need improvement. To this end, some adjustments were made on the general equations and different coefficients were used.

Sigmoidal Functions

Sigmoidal functions offer flexibility to describe the deterioration of a section in various shapes such as concave, convex, s-shaped or almost linear. This allowed modeling the faster decrease in CCI values of pavement sections with lower MSI values by fixing the parameters in the equation accordingly. The MSI and age parameters are placed in accordance with the theory that higher pavement age and lower MSI value lead to faster deterioration. The resulting best-fit sigmoidal function is presented in FIGURE 15. The figure also reports the 95% confidence range for the coefficients with various goodness-of-fit statistics.

$$CCI = 100 * \left[\frac{AGE^a}{b + AGE^a + MSI^c}\right] \quad (25)$$

The sigmoidal model meets the boundary conditions and seems to follow the general trend of pavement deterioration. Nevertheless, the adjusted coefficient of determination of 0.2942 indicates room for improvement.



General model:
f(x,y) = 100*((x^a)/(b+(x^a)+(y^c)))
Coefficients (with 95% confidence bounds):
a = -1.377 (-1.502, -1.252)
b = -0.9786 (-0.9839, -0.9733)
c = -0.008822 (-0.01208, -0.005565)

Goodness of fit:
SSE: 3.979e+05
R-square: 0.2949
Adjusted R-square: 0.2942
RMSE: 13.97

FIGURE 15 3D plot of the sigmoidal model and filtered data.

FIGURE 16 visualizes the model's sensitivity to MSI by displaying the deterioration for different MSI values. However, in this case the mathematical formulation of the model prevents discovering the deterioration of a hypothetical section with a theoretical infinite MSI. Therefore, it is not possible to comment on how much the model accounts for the factors that are not included as a parameter.

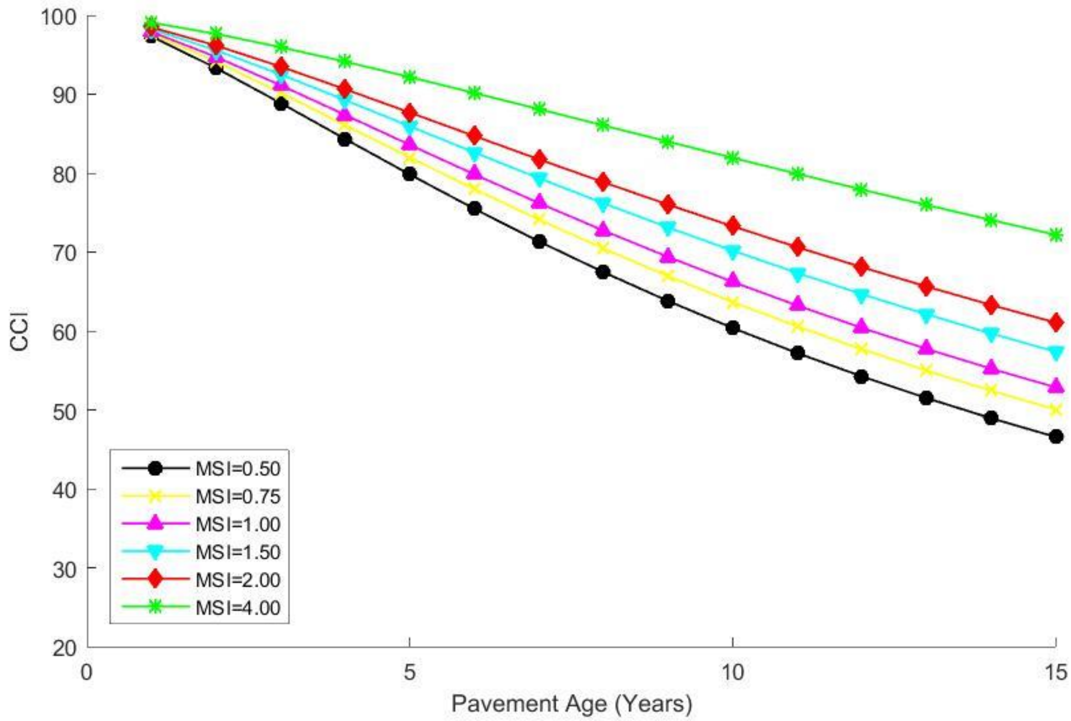
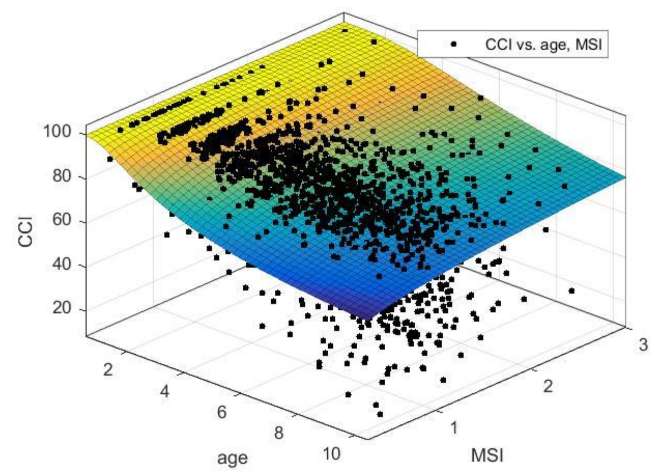


FIGURE 16 Pavement Deterioration with Sigmoidal Model for Different MSI Values.

Logistic Model

The logistic function is another widely used equation to describe deterioration and growth. An adjusted logistic curve equation is fitted to the data in FIGURE 17. It is generally known that higher pavement age and lower MSI values result in faster deterioration rates. Therefore, these parameters are placed inversely and coefficients are inserted to ensure the boundary conditions. As a result, the function, as given in equation (26), successfully meets the boundary conditions but the goodness-of-fit indicators are relatively weak.

$$CCI = \frac{100}{\left[1+e^{-a*\left(MSI^{b}/_{AGE^{c}}\right)}\right]} \qquad (26)$$



General model:
CCI=(100)/(1+exp(-a*(MSI^b)/(AGE^c)))
Coefficients (with 95% confidence bounds):
a = 6.472 (5.46, 7.484)
b = 0.3543 (0.2466, 0.4621)
c = 0.8939 (0.8065, 0.9813)

Goodness of fit:
SSE: 4.114e+05
R-square: 0.2709
Adjusted R-square: 0.2702
RMSE: 14.2

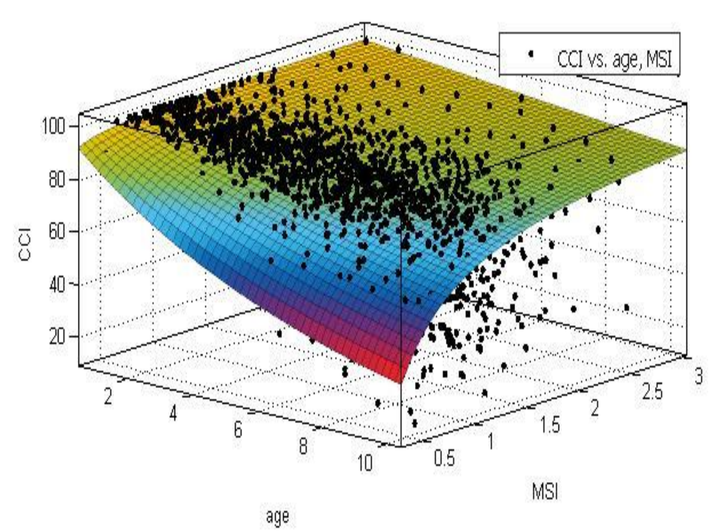
FIGURE 17 3D plot of the logistic model and filtered data.

Gompertz Model

The Gompertz equation (27), a special case of the generalized logistic function where the deterioration rate and the asymptote values can very easily be adjusted by the coefficients, seemed theoretically suitable to model the pavement deterioration:

$$y = ae^{-be^{-ct}} \tag{27}$$

This equation is generally used in biology and medicine to define aging or spreading of cancer cells. FIGURE 18 shows the results of fitting it to the pavement deterioration case where "Age" is the time, and MSI values are indirectly proportional to the rate of deterioration:



General model:
CCI = a*exp(-b*exp(-c*(AGE/MSI)))
Coefficients (with 95% confidence bounds):
a = 5.949 (-41.2, 53.1)
b = -2.799 (-10.71, 5.112)
c = 0.01255 (-0.02567, 0.05078)
Goodness of fit:
SSE: 3.795e+05
R-square: 0.2566
Adjusted R-square: 0.2559
RMSE: 13.9

FIGURE 18 3D plot of the Gompertz equation (27) and filtered data.

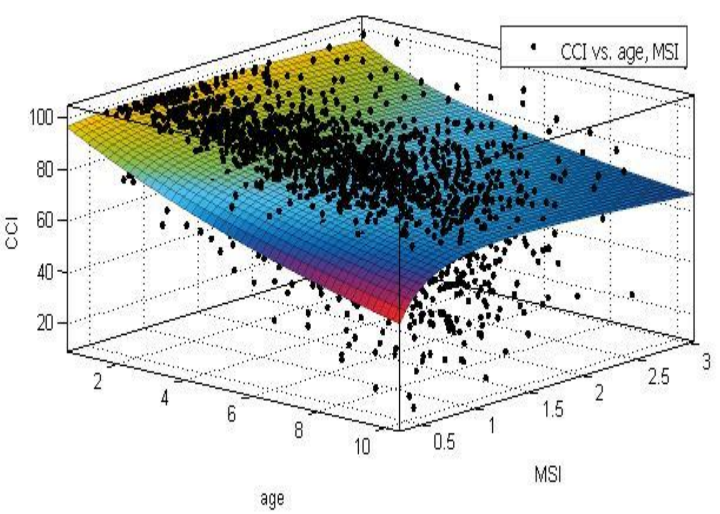
Richards' Curve

Another promising equation to describe the deterioration behavior is the Generalized Logistic Curve (Richards' Curve) given in equation (28). This equation is a commonly used growth model to fit a wide range of S-shaped curves. It generally is used with 4 or 5 parameters. The resulting curve is symmetrical about the point of inflection.

$$y = \beta + \frac{L_\infty}{\left(1+Te^{-k(t-t_m)}\right)^{1/T}} \qquad (28)$$

where β is the lower asymptote
L is the upper asymptote
k sets the growth rate
tm is the time of maximum growth
T is a parameter that sets the point of inflection

To include MSI and pavement age in a simpler way, an adjusted version of Richards' Curve is used to develop a model. MSI is inserted as the growth rate parameter defining the deterioration rate, in this case with lower asymptote being 0 and upper asymptote being 100. The pavement age parameter, which was relaxed by the addition of a coefficient, is set as the parameter that affects the point of inflection. The results are summarized in FIGURE 19. The resulting model is reasonable in appearance and satisfies the desired boundary conditions. The goodness-of-fit statistics indicate a relatively good statistical significance.



General model:
CCI =
100/((1+a*AGE*exp(MSI))^(b/MSI))
Coefficients (with 95% confidence bounds):
a = 0.02234 (0.01099, 0.0337)
b = 0.7203 (0.4195, 1.021)

Goodness of fit:
SSE: 4.069e+05
R-square: 0.2788
Adjusted R-square: 0.2785
RMSE: 14.12

FIGURE 19 3D plot of the Richards' Curve and filtered data.

Exponential Growth Function

Exponential growth function is frequently used in applied sciences and engineering when the growth rate is proportional to the function's current value. For this study, it is clear that the pavement condition of the next year for any given section is dependent on the pavement condition of the same section at the present year. For this reason, the negative exponential growth function was tested. The best-fit equation (29) is presented in FIGURE 20. FIGURE 21 shows the model's sensitivity to the MSI values.

$$CCI = 100 - [1 + MSI^{b}]^{AGE*a} \quad (29)$$

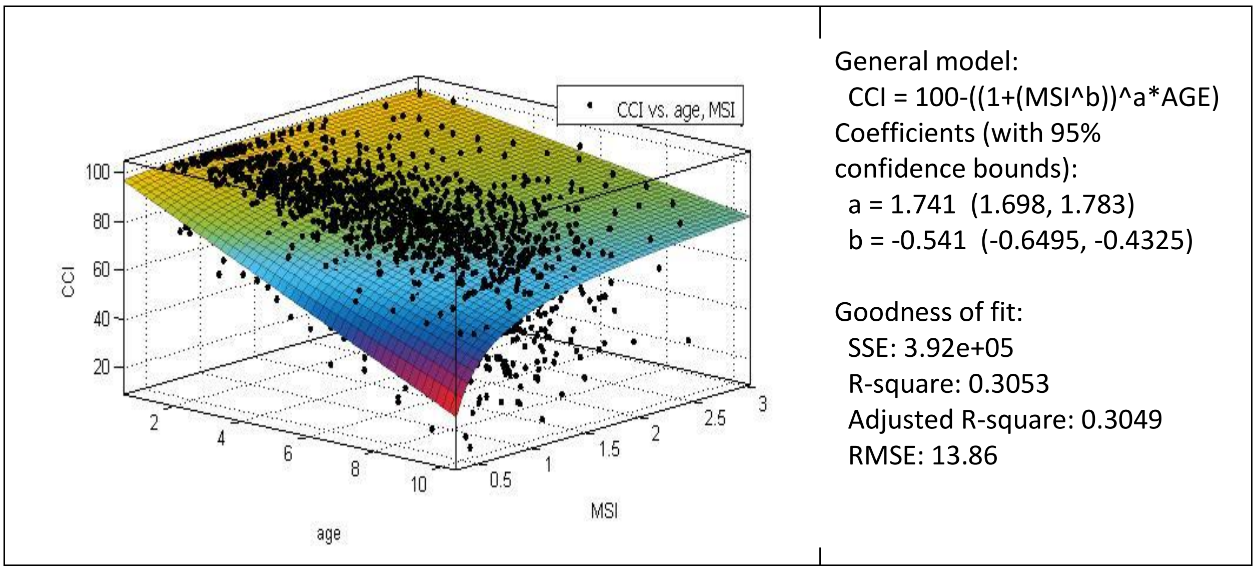


FIGURE 20 3D plot of the exponential growth function and filtered data.

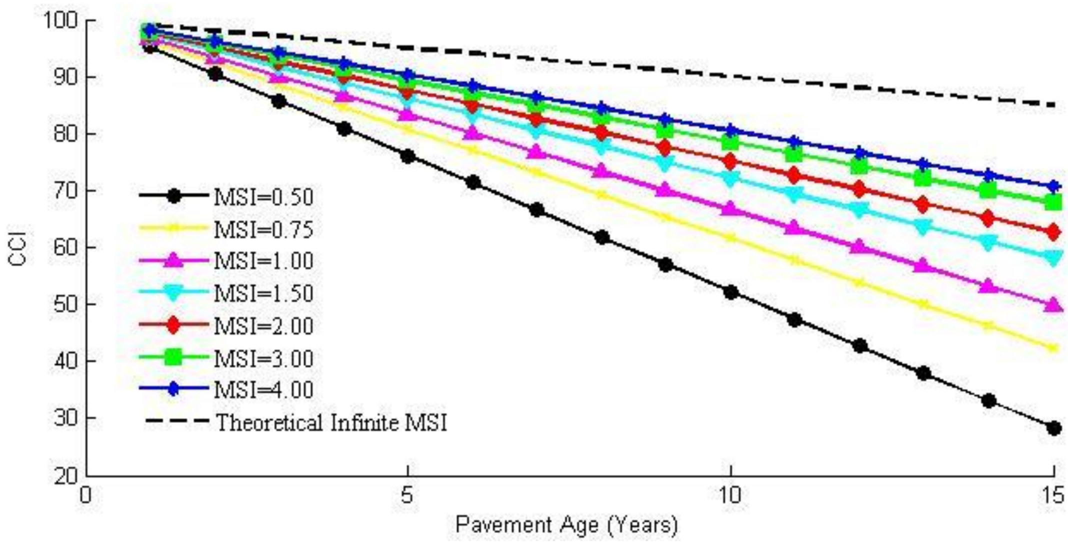


FIGURE 21: Pavement Deterioration according to the Exponential Growth Function for Different MSI Values.

Adjusted Stantec Model

Stantec (2007) developed the general form of equation (30) that is currently in use by VDOT to predict pavement deterioration for each combination of pavement type/performance index/treatment. This equation is as follows:

$$Index = I_0 - e^{(a-b*c^t)} \quad (30)$$

Where I_0 = Index immediately after rehabilitation (age zero)

a, b, c = model coefficients

t = Ln(1/Age)

The coefficients to the power "t" allow the equation to adapt to both, the rate of deterioration of a pavement section, and the change in the rate of deterioration in time. Therefore, this capability makes the equation a strong candidate to explain the behavior of the functional performance over the time while considering the structural capacity. The approach followed by Stantec was to use the windshield survey data and develop pavement performance models, assuming all pavement sections had corrective maintenance (CM) as a last treatment. The assumption of a last treatment was necessary, as the windshield data did not reflect the "strength" of the last treatment.

A very close correlation between coefficients "a" and "b" is visible in the report prepared by Stantec (2007). These coefficients seem to get very close numerical values in every case.

Therefore, a modified equation (31) was developed in this study to simplify the equation and avoid unnecessary coefficients:

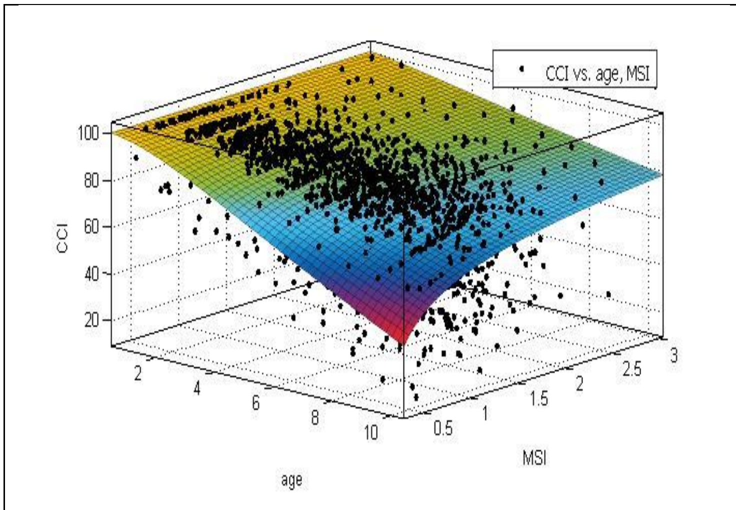
$$CCI = 100 - e^{(a(1-b^{t}))} \qquad (31)$$

Where CCI = Critical Condition Index
100 = CCI immediately after rehabilitation (age zero)
a,b = model coefficients
t = Ln(1/Age)

The data analysis of this study showed that pavement sections with a higher MSI tend to deteriorate slower. To this end, MSI should be a factor that effectively decelerates the decrease of CCI. This was implemented by including the inverse of MSI to a power coefficient ‘c’ to relax the rate of deceleration, as indicated in equation (32) and visualized in FIGURE 22.

$$CCI = 100 - exp\left[a * \left(1 - \left(\frac{b}{MSI^{c}}\right)^{ln\left(\frac{1}{AGE}\right)}\right)\right] \qquad (32)$$

Where CCI = Critical Condition Index
MSI = Modified Structural Index
100 = CCI immediately after rehabilitation (age zero)
a,b,c = model coefficients



General model:
CCI = 100-exp(a*(1-((b/(MSI^c))^log(1/AGE))))
Coefficients (with 95% confidence bounds):
a = 6.908 (5.326, 8.49)
b = 1.369 (1.23, 1.508)
c = 0.05514 (0.02811, 0.08217)
Goodness-of-fit:
SSE: 3.975e+05
R-square: 0.2955
Adjusted R-square: 0.2948
RMSE: 13.96

FIGURE 22 3D plot of the Adjusted Stantec model and filtered data.

FIGURE 23 illustrates the impact of MSI variability on the modeled deterioration rate. It is important to note that the model suggests a zero deterioration rate for pavement sections with a theoretical infinite MSI value. This is not a desired characteristic of a model since, in reality, there are many other factors that affect the deterioration even at the sections with perfect structural condition. It is practically impossible to have a zero deterioration rate with given VDOT traffic load and no maintenance activities.

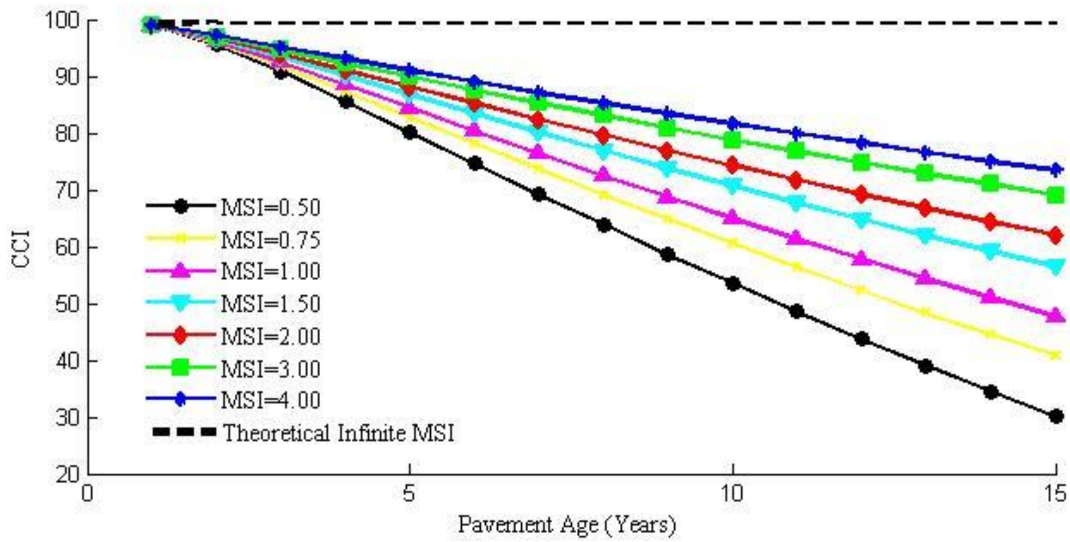


FIGURE 23: Pavement Deterioration according to the Adjusted Stantec model for Different MSI values.

4.3 Comparison of Regression Models and Selection

The various regression models are compared in TABLE 6.

TABLE 6 Comparison of Pavement Deterioration Regression Models

Model	Equation (CCI =)	# of Coefs	R^2	Adj. R^2	RMSE
Sigmoidal	100*((AGE^a)/(c+(AGE^a)+(MSI^b)))	3	0.295	0.294	13.97
Logistic	100/(1+exp(-a*(MSI^b/AGE^c)))	3	0.271	0.270	14.20
Gompertz	a*exp(-b*exp(-c*(AGE/MSI)))	3	0.257	0.256	13.90
Richards' Curve	100/((1+a*AGE*exp(MSI))^(b/MSI))	2	0.279	0.279	14.12
Exponential Growth	100-((1+(MSI^b))^a*AGE)	2	0.305	0.305	13.86
Adjusted Stantec	100-exp(a*(1-((b/(MSI^c))^log(1/AGE))))	3	0.296	0.295	13.96

Even though the goodness-of-fit indicators such as adjusted R-squared and RMSE are relatively poor compared to common statistical standards, they are quite comparable to other network-level models, considering the nature of pavement prediction models. In particular, the Exponential Growth model shows the best statistics, followed closely by the Adjusted Stantec and Sigmoidal models. The Adjusted Stantec model, being a modification of current VDOT deterioration models, is easy to implement. It is compared to the Exponential Growth model for different MSI values, as seen FIGURE 24.

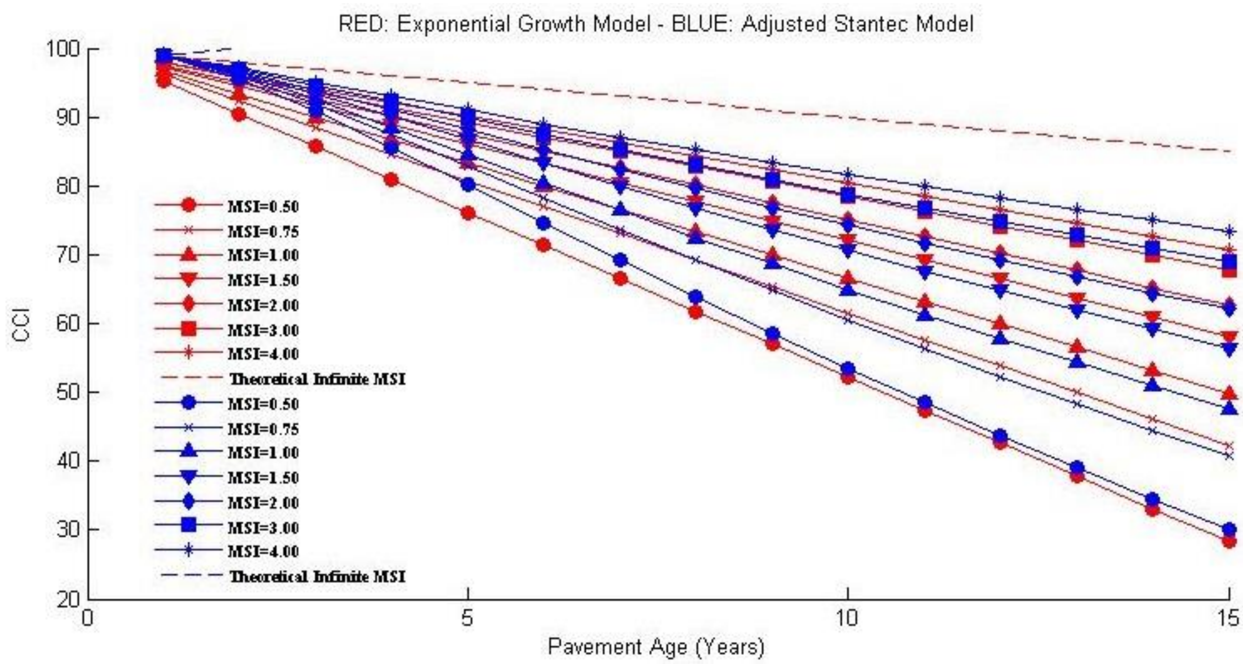
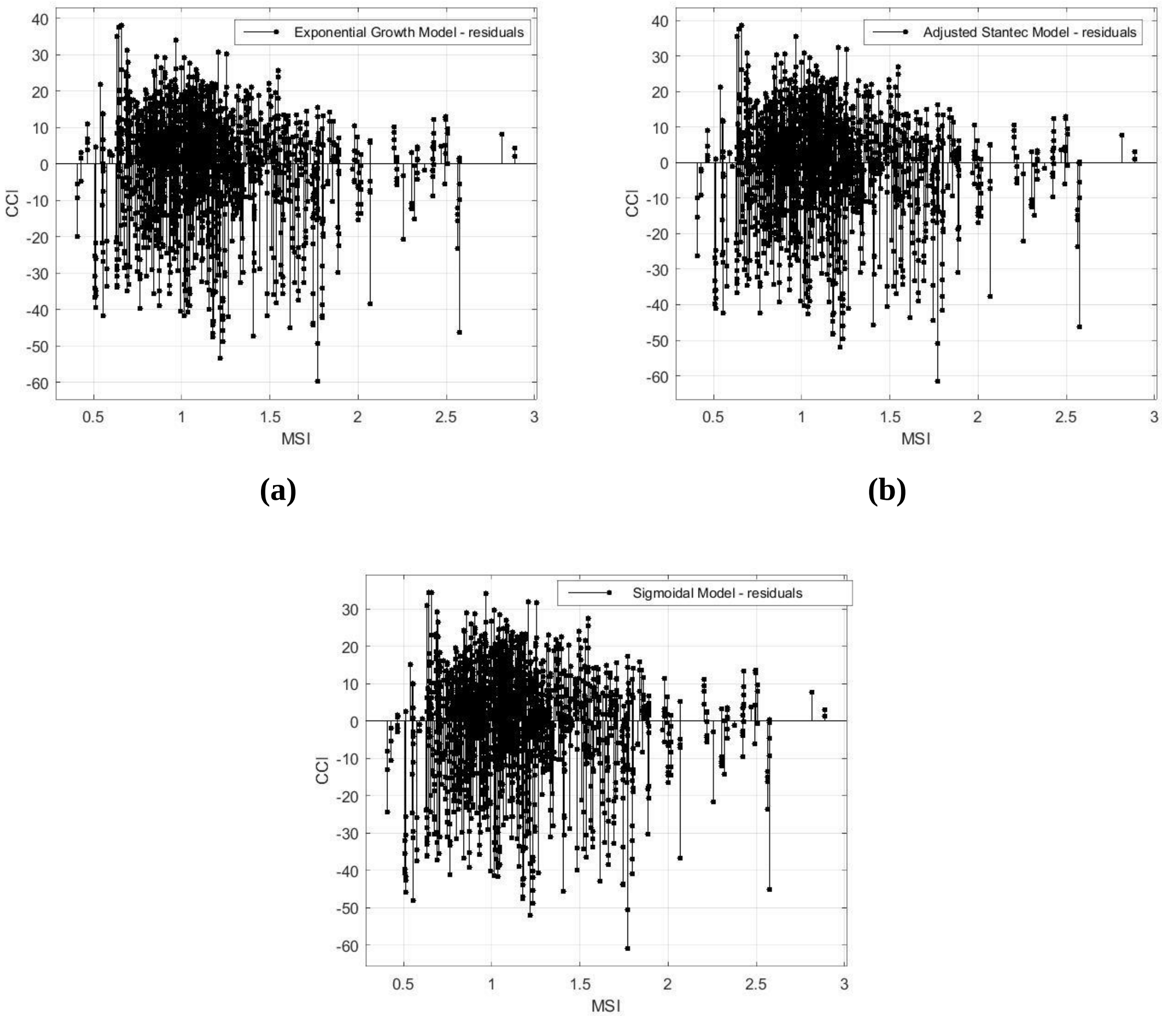


FIGURE 24: Pavement Deterioration for Different MSI Values with Exponential Growth and Adjusted Stantec Models.

Although the two models perform very similarly for pavements older than seven years, the Exponential Growth model shows a faster deterioration in the early life. Additionally, the deterioration curve of the theoretical infinite MSI in the Adjusted Stantec model produces an almost zero deterioration rate. FIGURE 25 poses the model residuals as CCI versus MSI in the three best performing models. The magnitudes of negative residuals are significantly higher than the positives. The reason behind this difference could be the data filtration and removal of extraordinarily well-performing aged pavement sections while keeping the extraordinarily poorly-performing sections. It should also be noted that having extraordinarily poorly-performing sections in this data was rare.



(a) (b) (c)

FIGURE 25 Plot of the CCI vs MSI Residuals for Three Best Performing Models

(a) Exponential Growth Model (b) Adjusted Stantec Model (c) Sigmoidal Model

4.3.1 Additional Validation on Data Filtering

The issue of censored data was further reviewed using two of the best performing models. At this point of study, it was easier to see the impact of filtering and verify that correct pavement age interval was chosen with a selected deterioration model. To validate the approach, the Adjusted Stantec Equation (being the most similar to the model currently in use) was modeled with 22 different datasets, each having pavement age limited to between 5 and 26. The changes in quantitative goodness-of-fit indicators, namely R-squared and RMSE, are noted along with the coefficients for each model determined through curve-fitting approach. These changes are plotted in FIGURE 26 and FIGURE 27.

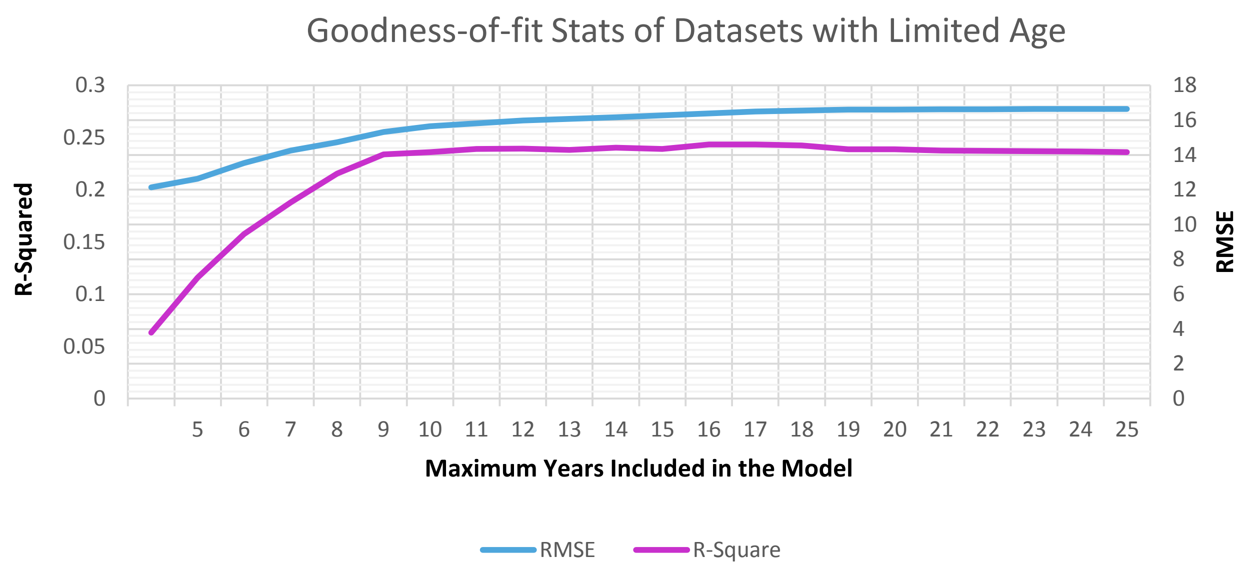


FIGURE 26 RMSE and R-squared values of the models with limited maximum ages.

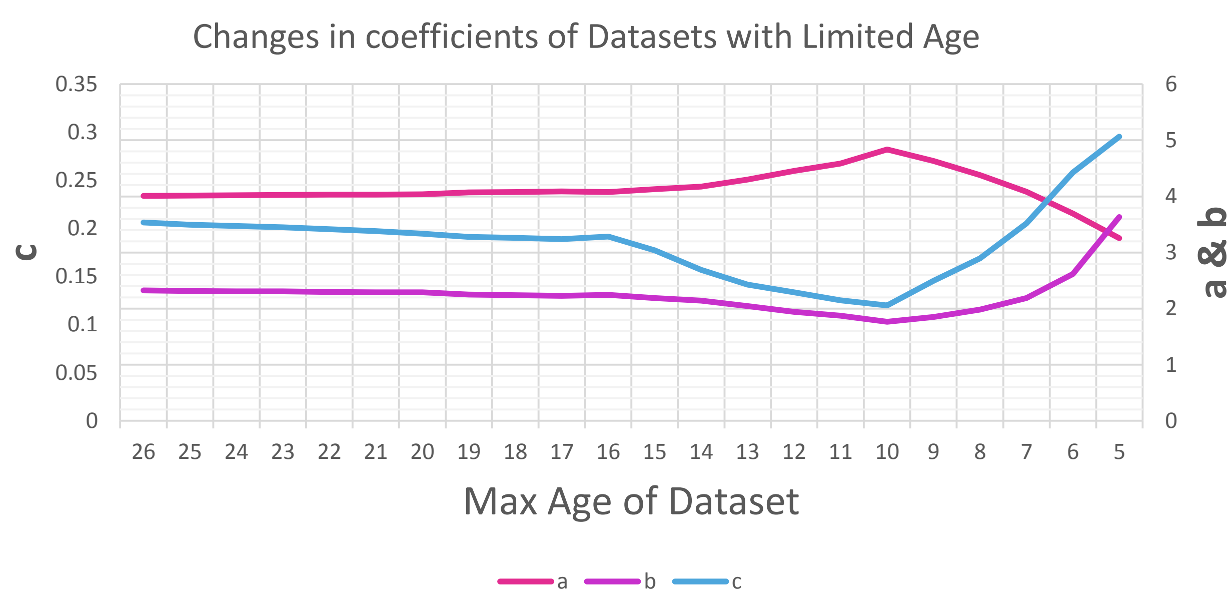
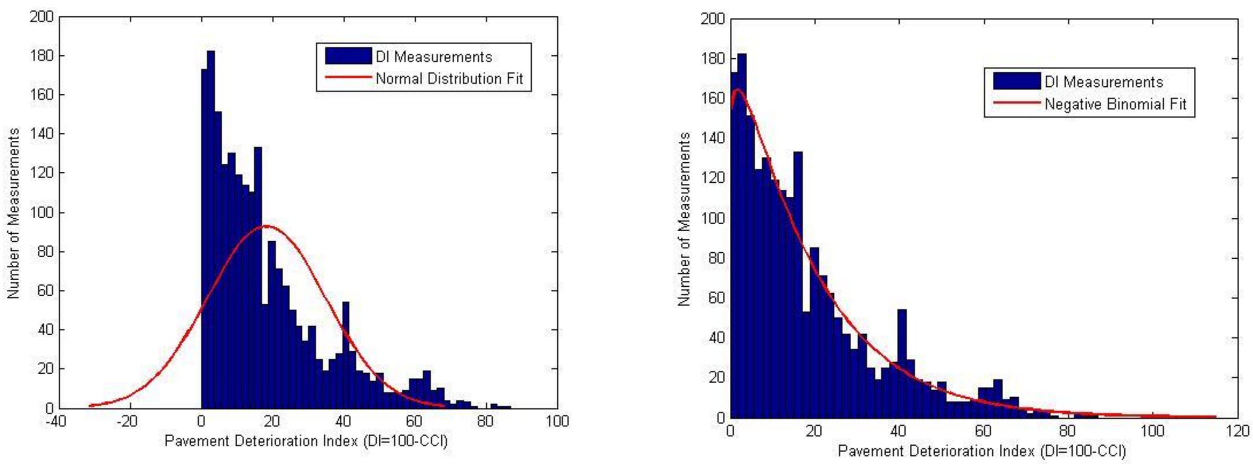


FIGURE 27 Coefficients in the models with limited maximum ages.

The pavement performance prediction framework, shown in **FIGURE 1**, indicates the importance of combining experiment results and expert judgement with observed pavement performance data. As both sources are available at this stage, engineering judgment and detailed analysis of the data allow refining of the input data. A change in the behavior of both models is observed when the maximum age to be included in the model is limited to 10 years. The coefficients a, b, and c hit their extreme values with maximum age limitation at 10 years, and the coefficient of determination decreases dramatically below 9-year limit.

4.4 Modeling Pavement Deterioration

Since the best performance models are in the form of the “100 minus” function, it seemed reasonable to model deterioration instead of CCI. To determine the most suitable approach for modeling, the statistical distribution of the pavement condition deterioration was evaluated.



(a)Normal Distribution **(b) Negative Binomial**

FIGURE 28 DI Histogram with Fitted Distributions.

To evaluate the *CCI* distribution, the *DI* (which is the complement of the *CCI*) was defined, as shown, in Equation 22. FIGURE 28 shows that the Negative Binomial distribution is a better fit to DI distribution than the normal distribution, making negative binomial more mathematically convenient to model. The probability density function of the Negative Binomial distribution, which is also a compound Poisson-gamma distribution, can be found in Appendix A along with the Poisson and Gamma distributions. FIGURE 29 shows that the negative binomial distribution matches closely with the empirical distribution of the pavement condition data, especially at early stages following construction. The condition data for the ten years after the pavement construction is slightly different from the negative binomial distribution.

Because the distribution of the pavement *DI* was well represented by a Negative Binomial distribution, the default pavement deterioration model was obtained using Negative Binomial regression, a form of the Generalized Linear Model (GLM) regression. In addition to providing a good fit to the observed data, the Negative Binomial model directly considers that variation in the observed pavement condition comes from two source; the first source is variation in performance of different pavement sections, while the second source is variation due to error in the measurement and reporting of the pavement condition.

Besides the fact that the *DI* values are distributed as a Negative Binomial distribution, one additional advantage of Negative Binomial regression is that its natural link function is the exponential function, which was used in 2007 by Stantec to develop the default pavement deterioration models. The final model used is given by Equation 33a and 33b. The equation has the same origin as the final model developed through the regression analysis.

$$\mu = e^{\left(\beta_0 + \beta_1 \frac{1}{MSI^3} + \beta_2 \ln(T)\right)} = T^{\beta_2} e^{\left(\beta_0 + \beta_1 \frac{1}{MSI^3}\right)} \tag{33a}$$

$$DI = AGE^a * \; exp\left[b + c * \left(1/_{MSI^3}\right)\right] \tag{33b}$$

Maximizing the likelihood function determined model parameters. After investigating different linear relationships in the exponential function, the final chosen model was the best fit with the maximized likelihood. Additionally, the model satisfies the boundary condition that the DI at year zero equals zero (CCI at year zero equals 100).

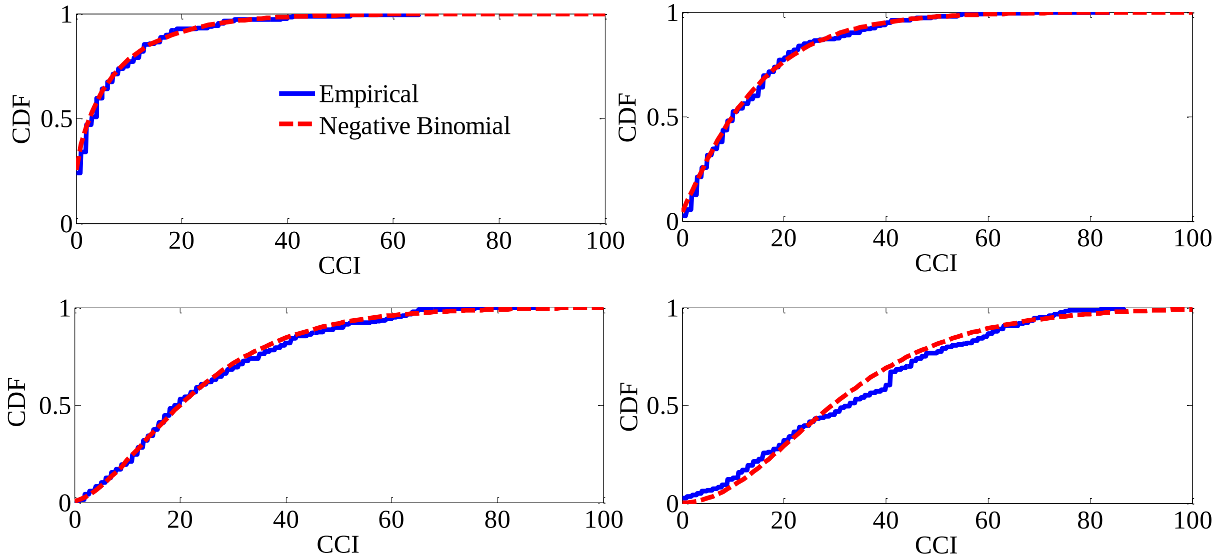


FIGURE 29 Comparison of Empirical Distribution of the Pavement DI Negative Binomial Distribution Fit for Different Pavement Ages; Top Left: 1 Year; Top Right: 4 Years; Bottom Left: 7 Years; Bottom Right: 10 Years.

The Negative Binomial regression model gives coefficients for the model parameters (*T*, MSI, and intercept) which, when substituted in Equation 33, give the mean response, μ, for the *DI*. Another parameter obtained from the Negative Binomial regression is the over-dispersion parameter, ϕ. This parameter considers the performance variability of different pavement sections. For the Negative Binomial model, the variance (σ_s^2) of the pavement sections condition can be calculated from the mean response (μ) and the over-dispersion parameter (ϕ), as shown in Equation 34. Under the Poisson error assumption, the variance of the error in *DI* reporting is equal to the mean response, μ. One way to justify the dependence of the variance on the pavement condition is to consider that it is easier to rate pavements in good condition than it is to rate pavements in bad condition. This will lead to ratings of pavements in a poorer condition having higher variability (i.e. error). This seems plausible, as there are many reasons that can cause a pavement section to be in poor condition but there is essentially one way for a pavement section to be in perfect condition. The data supports this observation as more variability is observed for pavement sections that are in worse condition. The standard deviation of the CCI measurements that are resulted lower than 70 points is 11.92 points, while the standard deviation of the measurements that are higher than 70 points is 8.20. Therefore, the total variance, σ^2_{mod}, of the model can be calculated as shown in Equation 35.

$$\sigma_s^2 = \phi\mu^2 \quad (34)$$

$$\sigma^2_{\text{mod}} = \mu(1+\phi\mu) \quad (35)$$

The estimate of the model parameters (Equation 33) β_0, β_1, and β_2 were 1.3672, 0.0830, and 0.8350, respectively, obtained from the weighted Negative Binomial regression with the data limited from 1 to 10 years (representing 76% of all the data). The over-dispersion parameter (ϕ) was equal to 0.3135.

The Akaike Information Criterion (AIC), which penalizes adding variables to the model, was used to validate incorporating the MSI in the model. The AIC assesses the fitness of a model based on the log-likelihood value of the model, L, and a penalty term related to the number of parameters, p. The AIC is calculated as shown in Equation 36.

$$AIC = -2\ln(L) + 2p \quad (36)$$

The AIC does not give an indication as to whether the model is the true model that generated the data. It can only be used to compare models and evaluate which one is more likely to be closer to the true model. This is done by calculating the exponential of half the relative difference between the AIC of two models being considered, as given in Equation 37.

$$w = \exp\left(\frac{AIC_{\min} - AIC_2}{2}\right) \quad (37)$$

Where "w" is the relative likelihood of model 2 being the model closer to the true (unknown) model that generated the data, compared to the model with the lowest AIC (model 1) (Burnham and Anderson, 2004). The two models evaluated in this thesis are the model with only the pavement age as a predictor of pavement condition (Equation 38) and the model with pavement age and MSI as predictors of pavement condition (Equation 39). The results of the AIC test are tabulated in TABLE 7.

TABLE 7 Akaike Information Criterion Results

Model	AIC
with MSI	10712.1
without MSI	10811.9
w=	2.14E-22

$$CCI = 100 - (AGE^a) * e^{\left(b + c\frac{1}{MSI^3}\right)} \quad (38)$$

$$CCI = 100 - (AGE^a) * e^b \quad (39)$$

The AIC weight, *w*, for the model with MSI and the model without MSI was less than 10^{-21}, indicating that the model with the *MSI* is at least 10^{21} more likely to be closer to the true pavement deterioration than the model without the MSI.

Deterioration curves for different MSI values are shown in FIGURE 30. Note that for MSI larger than 1, the predicted performance is practically the same whereas the performance changes significantly as the MSI decreases below 1. The limit for MSI increasing to infinity reflects the fact that pavement deterioration is not solely dependent on the MSI; other factors such as environmental loading and top down cracking also play a role in pavement deterioration, and these factors lead to deterioration no matter how strongly a pavement is designed.

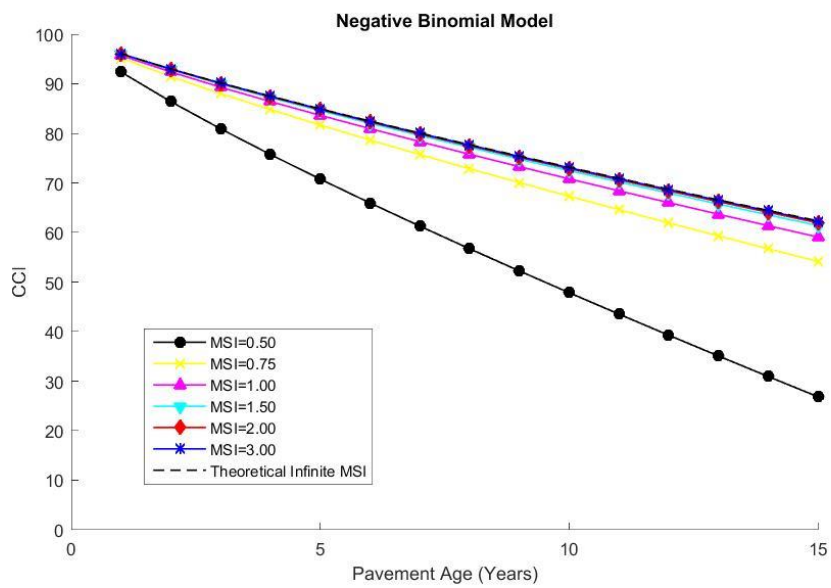


FIGURE 30 Pavement Deterioration for Different *MSI* values.

4.5 Application of Empirical Bayesian System to Negative Binomial Model

The negative binomial model with MSI values delivers improved estimates compared to those that do not account for MSI. However, the model does not consider the different performance of individual pavement sections that arises from other factors that are not included in the model. This difference in performance between individual sections can be observed in the individual measurements. The Empirical Bayes approach that estimates the prior distribution from the data itself can be useful to help accounting for this difference.

The Empirical Bayes approach arises from multi-parameter estimation problems where pooling information across the similar experiments results in a better estimate for each incident. By incorporating the variance and the condition information of all the sections into section-specific observations, every point estimate will be shifted towards the grand mean from its original section-specific estimate. The logic behind this shift is the similarity of the problem at each section. The prediction problems are independent for each section since each one shows different characteristics to affect the deterioration rate to some extent. However, the problem in each section is also very similar; the deterioration behavior model is general and replicable for each section, especially when considering that structural capacity is one of the parameters of the model and not predicted or fixed for the entire network.

In the Empirical Bayes approach, the parameters of the prior are estimated from the data, in this case by the Negative Binomial regression. The mean (μ), variance (σ^2_{mod}), and over-dispersion (ϕ) are related to the parameters of the Negative Binomial model as follows:

$$\mu = \frac{pr}{1-p} \tag{40}$$

$$\sigma^2_{\text{mod}} = \frac{pr}{(1-p)^2} \tag{41}$$

$$\phi = \frac{\sigma^2_{\text{mod}} - \mu}{\mu^2} = \frac{\frac{pr}{(1-p)^2} - \frac{pr}{(1-p)}}{\frac{p^2r^2}{(1-p)^2}} = \frac{\frac{pr - pr(1-p)}{(1-p)^2}}{\frac{p^2r^2}{(1-p)^2}} = \frac{1}{r} \tag{42}$$

The Poisson-Gamma model gives rise to a Bayesian model with the Gamma distribution prior. Practically, the prior represents the variability of the performance of different pavement sections. Once the DI data are observed, the posterior distribution of the true pavement condition can be calculated using Bayes' formula. The practical interpretation of Bayes' formula is that it combines the experience that can be learned from observing the historical performance of all pavement sections with specific observations to come up with an improved estimate of the pavement condition. The information from the prior and observation are combined using Equation 43, with details on how Equation 43 is obtained presented in Appendix A.

$$DI_{new}^{EB} = \mu_{posterior} = \frac{1}{\phi\mu_{prior}+1}\mu_{prior} + \left(1 - \frac{1}{\phi\mu_{prior}+1}\right)DI \quad (43)$$

$$CCI^{EB} = 100 - DI^{EB} \quad (44)$$

$$CCI_{+1}^{EB} = CCI^{EB} - \left(\mu_{prior_{+1}} - \mu_{prior}\right) \quad (45)$$

The EB updated DI (or CCI as noted in Equation 44) becomes the base to estimate for next year's performance. Subtracting the modeled deterioration from the updated CCI gives this estimation as shown in Equation 45. The EB estimate in Equation 43 assumes that the measurement error for reporting the *CCI* follows a Poisson distribution. The difference sequence method, described in Appendix B, was used to independently evaluate the variance of the error in reporting the *CCI*. It was found that this error variance is larger than what is predicted by the Poisson distribution. A concern could then be raised towards the applicability of the EB approach since the model assumptions (i.e. Poisson error distribution) are violated. It turns out that, even if the Poisson-Gamma model assumption is completely incorrect, the EB estimate is still a better estimate than the actual measurement as long as the variance of the error is not significantly overestimated. This is a results of the fact that linear Bayes estimators (e.g. the Poisson-Gamma model) guarantees to improve the estimate of the condition versus consideration of the measurement alone, regardless of the true distribution of pavement performance or the true distribution of the error in the measurement (Hartigan 1969 and Efron 1973). The improvement of the linear Bayes estimator is such that the mean square error is reduced by a factor of:

$$\frac{\sigma_s^2}{\sigma_{\text{mod}}^2} = \frac{\sigma_s^2}{\sigma_s^2 + \sigma_{Error}^2} \quad (46)$$

If the error variance, σ_{Error}^2, is underestimated (as is the case when assuming a Poisson error distribution) then the improvement of the EB estimate will be less than optimal. Therefore, the EB estimate in Equation 43 is conservative and can be improved if the appropriate value for the error variance is used. The linear EB estimator is calculated using Equation 47, which can be used for any two distributions and without knowledge of the appropriate distribution form.

$$\mu_{posterior} = \left(1 - \frac{\sigma_s^2}{\sigma_s^2 + \sigma_{Error}^2}\right)\mu_{prior} + \frac{\sigma_s^2}{\sigma_s^2 + \sigma_{Error}^2} DI \quad (47)$$

Note the similarities between Equation 43 and Equation 47; it can be shown that if $\sigma^2_{Error} = \alpha\sigma^2_{Poisson}$ (i.e. the error predicted from the Poisson distribution is different than the error in the data), then Equation 43 can still be used with ϕ replaced by $\phi_c = \phi/\alpha$ which is the form of Equation 48 (see Appendix A).

$$DI^{EB}_{new} = \mu_{posterior} = \frac{1}{\frac{\phi}{\alpha}\mu_{prior}+1}\mu_{prior} + \left(1 - \frac{1}{\frac{\phi}{\alpha}\mu_{prior}+1}\right)DI \tag{48}$$

The EB approach provides a mechanism to combine both sources of information that minimizes the mean square error (Bayes risk) between the estimated condition and the true condition (i.e., it gives the optimal method of combining both sources of information).

When the measurement error variance is assumed practically zero, the EB estimate of the condition is equal to the measured *CCI* and therefore the error is equal to 1. FIGURE 31 illustrates this case. As the estimate of the measurement error variance gets closer to the true measurement error variance, the mean square error of estimating the pavement condition using the EB method decreases. The best estimate is obtained when the measurement error variance is correctly estimated, in which case the mean square error of estimating the pavement condition is 0.75 (i.e. improvement of 25%). As the estimate of the measurement error variance starts to increase past the true measurement error variance, the error of the EB estimate starts to increase. The right side of FIGURE 31 represents the case where all the variance in the observed *CCI* measurement is wrongly assumed to be due to error in the measurements. In this case, the EB estimate coincides with the model estimate and the error increases to 2.60. FIGURE 31 also shows that the mean square error of 0.75 obtained with the correctly estimated measurement error variance coincides with the theoretical mean square error (based on Equation 40). The results presented in FIGURE 31 are based on leave-one-out cross-validation and incorporate an estimate of the pavement deterioration over a 1-year period obtained using the model. If the pavement deterioration is not taken into account, then the mean square error using the measurement is equal to 1.177.

Estimating the pavement condition from the model results in a significant error, that is 2.2 (2.6/1.177) times larger than the error of the measurement. However, the model is needed to account for the deterioration and obtain the EB estimate, which results in a 25% improvement in the estimate of the current pavement condition and a 36% improvement (0.36 = 1-0.75/1.177) in the estimate of the next year's pavement condition.

In FIGURE 32, the EB estimate combines the measured CCI with the model-predicted CCI to obtain a better estimate of the "true" CCI. The plots of the EB CCI and the model CCI are shifted for better visualization.

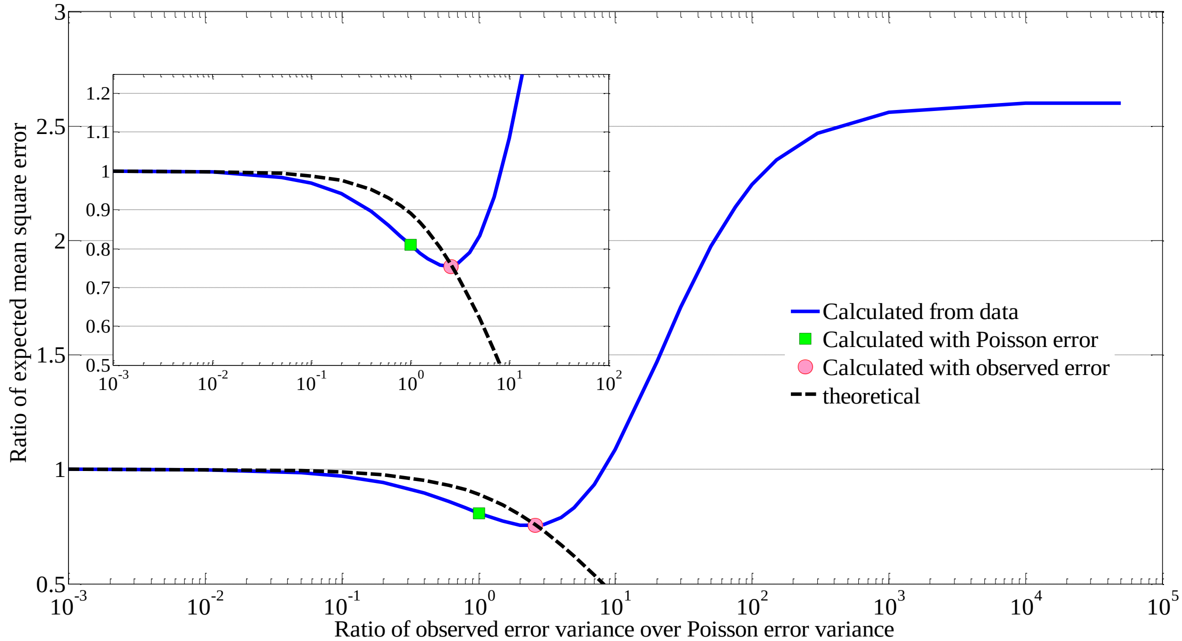


FIGURE 31: Error in estimating the Pavement Condition with the EB Method as a Function of Estimated Error Variance.

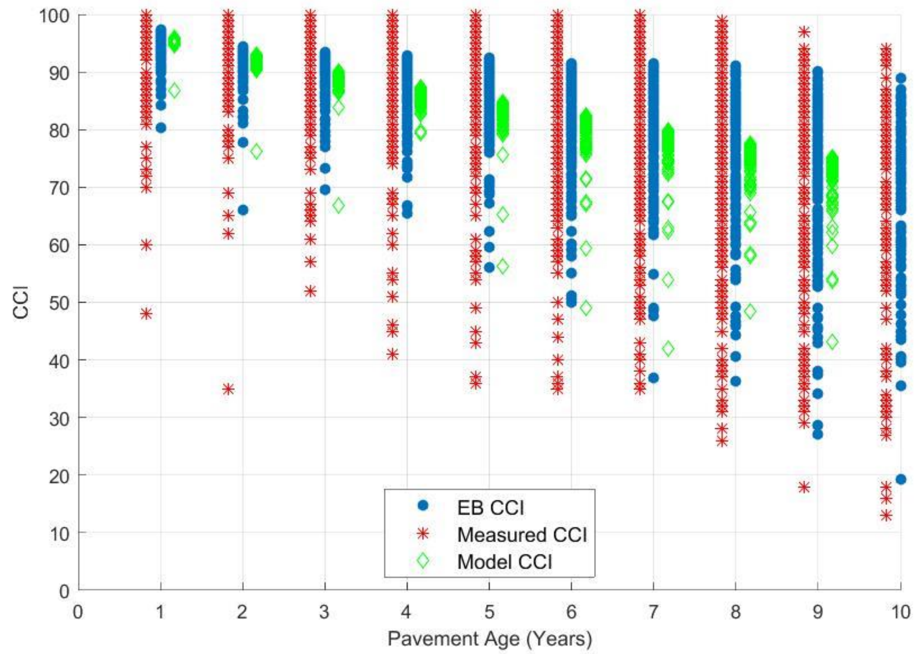


FIGURE 32 Comparison of measured CCI, model predicted CCI, and EB estimate of CCI.

4.5. Model Validation and Calibration

4.5.1. Dependence of Error in DI Measurement on Pavement Condition

Based on the difference sequence method, it was found that the Poisson distribution assumption underestimated the error variance of the observed data. A correction factor, α, was used to adjust for the discrepancy, as shown in Equation 49. However, since for the Poisson distribution the variance is equal to the mean, it was checked that the measurement error was related to the average of the observation, as shown in Equation 50. Finally, the total variance of the data was related to the variance of the pavement sections' performance and the error variance, as shown in Equation 51. σ^2_{mod} is the total variance which should be equal to the variance of the observed data.

$$\sigma^2_{Error} = \alpha \sigma^2_{Poisson} \tag{49}$$

$$\left|\text{Error}_i\right|^2 \approx \alpha \mu_{prior_i} \tag{50}$$

$$\sigma^2_{\text{mod}} = \sigma^2_s + \sigma^2_{Error} \tag{51}$$

4.5.2. Validating the Empirical Bayes Approach

The most accurate test to validate a modeling procedure would be to know the actual true quantity to be estimated (here the pavement deterioration represented by either the *DI* or the *CCI*) and verify that the chosen modeling procedure gives a better estimate of the true quantity compared to no modeling (i.e. just using the observations). Of course, for real data, the true value is never known and this approach cannot be followed. In some cases, however, alternative approaches that will tell whether the modeling approach improved the estimate can be used. When pavement condition data for two consecutive years exist, at each year the measurement consists of the true condition, *C*, plus a random error. The difference between two consecutive measurements can be decomposed into the difference between error terms and the difference between the true conditions (i.e. deterioration), as shown in Equation 52. The mean square of the differences can be calculated as shown in Equation 53. The quantities $\frac{1}{n-1}\sum_{i=1}^{n-1}\left(\varepsilon_{i+1}\right)^2$ and $\frac{1}{n-1}\sum_{i=1}^{n-1}\left(\varepsilon_{i}\right)^2$ are (almost identical) unbiased estimates of the error variance, σ^2_{Error}. Equation 53 quantifies the error made if the measurement DI_i is used to predict DI_{i+1}. However, Equation 53 can be used to estimate whether a new estimate, DI_i^{est}, of C_i is better than DI_i. Clearly, the smaller ε_i of the estimate is, the closer DI_i^{est} is to C_i, and the smaller the quantity D_i^2 is. Therefore, a DI_i^{est} that is better than DI_i reduces the quantity calculated using Equation 53. Equation 52 can also be used to evaluate the adequacy of a deterioration model, as detailed in Appendix B. This shows that Equation 53 (or modifications of it as presented in Appendix B) can be used to validate the modeling procedure without knowing the true (ground truth) pavement condition.

$$D_i = DI_{i+1} - DI_i = \left(C_{i+1} + \varepsilon_{i+1}\right) - \left(C_i + \varepsilon_{i1}\right) = \left(\varepsilon_{i+1} - \varepsilon_i\right) + \left(C_{i+1} - C_i\right) \quad (52)$$

$$\frac{1}{n-1}\sum_{i=1}^{n-1}\left(D_i\right)^2 = \frac{1}{n-1}\sum_{i=1}^{n-1}\left(\varepsilon_{i+1}\right)^2 + \left(\varepsilon_i\right)^2 + \left(C_{i+1} - C_i\right)^2 \quad (53)$$

The final verification consists of checking whether Equation 51 is approximately satisfied. σ_s^2 was calculated as 190 while σ^2_{Error} was calculated as 22 for a total σ^2_{mod} of 212. The total variance of the data was estimated as 273, showing that the model variance underestimates the data variance by 61 (273-212). While the estimate of the data's error variance was obtained for the positive error, since only positive errors were considered as the result of possibly missing recorded treatments, the total (both negative and positive) variance of the data includes observations with possibly missing recorded treatment. Therefore, the variance of the all the error (positive and negative) was evaluated as 72 and added to σ_s^2, to give a total variance for the model σ_{mod}^2 262 which is reasonably close to the total variance of 273 of the data. This shows that the modeling approach is consistent as the resulting model practically accounts for all the variance observed in the data.

4.6.Comparison of the Models

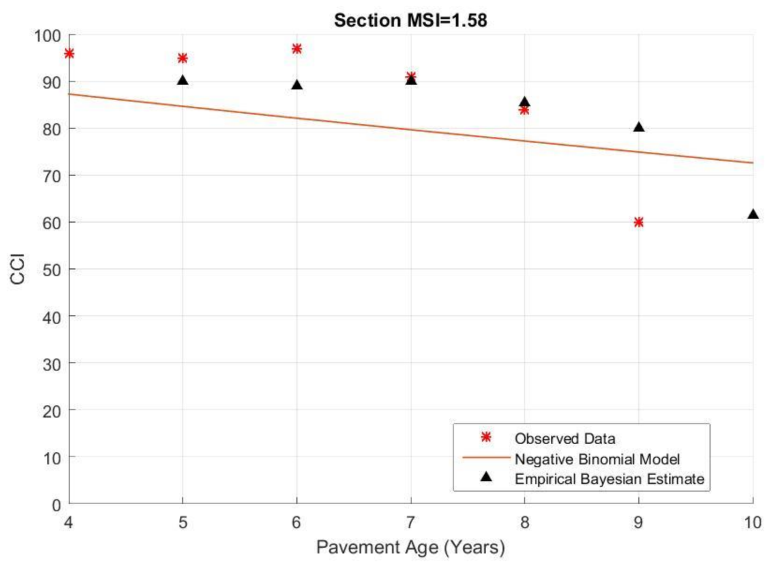
To have a statistical comparison between the different modeling approaches, the R-squared values of the final regression models are computed and listed in TABLE 8. Negative binomial regression does not have a natural equivalent to the R-squared measure found in ordinary least squares (OLS) regression; however, there exist many attempts to create one. Here, the R-squared value for Negative Binomial is calculated manually. Equation 33 was set with the determined coefficients of β_0=1.3672, β_1=0.0830, and β_2=0.8350. Computation of the residual sum of squares of this model and total variance of the final data set allowed the calculation of R-squared. The resulting coefficient of determination was found to be 0.312, as stated in TABLE 8. The Root Mean Squared Error of this form of the Negative Binomial model is 13.75.

TABLE 8 R-squared values of final EB and Regression models

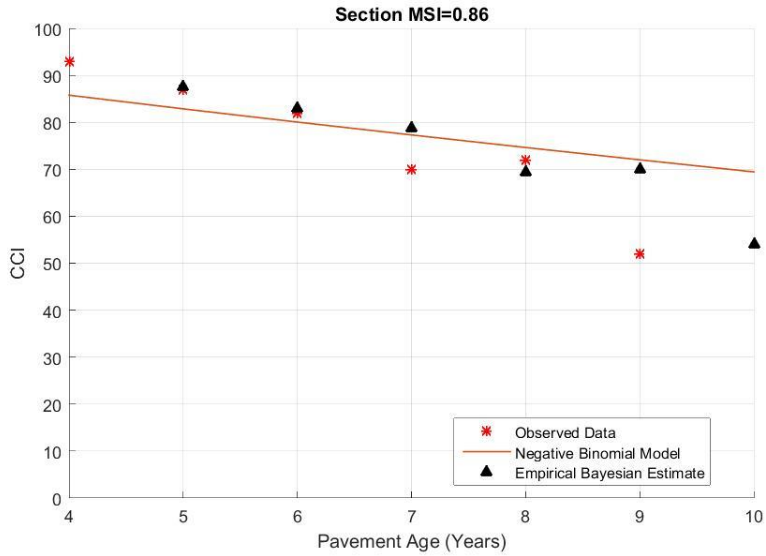
Model	Adjusted R2 values
Negative Binomial Model	0.312
Exponential Growth Model	0.305
Adjusted Stantec Model	0.295

FIGURE 34 shows the different deterioration curves with different MSI values over time. As indicated previously, Negative Binomial and Exponential Growth models are somewhat capable of considering the influential factors in pavement deterioration rate other than MSI by displaying a decreasing curve for the deterioration of sections with a theoretical infinite MSI, while the Adjusted Stantec Models fails to do so.

The implementation of Empirical Bayesian approach was proven to improve the overall accuracy of the model, with the optimum improvement in accuracy being 25%. Due to the computational practice, having a negative binomial or 'Poisson-Gamma' prior distribution was needed to apply the Empirical Bayesian. FIGURE 33 illustrates how the empirical Bayesian method estimate varies from the model prediction in example sections. It simply shifts the prediction towards the observations. Application of the EB method to Exponential Growth model requires complex computations that are not cost or time effective, especially in this case, where the statistical indicators favor the Negative binomial model as well. Therefore, the optimum accurate and feasible method to be used in modeling the pavement deterioration versus MSI for VDOT interstate network becomes the implementation of Empirical Bayesian method to the Negative Binomial model.



(a) MSI=1.58



(b) MSI=0.86

FIGURE 33 Example comparisons of EB estimates versus observed data and negative binomial model in two different sections (a) MSI=1.58 (b) MSI=0.86

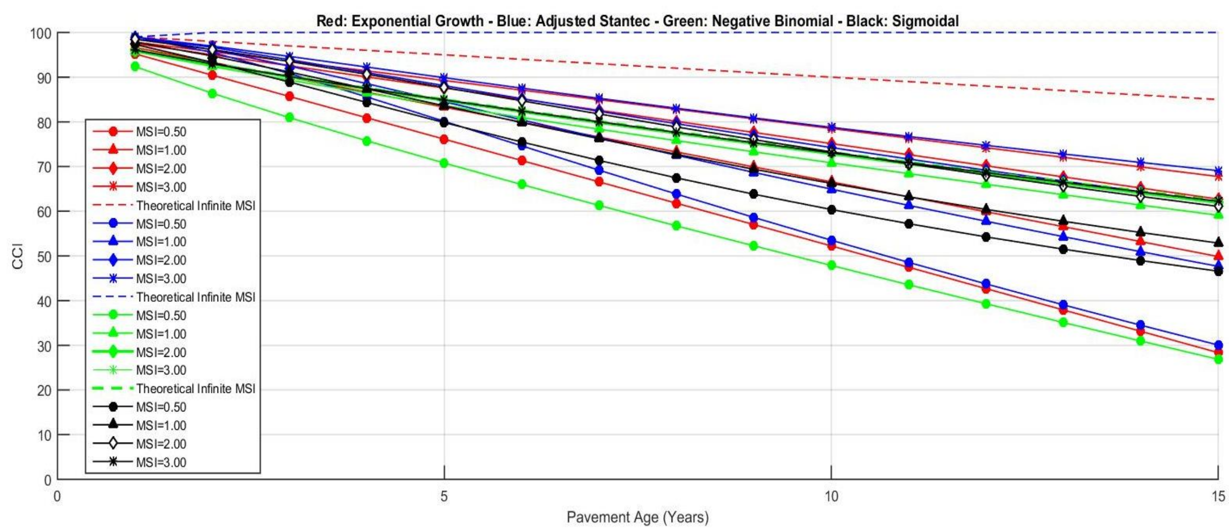


FIGURE 34: Pavement Deterioration for Different MSI Values with Different Models starting from Year 1.

4.7. Discussion

The statistics and plots show that employing the Empirical Bayesian method improves the performance of the overall model. It is noted that the Empirical Bayesian approach pulls the individual sections' deterioration curves towards the overall network average, resulting in a reduced variance. Consequently, this approach improves the model's accuracy to predict the pavement deterioration of the entire network, making the Empirical Bayesian method a useful tool for network-level predictions, and supporting network-level decision-making. On the other hand, by reducing the variance of the data, the EB method shifts away the input data for each section's future performance prediction from being section-specific to converge to the general network trend. Due to this EB approach characteristic, it does not help to improve the specific predictions for each section's future performance. Hence, the use of the Empirical Bayesian approach to predict how a specific section will deteriorate, and then to make section- or project-specific decisions accordingly, becomes questionable.

CHAPTER 5. FINDINGS, CONCLUSIONS, AND RECOMMENDATIONS

The literature agrees that accurate performance prediction is important in pavement management systems. In accordance, many approaches and emerging techniques have been used to model the pavement condition. Even though VDOT's current pavement management system operates satisfactorily to meet the demand, the addition of any improvement is valuable. The deterioration models used by VDOT were developed based on windshield pavement condition data and do not account for the structural capacity. This study investigated the applicable modeling approaches in order to incorporate the pavement structural capacity into the deterioration rate, then determined the suitability of such approaches, and exhibited the improvements that can be achieved.

5.1. Findings

- Deterioration models are crucial in pavement management since accurate performance prediction enables agencies to optimize their budget with better planning and scheduling. Many factors exist that contribute to pavement deterioration, such as traffic loading, environmental factors, material properties, and construction work quality. Structural condition of pavement is an important parameter that affects the deterioration rate, and accordingly, the maintenance and rehabilitation activities required to keep it at a serviceable level. The impact of structural condition on network-level decision-making should be taken into account for better planning and budget allocation.
- Many deterioration models have been developed, yet there is still need for improving accuracy and practicality. The number of deterioration models that include the structural condition is somewhat limited. On the other hand, several modeling techniques have recently been in use to improve the accuracy and facilitate the computation, such as Empirical Bayesian, Artificial Neural Networks, Fuzzy Logic systems and genetic algorithms. Additionally, various modeling approaches and developed equations have shown potential to be applicable in pavement deterioration modeling.
- Data filtering can be used to improve predictive capability. The data was censored because of maintenance and rehabilitation activities applied to poorly performing pavements before the end of the pavements' expected service, producing unrealistically good performances for the pavement sections that are older than 10 years. The analysis

showed that data from the first 10 years give the most adequate representation of pavement behavior. Filtering out all the measurements that were performed on sections older than 10 years resulted in the exclusion of 24% of the data points. With this filtered data, inclusion of MSI into prediction of CCI, even in a simple regression model, delivered better goodness-of-fit statistics. In part of the analysis, further filtering was also applied by removing the pavement sections that recorded more than a 10-point increase in two consecutive CCI measurements.

- Implementation of many different approaches and functions, which were developed to model deterioration and growth in various fields of science, showed in this study that they are (to some extent) capable of explaining the pavement deterioration characteristics. A modification of VDOT's currently used model, prepared by Stantec, has delivered a good fit with improved statistical indices and satisfied boundary conditions, in addition to its adaptability.
- In accordance with the objective of a more accurate model, Empirical Bayesian approach was applied to the modeling. Due to the over-dispersion observed in the data, Negative Binomial regression (a form of a Generalized Linear Model) was found to be a good representation of pavement deterioration. This allows for a better understanding and modeling approach to pavement condition, where variability in pavement condition can be decomposed into variability due to different performance of different pavement sections and variability due to error in measuring the pavement condition.
- Using the models with a theoretical infinite MSI still yielded to a significant deterioration over the years for most of the models, since MSI is not the only factor that affects the pavement deterioration. The pavement deterioration with a theoretical infinite MSI showed that the model realistically accounts for other factors, though roughly, since only two parameters exist.
- Negative Binomial, Adjusted Stantec, Exponential Growth and Sigmoidal models were found to have the highest R-squared values with slight differences. Negative binomial model performs more realistically when the impact of other factors affecting the pavement deterioration is considered. This was reflected by the significant deterioration of pavement sections with a theoretical infinite MSI value.

- The Empirical Bayesian updated model clearly exposes the improvement by shifting the model predictions closer to the observed data. The Negative Binomial model provides a more computable prior distribution that facilitates the application of Empirical Bayesian method. In case of optimal improvement done by EB, the accuracy of the future predictions increases by 25%.

5.2. Conclusions

The pavement structural condition is a significant parameter that affects the pavement condition; using the AIC criterion, the model that incorporates the pavement structural parameter as an explanatory variable of the pavement condition was more than 10^{21} times more likely to be close to the “true model”, compared to the model that does not incorporate the pavement structural condition.

The optimal estimate of the pavement condition is one that combines the observed condition with the model predicted condition. The estimate is obtained by an EB approach, which combines the model estimate with the observed condition through a weighted average. The weight is determined by the relative variability of the error in the measurement of the pavement condition and the variability of the performance of different pavement conditions. The model on its own gives an inaccurate estimate of the pavement condition with a mean square error that is about 2.22 times the mean square error of the observations. However, combining the observations with the model resulted in an estimated mean square error that is about 0.75 times the mean square error of the observations. The estimate of the improvement is based on cross validation where observations were held out and used to estimate the mean square error prediction.

5.3. Recommendations for Implementation

VDOT should continue performing network-level pavement structural evaluation. The pavement structural condition summarized in terms of the MSI was found to affect the rate of pavement deterioration. To include the structural condition into deterioration modeling permanently, MSI should be computed and added to the pavement management system database after structural evaluations. Employing emerging technologies that allow faster and easier structural evaluations on a wide network, such as TSD, may be taken into consideration in order to improve data collection practices.

VDOT's Maintenance Division should implement the Empirical Bayes method to determine the pavement condition of Interstate roads. The approach to estimate the pavement deterioration proposed in this research can be readily implemented into the VDOT PMS and results in an estimated 25% improvement (mean square error) in the predicted pavement condition. In this study, the EB method was implemented for Interstate roads.

5.4. Recommendations for Future Research

The developed model should be reevaluated as more data becomes available in the PMS. The data used in this study spanned from 2007 to 2012. For some of the pavement sections, the accuracy of the reported age is questionable. Hopefully, as more data are collected, the history of the different pavement sections will be more accurate, warranting model reevaluation.

The varying section lengths prevent application of some different approaches, or bring extra difficulties in calculations and analysis. Computing both the structural index and the conditional index over equally long sections will be helpful in further studies.

The implementation of the EB method was shown to improve the overall accuracy of the model; however, the assumptions lying behind the method and its impact on section-specific applications should be further evaluated, and adjustments should be developed in order to adapt the model for section-specific predictions.

VDOT's Maintenance Division should develop a similar approach for the pavement condition of primary and secondary roads. The implementation for secondary roads that are only evaluated at 5-year cycles is especially needed. The EB method combined with the modeled deterioration should provide for a better prediction of the conditions of secondary roads during years when the conditions are not collected.

REFERENCES

Administration, F. H. (2015). "Long Term Pavement Performance Program Database."

Ammons, D. N. (2012). Municipal benchmarks: Assessing local performance and establishing community standards, ME Sharpe.

Attoh-Okine, N. (1995). "Using adaptive neural networks to identify significant variables in pavement performance prediction." Intelligent Engineering Systems through Neural Network, 5, 803-808.

Attoh-Okine, N. O. (2005). "Modeling incremental pavement roughness using functional network." Canadian Journal of Civil Engineering, 32(5), 899-905.

Belsley, D. A. (1991). Conditioning diagnostics, Wiley Online Library.

Bryce, J., Flintsch, G. W., Katicha, S. W., and Diefenderfer, B. K. (2013). "Developing a Network-Level Structural Capacity Index for Structural Evaluation of Pavements."

Burnham, K. P., and Anderson, D. R. (2004). "Multimodel inference understanding AIC and BIC in model selection." Sociological methods & research, 33(2), 261-304.

Byers, A. L., Allore, H., Gill, T. M., and Peduzzi, P. N. (2003). "Application of negative binomial modeling for discrete outcomes: a case study in aging research." Journal of clinical epidemiology, 56(6), 559-564.

Dimitrova, M. (1996). "A Neural Compound Neural Network for User Identification In Adaptive Human Computer Interfaces." J. Complex Control Systems, 1.

Draper, N., and Smith, H. (1981). "Applied regression analysis, 709 pp." John Wiley, New York.

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.

Federal Highway Administration, Miller, S., Bellinger Y. (2014). “Distress Identification Manual for the Long-Term Pavement Performance Program (Fifth Revised Edition)” Report FHWA-HRT-13-092

Flintsch, G., Katicha, S., Bryce, J., Ferne, B., Nell, S., and Diefenderfer, B. (2013). Assessment of continuous pavement deflection measuring technologies.

Flintsch, G. W., and Chen, C. (2004). "Soft computing applications in infrastructure management." Journal of Infrastructure Systems, 10(4), 157-166.

Flintsch, G. W., and McGhee, K. K. (2009). Quality management of pavement condition data collection, Transportation Research Board.

Flora, W. F. (2009). Development of a structural index for pavement management: An exploratory analysis, Purdue University.

Fwa, T. F. (2005). The Handbook of Highway Engineering, CRC Press.

Gedafa, D. S., Hossain, M., Miller, R., and Van, T. (2009). "Estimation of remaining service life of flexible pavements from surface deflections." Journal of Transportation Engineering, 136(4), 342-352.

George, K. (2000). "MDOT pavement management system: prediction models and feedback system." Department of Civil Engineering The University of Mississippi. Report FHWA/MS-DOT-RD-00-119.

George, K., Rajagopal, A., and Lim, L. (1989). "Models for predicting pavement deterioration." Transportation research record(1215).

Gompertz, B. (1825). "On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies." Philosophical transactions of the Royal Society of London, 513-583.

Haas, R., Hudson, W. R., and Zaniewski, J. P. (1994). Modern pavement management.

Hadidi, R., and Gucunski, N. (2010). "Comparative study of static and dynamic falling weight deflectometer back-calculations using probabilistic approach." Journal of Transportation Engineering, 136(3), 196-204.

Hajek, J. J., Phang, W., Wrong, G., Prakash, A., and Stott, G. (1986). "Pavement Condition Index (PCI) for Flexible Pavements."

Hartigan, J. (1969). "Linear bayesian methods." Journal of the Royal Statistical Society. Series B (Methodological), 446-454.
Huang, Y. H. (1993). Pavement analysis and design.
Kaur, D., and Pulugurta, H. (2008). "Comparative analysis of fuzzy decision tree and logistic regression methods for pavement treatment prediction." WSEAS Transactions on Information Science and Applications, 5(6), 979-990.
Lee, Y.-H., Mohseni, A., and Darter, M. I. (1993). "Simplified pavement performance models." Transportation Research Record, 7-7.
Li, N., Haas, R., and Xie, W.-C. (1997). "Development of a new asphalt pavement performance prediction model." Canadian Journal of Civil Engineering, 24(4), 547-559.
Lytton, R. L. "Concepts of pavement performance prediction and modeling." Proc., Proc., 2nd North American Conference on Managing Pavements.
McCulloch, W. S., and Pitts, W. (1943). "A logical calculus of the ideas immanent in nervous activity." The bulletin of mathematical biophysics, 5(4), 115-133.
Molenaar, A. A. "Pavement performance evaluation and rehabilitation design." Proc., Maintenance and Rehabilitation of Pavements and Technological Control.
Mubaraki, M. (2010). "Predicting Deterioration for the Saudi Arabia Urban Road Network." University of Nottingham.
Mubaraki, M., and Sallam, H.-D. (2014). "Predicting Deterioration for the Saudi Secondary Urban Road Network." IRF, 40.
National Research Council . Highway Research, B., and American Association of State Highway, O. (1961). The AASHO road test : history and description of project, National Academy of Sciences-National Research Council, Washington, D.C. :.
National Research Council . Highway Research, B., and American Association of State Highway, O. (1962). The AASHO road test : report 5, pavement research, National Academy of Sciences-National Research Council, Washington, D.C. :.
Noureldin, A. S., Zhu, K., Li, S., and Harris, D. (2003). "Network pavement evaluation with falling-weight deflectometer and ground-penetrating radar." Transportation Research Record: Journal of the Transportation Research Board, 1860(1), 90-99.
Ortiz-García, J. J., Costello, S. B., and Snaith, M. S. (2006). "Derivation of transition probability matrices for pavement deterioration modeling." Journal of Transportation Engineering, 132(2), 141-161.
Park, K., Thomas, N. E., and Wayne Lee, K. (2007). "Applicability of the International Roughness Index as a Predictor of Asphalt Pavement Condition 1." Journal of Transportation Engineering, 133(12), 706-709.
Paterson, W. D. (1987). Road deterioration and maintenance effects: Models for planning and management.
Poch, M., and Mannering, F. (1996). "Negative binomial analysis of intersection-accident frequencies." Journal of Transportation Engineering, 122(2), 105-113.
Prozzi, J., and Madanat, S. (2000). "Using duration models to analyze experimental pavement failure data." Transportation Research Record: Journal of the Transportation Research Board(1699), 87-94.
Prozzi, J. A. (2001). "Modeling pavement performance by combining field and experimental data." University of California Transportation Center.
Rada, G. R. (2013). Guide for Conducting Forensic Investigations of Highway Pavements (with supplemental material on CD-ROM), Transportation Research Board.
Ratkowsky, D. A., and Giles, D. E. (1990). Handbook of nonlinear regression models, Marcel Dekker New York.
Raymond, C., Tighe, S., Haas, R., and Rothenburg, L. (2003). "Development of Canadian asphalt pavement deterioration models to benchmark performance." Canadian Journal of Civil Engineering, 30(4), 637-643.

Richards, F. (1959). "A flexible growth function for empirical use." Journal of experimental Botany, 10(2), 290-301.
Robinson, C., Beg, M. A., Dossey, T., and Hudson, W. R. (1996). "Distress prediction models for rigid pavements for Texas pavement management information system." Transportation Research Record: Journal of the Transportation Research Board, 1524(1), 145-151.
Rowe, G., Baumgardner, G., and Sharrock, M. (2009). "Functional forms for master curve analysis of bituminous materials." Advanced testing and characterization of bituminous materials. CRC Press, Boca Raton, 81-91.
Saarenketo, T., and Scullion, T. (2000). "Road evaluation with ground penetrating radar." Journal of Applied Geophysics, 43(2–4), 119-138.
Sadek, A. W., Freeman, T. E., and Demetsky, M. J. (1996). "Deterioration prediction modeling of Virginia's interstate highway system." Transportation Research Record: Journal of the Transportation Research Board, 1524(1), 118-129.
Saltan, M., Saltan, S., and Şahiner, A. (2007). "Fuzzy logic modeling of deflection behavior against dynamic loading in flexible pavements." Construction and Building Materials, 21(7), 1406-1414.
Saltan, M., and Terzi, S. (2008). "Modeling deflection basin using artificial neural networks with cross-validation technique in backcalculating flexible pavement layer moduli." Advances in Engineering Software, 39(7), 588-592.
Sayers, M. W., Gillespie, T. D., and Paterson, W. D. (1986). Guidelines for conducting and calibrating road roughness measurements.
Shekharan, A. (2000). "Solution of pavement deterioration equations by genetic algorithms." Transportation Research Record: Journal of the Transportation Research Board(1699), 101-106.
Smith, E. P., and Rose, K. A. (1995). "Model goodness-of-fit analysis using regression and related techniques." Ecological Modelling, 77(1), 49-64.
Stantec Consulting Services, I. H. W. L., Inc. (2007). "Development of Performance Prediction Models for Virginia Department of Transportation Pavement Management System, Department of Transportation, Richmond, VA, Virginia".
Tigdemir, M., Karasahin, M., and Sen, Z. (2002). "Investigation of fatigue behaviour of asphalt concrete pavements with fuzzy-logic approach." International Journal of Fatigue, 24(8), 903-910.
Tighe, S., Haas, R., and Ponniah, J. (2003). "Life-cycle cost analysis of mitigating reflective cracking." Transportation Research Record: Journal of the Transportation Research Board(1823), 73-79.
Tran, N., and Hall, K. (2005). "Evaluating the predictive equation in determining dynamic moduli of typical asphalt mixtures used in Arkansas."
Virginia Department of Transportation, C. O., Asset Management Staff (2006). "State of Pavement - Interstate and Primary Flexible Pavements." URL: http://virginiadot.org/info/resources/2006_Condition_Report.pdf
Weibull, W. (1951). "Wide applicability." Journal of applied mechanics, 103.
Zadeh, L. A. (1965). "Fuzzy sets." Information and control, 8(3), 338-353.
Zaghloul, S., He, Z., Vitillo, N., and Kerr, J. (1998). "Project scoping using falling weight deflectometer testing: New Jersey experience." Transportation Research Record: Journal of the Transportation Research Board(1643), 34-43.
Zellner, A. (1971). An introduction to Bayesian inference in econometrics, Wiley New York.
Zhang, G., Patuwo, B. E., and Hu, M. Y. (1998). "Forecasting with artificial neural networks:: The state of the art." International journal of forecasting, 14(1), 35-62.
Zhang, Z., Claros, G., Manuel, L., and Damnjanovic, I. (2003). "Development of structural condition index to support pavement maintenance and rehabilitation decisions at network level." Transportation Research Record: Journal of the Transportation Research Board, 1827(1), 10-17.

APPENDIX A: Calculating Posterior Mean of Empirical Bayesian Update with Corrected Overdispersion Parameters

The probability density functions of the Negative Binomial Distribution, the Poisson distribution, and the Gamma distribution are given in Equation A.1, Equation A.2 and Equation A.3, respectively. An alternative parametrization of the Gamma distribution is given in Equation A.4. The definition of the Negative Binomial distribution in terms of the Poisson and Gamma distributions is given in Equation A.5.

$$f_{NB}(x;r,p)=\frac{\Gamma(r+x)}{x!\Gamma(r)}p^x(1-p)^r \quad \text{(A.1)}$$

$$f_P(x;\lambda)=\frac{\lambda^x}{x!}e^{-\lambda} \quad \text{(A.2)}$$

$$f_G(\lambda;r,p)=\frac{\lambda^{r-1}}{\left(\frac{p}{1-p}\right)^r\Gamma(r)}e^{-\lambda(1-p)/p} \quad \text{(A.3)}$$

$$f_G(x;r,\theta)=\frac{x^{r-1}}{\theta^r\Gamma(r)}e^{-\frac{x}{\theta}} \quad \text{(A.4)}$$

$$f_{NB}(x;r,p)=\int_0^\infty f_P(x;\lambda)f_G(\lambda;r,p)d\lambda \quad \text{(A.5)}$$

The Gamma distribution is the conjugate prior of the Poisson distribution, therefore the posterior distribution is also a Gamma distributions with parameters depending on the prior distribution and the observation *DI* as shown in Equation A.6.

$$f_{G_{Posterior}}\left(\lambda;r+DI,\frac{\theta}{\theta+1}\right)=\frac{\lambda^{r+DI-1}}{\left(\frac{\theta}{\theta+1}\right)^{r+DI}\Gamma(r+DI)}e^{-\lambda(\theta+1)/\theta} \quad \text{(A.6)}$$

A point estimate of the pavement deterioration is the posterior mean which is a weighted average of the prior mean and the observation as shown in Equation A.7 or Equation A.8.

$$\mu_{posterior}=(r+DI)\frac{\theta}{\theta+1}=\frac{r\theta}{\theta+1}+\frac{\theta DI}{\theta+1}=\frac{\mu_{prior}+\theta DI}{\theta+1}=\mu_{prior}(1-p)+pDI \quad \text{(A.7)}$$

$$\mu_{posterior}=\frac{1}{\phi\mu_{prior}+1}\mu_{prior}+\left(1-\frac{1}{\phi\mu_{prior}+1}\right)DI \quad \text{(A.8)}$$

If we label the factor that multiplies μ_{prior} in Equation A.8 by *w*, then we have

$$w = \frac{1}{\phi\mu_{prior} + 1} = \frac{\mu_{prior}}{\phi\mu_{prior}^2 + \mu_{prior}} = \frac{\phi\mu_{prior}^2 + \mu_{prior} - \phi\mu_{prior}^2}{\phi\mu_{prior}^2 + \mu_{prior}}$$
$$= 1 - \frac{\phi\mu_{prior}^2}{\phi\mu_{prior}^2 + \mu_{prior}} = 1 - \frac{\sigma_s^2}{\sigma_s^2 + \sigma_{Poisson}^2} \quad \text{(A.9)}$$

Equation A.8 can be written as show in Equation A.9 which has the same form as the linear EB estimate shown in Equation 43. Therefore, if the variance of the error in the *DI* is larger than the variance predicted from the Poisson model by a factor, α, then we can calculate the posterior mean as shown in Equation A.10. It can be easily shown that this is equivalent to using Equation 41 with ϕ replaced with a corrected overdispersion parameters, ϕ_c as shown in Equation A.12.

$$\mu_{posterior} = w\mu_{prior} + (1-w)DI = \left(1 - \frac{\sigma_s^2}{\sigma_s^2 + \sigma_{Poisson}^2}\right)\mu_{prior} + \left[1 - \left(1 - \frac{\sigma_s^2}{\sigma_s^2 + \sigma_{Poisson}^2}\right)\right]DI$$
$$= \mu_{prior} + \frac{\sigma_s^2}{\sigma_s^2 + \sigma_{Poisson}^2}\left(DI - \mu_{prior}\right) \quad \text{(A.9)}$$

$$\mu_{posterior} = \mu_{prior} + \frac{\sigma_s^2}{\sigma_s^2 + \alpha\sigma_{Poisson}^2}\left(DI - \mu_{prior}\right) \quad \text{(A.10)}$$

$$\phi_c = \frac{\phi}{\alpha} \quad \text{(A.12)}$$

APPENDIX B: Estimating the Error in Predicting the Future Condition through the Model and the Observations

Equation 51 can be rewritten as shown in Equation B.1, where $\frac{1}{n-1}\sum_{i=1}^{n-1}(\varepsilon_{i+1})^2+(\varepsilon_i)^2$, is replaced with $2\hat{\sigma}^2_{Error}$, which is an unbiased estimate of σ^2_{Error}. The quantity, $2\hat{\sigma}^2_{Error}$, cannot be directly estimated however $\frac{1}{n-1}\sum_{i=1}^{n-1}(D_i)^2$ can be directly estimated from the data and can be used as an upper limit (biased upward) estimate of $2\hat{\sigma}^2_{Error}$. We can improve the estimate of $2\hat{\sigma}^2_{Error}$ by estimating $\frac{1}{n-1}\sum_{i=1}^{n-1}(C_{i+1}-C_i)^2$ with $\frac{1}{n-1}\sum_{i=1}^{n-1}(\mu_{\text{Prior}_{i+1}}-\mu_{\text{Prior}_i})^2$ to take into account the effect of pavement deterioration, where μ_{Prior} is the condition given by the Negative Binomial regression model. Note that in general, $(\mu_{\text{Prior}_{i+1}}-\mu_{\text{Prior}_i})$ does not have to be a good estimate of $(C_{i+1}-C_i)$ for individual observations. This is because some pavement sections will be in a condition that is better than what is predicted by the model while other pavement sections will be in a condition that is worse than what is predicted by the model. However, the model represent the average condition of all pavement sections and therefore $\frac{1}{n-1}\sum_{i=1}^{n-1}(\mu_{\text{Prior}_{i+1}}-\mu_{\text{Prior}_i})^2\approx\frac{1}{n-1}\sum_{i=1}^{n-1}(C_{i+1}-C_i)^2$ is a good assumption. With this assumption, the measurement error can be estimated as shown in Equation B.2. Equation B.1 can then be used to independently check that $\hat{\sigma}^2_{Error}$ is consistent with σ^2_{mod} and σ^2_s.

$$\frac{1}{n-1}\sum_{i=1}^{n-1}(D_i)^2=2\hat{\sigma}^2_{Error}+\frac{1}{n-1}\sum_{i=1}^{n-1}(C_{i+1}-C_i)^2 \tag{B.1}$$

$$\hat{\sigma}^2_{Error}\approx\frac{1}{2}\left(\frac{1}{n-1}\sum_{i=1}^{n-1}(D_i)^2-\frac{1}{n-1}\sum_{i=1}^{n-1}(\mu_{\text{Prior}_{i+1}}-\mu_{\text{Prior}_i})^2\right) \tag{B.2}$$

Recall in Equation 51 that $D_i=DI_{i+1}-DI_i$ is a measure of the error made if DI_i (the measurement obtained in year *i*) is used to estimate DI_{i+1} (the measurement the following next year). Suppose instead of , DI_i, we use an estimate DI_i^{est} of C_i in the hope that this estimate will give a better prediction for DI_{i+1}. Clearly the best estimate that will minimize $\frac{1}{n-1}\sum_{i=1}^{n-1}(D_i)^2$ is $DI_i^{est} = C_i$ because in this case, $\varepsilon_i^2 = 0$. This shows that the best possible improvement on the estimate (i.e. in the case we can somehow exactly estimate the pavement condition at a particular year), then the prediction error can be reduced by a maximum amount that is equal to $\hat{\sigma}^2_{Error}$. This scenario is detailed in Equation B.3 with DI_i^{est} set to the empirical Bayes estimate, DI_i^{EB}. Note that the factor $(C_{i+1} - C_i)^2$ is not affected by DI_i^{est} as DI_i^{est} is only attempting to improve the estimate of C_i without any consideration to the deterioration. We can therefore try and estimate the deterioration $(C_{i+1} - C_i)$ by, for example, $(\mu_{i+1} - \mu_i)$, to obtain a better prediction of the future condition. Again the best possible scenario in the case is obtained when $(C_{i+1} - C_i)$ can be exactly estimated. This scenario is detailed in Equation B.4. A natural way to proceed is to combine both approaches to potentially obtain still a better estimate of the future condition. This combined scenario is detailed in Equation B.5. Finally, we can directly estimate the future condition using model predicted condition as shown in Equation B.6.

$$\frac{1}{n-1}\sum_{i=1}^{n-1}\left(D_i^{EB}\right)^2 = \hat{\sigma}^2_{Error} + \frac{1}{n-1}\sum_{i=1}^{n-1}\left(\varepsilon_i^{EB}\right)^2 + \left(C_{i+1} - C_i\right)^2 \geq \hat{\sigma}^2_{Error} + \frac{1}{n-1}\sum_{i=1}^{n-1}\left(C_{i+1} - C_i\right)^2 \qquad \text{(B.3)}$$

$$\frac{1}{n-1}\sum_{i=1}^{n-1}\left(D_i^{\det}\right)^2 = 2\hat{\sigma}^2_{Error} + \frac{1}{n-1}\sum_{i=1}^{n-1} + \left[\left(C_{i+1} - C_i\right) - \left(\mu_{i+1} - \mu_i\right)\right]^2 \geq 2\hat{\sigma}^2_{Error} \qquad \text{(B.4)}$$

$$\frac{1}{n-1}\sum_{i=1}^{n-1}\left(D_i^{combined}\right)^2 = \hat{\sigma}^2_{Error} + \frac{1}{n-1}\sum_{i=1}^{n-1}\left(\varepsilon_i^{est}\right)^2 + \left[\left(C_{i+1} - C_i\right) - \left(\mu_{i+1} - \mu_i\right)\right]^2 \geq \hat{\sigma}^2_{Error} \qquad \text{(B.5)}$$

$$\frac{1}{n}\sum_{i=1}^{n}\left(D_i^{\text{mod}}\right)^2 = \frac{1}{n}\sum_{i=1}^{n}\left(DI_{i+1} - \mu_{i+1}\right)^2 \geq \hat{\sigma}^2_{Error} \qquad \text{(B.6)}$$

Equation B.1 gives the error of predicting the future (next year's) observation using the current observation while Equation B.6 gives the error of predicting future (next year's) observations using the model. Equations B.3 to B.5 combine the model and observations. In Equation B.3 the model and observations are combined using the EB approach which does not take into account pavement deterioration. Equation B.4 uses the available (this year's) observation adjusted by an estimate of the deterioration obtained from the model to predict future (next year's) observation. Equation B.5 combines the approaches used in Equation B.3 and Equation B.4. To compare the different methods, Equation B.4 was used as the reference while

$\hat{\sigma}_{Error}^2$ is the lowest achievable value of the mean square difference between the estimate and observation and is attained when the estimate is equal to the true pavement condition.