Pavement deterioration prediction models in the absence of archived data

by

Abdussalam HEBA

MANUSCRIPT-BASED THESIS PRESENTED TO ÉCOLE DE TECHNOLOGIE SUPÉRIEURE IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY
Ph.D.

MONTREAL, JUNE 29TH 2022

ÉCOLE DE TECHNOLOGIE SUPÉRIEURE
UNIVERSITÉ DU QUÉBEC

BOARD OF EXAMINERS

THIS THESIS HAS BEEN EVALUATED

BY THE FOLLOWING BOARD OF EXAMINERS

Mr. Assaf, Gabriel J. Thesis Supervisor
Département de génie de la construction, École de technologie supérieure

Mr. Saad, Maarouf, President of the Board of Examiners
Département de génie électrique, École de technologie supérieure

Mr. Katsanis, Constantine J., Member of the jury
Département de génie de la construction, École de technologie supérieure

Mr. Bernard Clément, External Evaluator
Département de mathématiques et de génie industriel, polytechnique montréal

THIS THESIS WAS PRENSENTED AND DEFENDED

IN THE PRESENCE OF A BOARD OF EXAMINERS AND PUBLIC

ON JUNE 29TH 2022

AT ÉCOLE DE TECHNOLOGIE SUPÉRIEURE

FOREWORD

To my parents,

To my dear siblings,

To my beloved wife and children……. I dedicate this thesis

ACKNOWLEDGMENT

I would like to thank my supervisor Professor Assaf, Gabriel J. for his consistent support and guidance during the running of this research.

Modèles de prédiction de la détérioration des chaussées en l'absence de données archives

Abdussalam HEBA

RESUME

La planification stratégique et à long terme d`un système de gestion des chaussées (SGC) repose principalement sur des modèles de prévision de détérioration, pour assurer une gestion efficace et préventive et pour définir les besoins budgétaires actuels et futurs. Un SGC comprend de nombreuses activités essentielles au niveau du réseau et des projets. La modélisation de la détérioration des chaussées est l'une des activités les plus importantes au niveau du réseau SGS qui doit être réalisée de manière coordonnée avec d'autres activités pour disposer d'un SGC adapté. Évidemment, l'entretien préventif de la route coûte moins cher que sa reconstruction après avoir été endommagée. Dans de nombreux pays en développement, les routes sont de plus en plus endommagées en raison du manque d'entretien régulier. Cela renforce la nécessité de développer un système de prévision de la détérioration des routes afin de déterminer les stratégies d'intervention optimales pour le réseau routier. De plus, l'analyse de la progression de la détérioration de la chaussée au fil du temps donne un aperçu du comportement fonctionnel de la chaussée pour soutenir efficacement le SGC. En général, les modèles de détérioration des chaussées peuvent être développés de manière déterministe ou probabiliste. Dans des circonstances normales, la modélisation déterministe de la détérioration de la chaussée nécessite des mesures régulières de l'état de la chaussée au fil du temps. Cependant, en l'absence de telles informations dans de nombreux cas, comme dans les pays en développement, une telle méthode ne peut pas être utilisée, et une alternative consiste à utiliser la modélisation probabiliste de l'opinion des experts locaux. Cette recherche présente trois méthodologies pour prédire et analyser l'état de la chaussée et son évolution lorsque les données historiques d'indice de chaussée ne sont pas disponibles.

La première méthode suggérée est une approche probabiliste de régression linéaire bayésienne pour développer un modèle de performance lorsque les données sur l'historique des chaussées

ne sont pas disponibles. Au lieu de cela, le modèle utilise les connaissances d'experts comme une distribution préalable. A ce titre, des experts qui ont longtemps travaillé avec les administrations routières et de transport ont été interrogés pour développer une partie des données d'entrée pour alimenter le modèle Bayésien. La distribution a posteriori a été calculée à l'aide de la fonction d'estimation de la vraisemblance basée sur les inspections de l'état des routes selon un protocole prédéfini. Dans cette étude, les paramètres du modèle ont été estimés et des intervalles de confiance à 95% et ont été établis autour de ces paramètres estimés. Les résultats sont des modèles de prédiction de la détérioration des chaussées basés sur un échantillon de quelques inspections sur place en interaction avec des connaissances d'experts.

La deuxième méthode est une technique probabiliste pour analyser la progression de la détérioration de la chaussée au fil du temps afin de permettre une meilleure compréhension du comportement fonctionnel de la chaussée pour soutenir efficacement le SGC. Le but de cette méthode est d'étudier et de prévoir les tendances de la détérioration de la chaussée à l'aide d'une analyse de séries temporelles par moyenne mobile intégrée autorégressive (ARIMA). Les données utilisées dans cette étude sont l'indice de rugosité international (IRI) qui a été estimé à l'aide d'inspections visuelles de la chaussée. La pertinence des modèles a été évaluée en traçant les résidus de la fonction d'auto-corrélation (ACF) et de la fonction d'auto-corrélation partielle (PACF). Tous les résidus se trouvent dans les bandes, ce qui signifie que tous les modèles sont appropriés. Les statistiques de Ljung – Box (Q) ont été appliquées à tous les modèles et les résultats ont montré que tous les coefficients des modèles étaient significativement différents de zéro. Tous les modèles ajustés sont adéquats car ils ont épousé toutes les données.

La troisième méthode consiste à développer un modèle de prévision de la détérioration pour anticiper les conditions futures de la chaussée à l'aide du modèle de Markov caché (HMM). Cette étude a expliqué comment estimer la séquence la plus probable des états de la chaussée qu'une chaussée spécifique va suivre en se détériorant selon 10 étapes de transition. Une base de données initiale représentant l'état de la chaussée pour une période de temps donnée est utilisée dans le processus de développement. Les probabilités de transition et les probabilités d'émission sont également calculées.

Mots-clés: Système de gestion de la chaussée (SGC), régression linéaire bayésienne, détérioration de la chaussée, indice de rugosité international (IRI), moyenne mobile intégrée autorégressive (ARIMA), modèle de Markov caché (HMM), matrice de transition de probabilité.

Pavement deterioration prediction models in the absence of archived data

Abdussalam HEBA

ABSTRACT

Strategic and long-term planning in Pavement Management System (PMS) relies mainly on deterioration prediction models, to ensure efficient and forward-looking management and for setting present and future budget requirements. PMS consist of many essential network and project levels activities. Modeling of pavement deterioration is one of the most important PMS network level activities that must be coordinately executed with other activities to have a functional PMS. Obviously, preventive maintenance of a pavement is less expensive than reconstructing it after being deteriorated. In many developing countries, roads face increasing damage because of the lack of regular maintenance. This reinforces the need to develop a system to predict the deterioration of roads in order to determine the Optimal Intervention Strategies (OIS) for the road network. Additionally, analyzing the progression of pavement deterioration over time enables better understanding for the pavement functional behavior to efficiently support a PMS. In general, pavement deterioration models can be developed deterministically or probabilistically. Under normal circumstances, pavement deterministic deterioration modelling requires regular measurements of the pavement condition over time. However, in the absence of such information and records in many cases such as in developing countries, such method cannot be used, and alternative is to use probabilistic modeling. This research presents three methodologies to predict and analyze pavement condition and its progression when archived pavement indices data is not available.

First suggested method is a probabilistic approach of Bayesian linear regression to develop a deterioration model when archived data about pavement history is not available. Instead, the model uses expert knowledge as a prior distribution. As such, experts who have worked for a long time with the road and transportation agencies have been interviewed to develop a portion of the input data to feed the Bayesian model. The posterior distribution was calculated using the likelihood estimation function based on road condition inspections according to a predefined protocol. In this study, model parameters were estimated, and 95% confidence intervals established around these estimated parameters. The results are forecasting models of pavement deterioration prediction model based on a mixture of few on-site inspections interacting with expert knowledge.

Second method is a probabilistic technique to analyze the progression of pavement deterioration over time to enable better understanding for the pavement functional behavior to efficiently support a PMS. The aim of this method is to investigate and forecast the trends of the pavement deterioration using Autoregressive Integrated Moving Average (ARIMA) time series analysis. Data used in this study is the International Roughness Index (IRI) which was estimated using visual pavement inspections. Models' appropriateness are evaluated by plotting the residuals of Auto Correlation Function ACF and Partial Auto Correlation Function PACF. All residuals were within the bands which meant that all models are appropriate. Ljung–Box (Q) statistics was applied for all models and the results showed that all model coefficients are not significantly different from zero.

The third method was to develop a deterioration prediction model to anticipate future pavement conditions using a Hidden Markov Model (HMM). This study explained how to estimate the most likely sequence of pavement condition states that a specific pavement goes through to failure, using 10 transitioning steps to run the pavement HMM. Viterbi algorithm was used to compute this sequence of pavement condition states. An initial database representing the pavement condition for a given period of time is used in the development process. Transition probabilities and emission probabilities are also calculated.

Keywords: Pavement Management System (PMS), Bayesian linear regression, Pavement deterioration, International Roughness Index (IRI), Autoregressive Integrated Moving Average (ARIMA), Hidden Markov Model (HMM), Probability Transition Matrix.

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

Page

LIST OF ABREVIATIONS

ACF	Autocorrelation function
AR	Autoregressive
BIC	Bayesian information criterion
C-LTPP	Canadian long-term pavement performance
CSHRP	Canadian Strategic Highway Research Program
EM	Emission matrix
HMM	Hidden Markov model
IRI	International roughness index
LCL	Lower credibility level
MA	Moving average
MCMC	Monte Carlo Markov model
MLE	Maximum likelihood estimation
PACF	Partial autocorrelation function
PDF	Probability density function

PMS	Pavement management system
PSI	Pavement serviceability index
PSR	Pavement serviceability rating
RCI	Riding Comfort Index
SE	Standard error
SSE	Error sum of squares
SSR	Regression sum of squares
SST	Total sum of squares
TPM	Transition probability matrix
UCL	Uber credibility level

LIST OF SYMBOLS

$\hat{Y}_i$	Y Estimated value
$\hat{y}_t$	ARIMA model estimated value
$\hat{\alpha}_i$	α Estimated value
$\hat{\beta}_i$	β Estimated value
A_k	District random variables
$\acute{P}(x)$	n degree polynomial first order differentiation
R^2	Determination coefficient
$TPM_{n\times n}$	The one-step transition probability matrix
$\bar{X}$	X Average value
X_i	Independent variable (time)
$\bar{Y}$	Y Average value
Y_i	Dependent variable (pavement condition indices)
y_p	The coefficient for the lagged variable in time $t-p$
α_i	Regression parameters

β_i	Regression parameters
ε_i	Prediction error
θ_i	Lagged error coefficient
φ_i	ARIMA model constants
ϵ_i	Model error
$\propto$	Proportional to operator
A, B	Regression coefficients
A_i	Model parameter
B	Arbitrary event $P(B) > 0$
e	Mechanical wheel loading
E_i	Model error
n	Number of observed data points
p,d,q	ARIMA model parameter order
X	Observed value of a random variable
X_i	*PSR* correlates

Y	*PSR* average
AGE	Pavement age
$E(Y_i)$	Estimated value
N	Asphalt pavement fatigue life
$P(x)$	*n* degree polynomial
S	Y_i sum square of errors
$V(t)_{n\times 1}$	The predicted condition state matrix at year *t*
α	Constant controls the pavement age variable
β	Curve bend constant
θ	Prior variable
μ	Moving average terms constant
ρ	Location constant
σ	Standard deviation
A_k	District random variables
B	Arbitrary event $P(B) > 0$

XXX

θ	Prior variable
X	Observed value of a random variable
$\propto$	Proportional to operator
$V(t)_{n\times 1}$	The predicted condition state matrix at year *t*
$TPM_{n\times n}$	The one-step transition probability matrix
p,d,q	ARIMA model parameter order
φ_i	ARIMA model constants
y_p	The coefficient for the lagged variable in time $t-p$
θ_i	Lagged error coefficient
ϵ_i	Model error
μ	Moving average terms constant
$\hat{y}_t$	ARIMA model estimated value
IRI_j	Road *i* section condition
t_i	Time *i*
w	Prior distribution vector

σ	*IRI* standard deviation
a	$\frac{1}{\sigma^2}$
b	*IRI* covariance matrix
D	*IRI* data
Λ	*w* MLE estimations
ΔIRI_t	Differenced variable
α_f, β_f	Pavement family *f* ARIMA model parameters
$e_{t,f}$	Pavement family *f* ARIMA model error
H_0	Null hypothesis
H_1	Alternative hypothesis
ρ_i	Pavement *i* autocorrelation
$\chi^2_{1-\alpha,df}$	Chi square table value
Q	Chi square calculated value
df	Degree freedom
Z_n	Pavement hidden deterioration variable

X_n	Pavement observed deterioration variable
π_{ij}	TPM (*i,j)* element
$\varepsilon_i(x)$	Emission probabilities
$cov(\mu, \sigma)$	(μ, σ) Covariance
$\mu\sigma_{std.Err}$	Medium standard deviation error

INTRODUCTION

Context

Under the combination of traffic load, pavement materials, and environment, road pavements deteriorate gradually, reducing their quality of ride (Shahin, 2005). Pavement performance models are used to predict pavement deterioration indicators including future condition (Haas, 1994). Pavement deterioration models are a comprehensive term that expresses how pavement conditions deteriorate under different use circumstances over time (N. Y. Li, Haas, & C Xie, 2011). In addition, pavement performance models must reflect the best possible representation of pavement deterioration. Thus, they can be used to anticipate pavement future condition as part of pavement management system (PMS) activities (Shahin, 2005). Prediction models in PMS, are the milestone for many activities such as: determining the pavement remaining time of service; predicting best pavement maintenance or rehabilitation intervention strategies; analyzing the pavement life cycle cost (Amin, 2015).

Pavement performance prediction models can be developed using two methods: deterministic and probabilistic (Kobayashi, Kaito, & Lethanh, 2012; Ralph, W. Ronald, & Lynne, 2015). Deterministic models predict a single number for the life of a pavement, its level of distress, or any other measure of its condition. In contrast, probabilistic models predict a distribution of such events. There are many probabilistic techniques used to develop pavement performance models some of them are Bayesians, Markovian and time series analysis and forecasting. Each technique depends primarily on data availability and quality (Z. Li, 2005). For instance, Bayesian modeling is considered a proper technique when historical data is insufficient. In this case, experts' knowledge is used to recover historical data limitation (Feng Hong & Prozzi, 2006). Hidden Markov Model (HMM) can be used to know the sequence of conditions that a pavement section has gone through to reach a specific condition. Pavement roughness indices are equally spaced time-correlated values so their general trend can be investigated and analyzed. Consequently, prediction model can be developed to forecast future using time series analysis and forecasting (Lethanh & Adey, 2013).

Many studies using various methods have developed large number of pavement prediction models. However, most of these methods were based on real archived data.

This research aims to develop a methodology to predict pavement future condition when historical data on road network sections is insufficient or unavailable. This is the specific scientific contribution of this research.

In such a situation, probabilistic techniques are considered appropriate and able to accommodate the limitations of the research data.

Probabilistic and deterministic models require however records of pavement condition indices which is not available for this study. Because of the lack of archived data and the absence of any type of pavement performance measurements, three probability approaches were used. To illustrate the effectiveness of these methods, the Libyan road network was chosen as a case study.

Research Objectives

The research objectives are:

1. Develop a pavement performance prediction methodology when no data is available on road condition based solely on local expert interviews.
2. Develop a pavement roughness prediction model based solely on visual observation of the road condition.
3. Develop a pavement Markovian deterioration model to investigate pavement condition transitions.

Organization of the Thesis

The thesis is organized as follows. Chapter 1 is the literature review. In this chapter, the general concept of pavement management systems is discussed, with emphasis and literature review on pavement prediction models. Additionally, most causes and types of pavement deterioration are defined. Pavement roughness indices are also defined and discussed. Moreover, the deterministic and probabilistic pavement prediction models and the techniques and priorities for selecting the appropriate model according to the type, quality, and availability of the data are summarized. Bayesian prediction methods, Markov chains, transition matrices, and Autoregressive Integrated Moving Average time series models are presented.

Chapter 2 presents a methodology to develop new models for various selected pavement sections to predict their future condition. The predicted future condition of the pavements could be used in estimating their remaining service life to failure, which will consequently be used to help find the best ways to intervene in the maintenance and rehabilitation (M&R) activities for a given network. In this chapter, a Bayesian linear regression is used to come over the lack of archived data, by replacing it with the knowledge of experts, and use it as the model prior distribution.

A Markovian prediction model is presented in Chapter 3 and how it is applied in the PMS network level to predict pavement future condition. In the pavement deterioration process, pavements go through multiple condition state until they reach the failure state. It is easy to approximately know when pavements fail and become unusable from their life spin. However, it is difficult to know the condition states that a specific pavement went through to fail. This Chapter explains how to estimate the most likely sequence of pavement condition states that a specific pavement has went through to fail. A 10 transitioning steps were taken to run the pavement hidden Markov model (HMM). however, it is recommended to use several transitioning steps that are equal or close to the pavement life spin.

Pavement deterioration model using time series forecasting concept is discussed in Chapter 4. The aim of this chapter is to investigate the general trend of the International Roughness Index (*IRI)* and develop a prediction model to forecast the pavement future condition. The research *IRI* data is nonstationary, a differencing process is applied to transform the data to stationary status. The fitted ARIMA model was used to forecast values of the IRI for years from 2021 up to 2025. The results of the ARIMA analysis showed that the forecasted values of IRI are in an increasing trend.

CHAPTER 1

LITERATURE REVIEW

1.1 Pavement Management System (PMS)

A Pavement Management System (PMS) is an effective and efficient process of a multiple activities involved in providing and sustaining pavements in a condition acceptable to the traveling public at the least life cycle cost (Haas, 1994). A Pavement Management System is a set of defined procedures for collecting, analyzing, maintaining, and reporting pavement data, to support the decision makers to reach optimum strategies for maintaining pavements in serviceable condition over a given period of time for the least cost (Shahin, 2005).

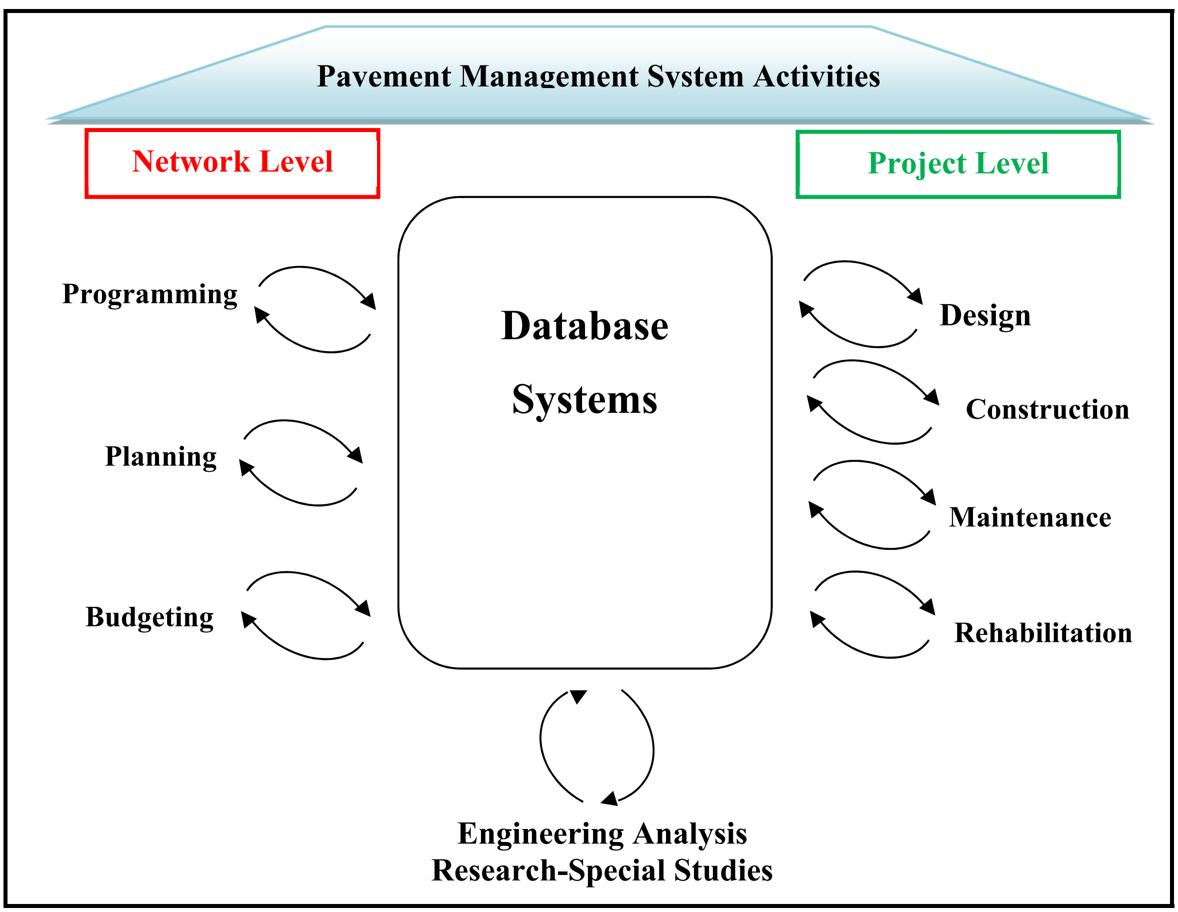

Figure 1.1 Pavement Management System Activities. Taken from Haas (1994, p. 14)

A PMS is designed to provide objective information and useful data for analysis so that road managers can make more consistent, cost-effective, and defensible decisions related to the preservation of a pavement network. While a PMS cannot make final decisions, it can provide the basis for an informed understanding of the possible consequences of alternative decisions (Haas, 1994).

1.1.1 Pavement Management Systems Components

A PMS consists of several interacting components which are planning, programming, design, construction, maintenance, and rehabilitation. These components are mainly functioning under two operational levels: network and project as shown in Fig 1.1.

1.1.1.1 Network level

The main task of network management level is to develop a prioritized program and work schedule, within overall available budget and legislations constraints. The primary goal of the pavement management network level is to optimize the cost-benefit ratio for the entire network. However, this requires coordination with project level decisions. When projects are already planned before high-level decisions are made, project-level decisions and priorities may not be consistent with network-level decisions and priorities. This can lead to a suboptimal system solution that is driven by individual project-level decisions instead of network-level decisions (Ralph et al., 2015).

In network level, conditional scenarios can be produced quickly and accurately using software tools. These tools allow the user to adjust top-level budget and policy inputs and then quickly calculate the resulting network-wide effects because these models are driven by top-level (network-level) decisions.

Network-level analysis can determine pavement Maintenance and Rehabilitation (M&R) treatments and costs; these targets can be easily and consistently applied to individual projects. In order to accomplish the same thing with a project-level approach, network-level targets need to be provided in advance such that project-level decision can be made with network-level targets in mind. As a result, broader areas of maintenance, rehabilitation and reconstruction can be prioritized.

1.1.1.2 Project level

Project level presents the actual physical implementation of the network decisions. All project level tasks must be consecutive in accordance with the assigned schedule. In contrast with network level, project level approach requirements are considered less complicated. Project level needs fewer data and resource requirements, less reliance on feedback for success and better understanding. Strong communication between network level and project level elevates the quality of decisions made by management.

1.2 Pavement Deterioration

Pavement deterioration is the process by which distress develop in the pavement under the combined effects of traffic loading and environmental conditions (refer to Fig, 1.2). The four major categories of common asphalt pavement surface distresses are: cracking, surface deformation, disintegration, and surface defects.

There are many possible causes of deterioration in pavements which can vary from one geographic area to another some of them are: sudden increase in traffic loading especially on new roads where the design is based on lesser traffic is a major cause of cracking. After construction of good road, traffic of other roads also shifts to that road. This accelerates the fatigue failure. Secondly, temperature variation ranging leads to bleeding and cracking. Third, provision of poor shoulders leads to edge failures (Shahin, 2005).

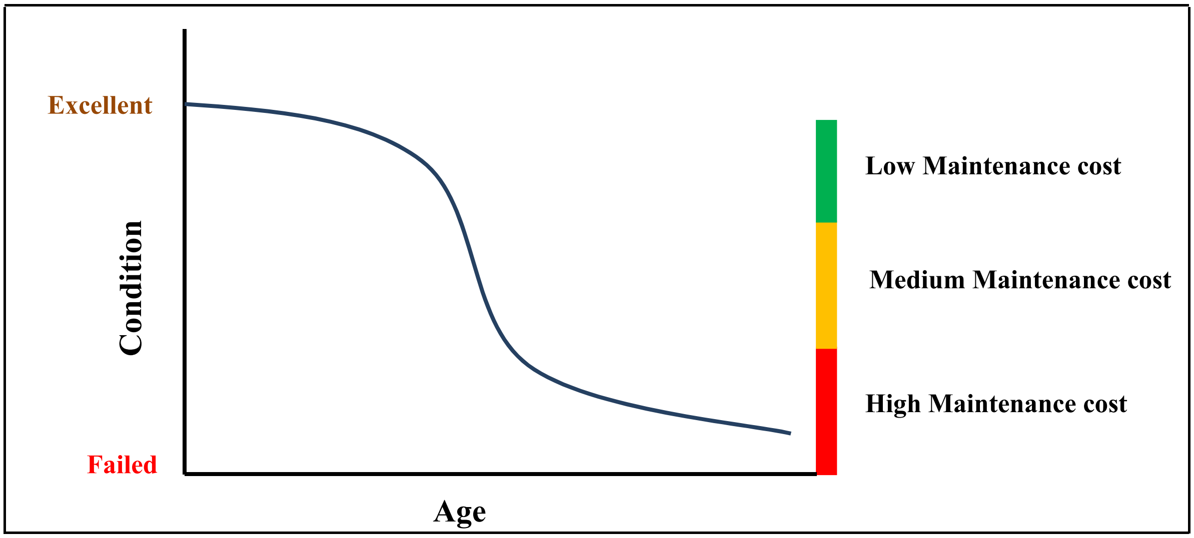

Figure 1.2 Typical Pavement Deterioration Curve. Taken from Haas (1994, p. 193)

Forth, provision of poor clayey subgrade results in corrugation at the surface and increase in unevenness. Fifth, poor drainage conditions especially during rainy seasons, force the water to enter the pavement from the sides as well as from the top surface. In case of open graded bituminous layer, this phenomenon becomes more dangerous, and the top layer gets detached from the lower layers. Sixth, if the temperature of bitumen/bituminous mixes is not maintained properly, then it also leads to pavement failure. Overheating of bitumen reduces the binding property of bitumen. If the temperature of bituminous mix has been lowered down, then the compaction will not be proper leading to longitudinal corrugations (Zumrawi, 2016).

It is clear from the number of possible deteriorations causes and mechanisms that pavement deterioration is not a straightforward process to measure and monitor. It is therefore necessary to make sure that measurements and investigations are aimed at finding out the deterioration causes and the pavement condition in future. In this way maintenance can be targeted. That is why; one of the most important activities in PMS is to develop pavement performance models to anticipate the future condition of the pavements. And therefore, road agencies will be able to a make a proper decision to make an optimum intervention strategy for maintenance or rehabilitation (Sultan Tarawne & Sarireh, 2013).

1.2.1 Pavement Roughness Indicators

Pavement roughness is generally defined as an expression of irregularities in the pavement surface that adversely affect the ride quality of a vehicle (and thus the user). Roughness is an important pavement characteristic because it affects not only ride quality but also vehicle delay costs, fuel consumption and maintenance costs. Several roughness indices are currently in use by highway and airport agencies. These indices are based either on pavement surface profile or on output from a road meter installed in a vehicle. Roughness is typically quantified using many measurement tools (Michael W. Sayers, Thomas D. Gillespie, & Queiroz, 1986b).

1.2.2 International Roughness Index (IRI)

International Roughness Index was developed in 1986 using the results of the International Road Roughness Experiment performed in Brazil in 1982 (Michael W. Sayers, Thomas D. Gillespie, & Queiroz, 1986a). Since then, the IRI has become a well-recognized standard for the measurement of road roughness (Du, Liu, Wu, & Jiang, 2014).The main advantages of the IRI are that it is stable over time and transferable throughout the world (Michael W. Sayers et al., 1986a). The index measures pavement roughness in the wheel paths in terms of the number of rough meters per kilometer. The most common method uses a laser that is mounted on a specialized van. The laser is trained on the road surface, like a laser pointer. As the van drives along a road, the beam jumps unexpectedly at rough patches, just as a laser pointer; these jumps are measured and used for analysis (Mašović & Hajdin, 2013). The lower the IRI number at a given speed, the smoother the ride felt by the road user. Moreover, this roughness statistic is suitable for any road surface type and covers all levels of roughness(Kobayashi et al., 2012). IRI can be treated as a random variable and therefore it can be described as a probability distribution (Shahin, 2005). IRI can also be used as a measure of pavement conditions and the data can be easily shared among researchers. It can also be directly related to vehicle operating costs (Shahin, 2005). Figure 1.3 shows the IRI range scaling pavement condition from new pavement to damaged.

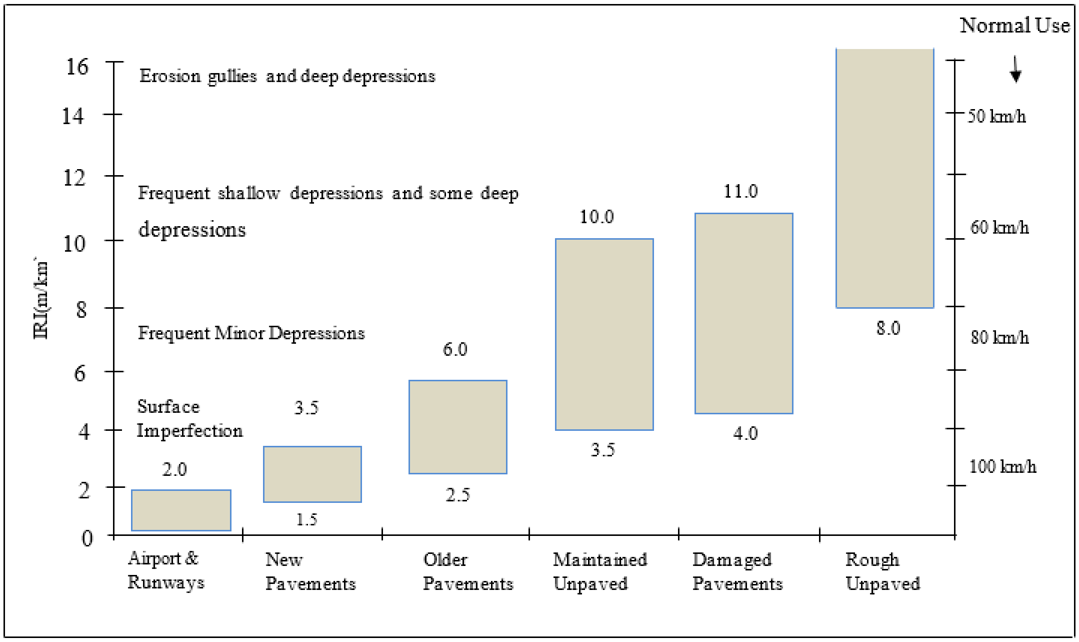

Figure 1.3 IRI Roughness Scale. Taken from Shahin (2005, p. 98)

1.2.2.1 Present Serviceability Rating (PSR)

The PSR of a pavement is a user's judgment of the level of service that a particular pavement provides at a point in time. In the AASHO Road Test studies, raters were asked to judge present serviceability in 1 of 5 categories: 4.0 to 5.0, very good; 3.0 to 4.0, good; 2.0 to 3.0, fair; 1.0 to 2.0, poor; or 0.0 to 1.0, very poor (see Fig. 1.4). It was found that the average PSR given by a panel of raters was reproducible within and among various panels (AASHO, 1962).

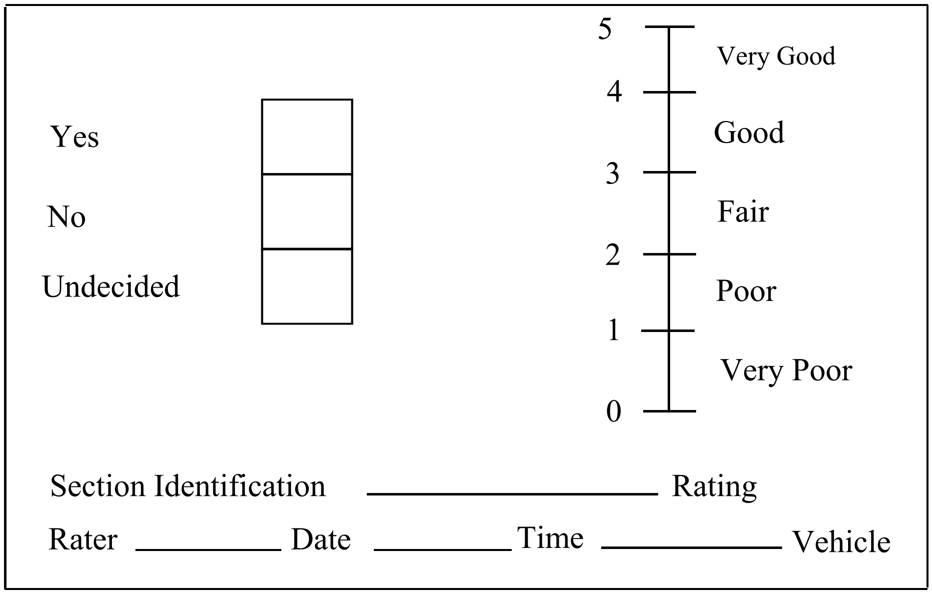

Figure 1.4 Reproduction of an Individual Present Serviceability Rating Form. Taken from Haas (1994, p. 80)

1.2.2.2 Present Serviceability Index (PSI)

A PSI was defined to be an algebraic function of PSR correlates. Moreover, the PSI concept incorporated the view that coefficients in the function should be determined through multiple regression analysis of the form as in Eq. 1.1:

$$Y = A_0 + A_1X_1 + A_2X_2 + \cdots + E \tag{1.1}$$

Where *Y* is an average *PSR* for a pavement section; X_i are *PSR* correlates; $A_0 + A_1X_1 + A_2X_2 + \cdots$ is a *PSI* determined by the regression analysis; and *E* is a discrepancy between the *PSR* and the *PSI* (see Eq. 1.2). In other words,

$$PSR = PSI + E \tag{1.2}$$

Multiple regression procedures were used to find the set of correlates (X_i) that minimized the sum of E^2 and that excluded correlates that did not contribute significantly to the goodness of fit provided by those correlates that were included (Prozzi & Madanat, 2003).

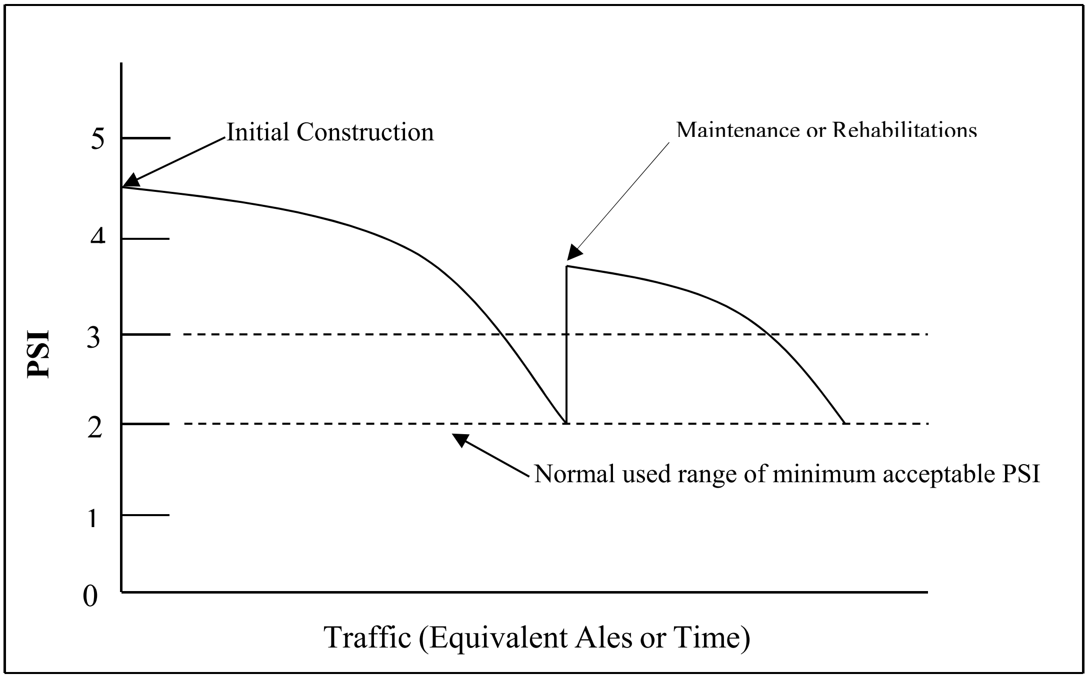

Figure 1.5 Concept of pavement performance using Present Serviceability Index (PSI). Taken from Haas (1994, p. 215)

1.2.3 Pavement Deterioration Causes

There are four major categories of common asphalt pavement surface distresses are:

1.2.3.1 Cracking

Cracks occur directly over the underlying rigid pavement joints. There are several different types of cracks that can develop on pavements.

- **Fatigue Cracking (Alligator cracking)**

Fatigue cracking is commonly called alligator cracking. This is a series of interconnected cracks creating small, irregular shaped pieces of pavement. It is caused by failure of the surface layer or base due to repeated traffic loading (fatigue). Eventually the cracks lead to disintegration of the surface. Alligator cracking is usually associated with base or drainage problems. Small areas may be fixed with a patch or area repair. Larger areas require reclamation or reconstruction. Drainage must be carefully examined in all cases (Shiferaw Wayessa & Abuye, 2019).

- **Longitudinal cracking**

Longitudinal cracks are long cracks that run parallel to the center line of the roadway. These may be caused by frost heaving or joint failures or they may be load induced. Understanding the cause is critical to selecting the proper repair. Multiple parallel cracks may eventually form from the initial crack (S.S.Adlinge & A.K.Gupta, 2013).

- **Transverse cracking**

Transverse cracks form at approximately right angles to the centerline of the roadway. They are regularly spaced and have some of the same causes as longitudinal cracks. Transverse cracks will initially be widely spaced. They usually begin as hairline or very narrow cracks and widen with age. If not properly sealed and maintained, secondary or multiple cracks develop, parallel to the initial crack. The reasons for transverse cracking, and the repairs, are similar to those for longitudinal cracking (Zumrawi, April 2016).

- **Block cracking**

Block cracking is an interconnected series of cracks that divides the pavement into irregular pieces. This is sometimes the result of transverse and longitudinal cracks intersecting. They can also be due to lack of compaction during construction. Low severity block cracking may be repaired by a thin wearing course (Reem Alaamri, Rafeeq Kattiparuthi, & Koya, 2017).

- **Slippage cracking**

Slippage cracks are half-moon shaped cracks with both ends pointed towards the oncoming vehicles. They are created by the horizontal forces from traffic. They are usually a result of poor bonding between the asphalt surface layer and the layer below.

The lack of a tack coat is a prime factor in many cases. (Shiferaw Wayessa & Abuye, 2019).

- **Reflective cracking**

Reflective cracking occurs when a pavement is overlaid with hot mix asphalt concrete and cracks reflect up through the new surface. It is called reflective cracking because it reflects the crack pattern of the pavement structure below. As expected from the name, reflective cracks are covered over cracks reappearing in the surface (Sultan Tarawne & Sarireh, 2013).

- **Edge cracking**

Edge cracks typically start as crescent shapes at the edge of the pavement. They will expand from the edge until they begin to resemble alligator cracking. This type of cracking results from lack of support of the shoulder due to weak material or excess moisture. They may occur in a curbed section when subsurface water causes a weakness in the pavement. (S.S.Adlinge & A.K.Gupta, 2013).

1.2.3.2 Surface Deformation

Pavement deformation is the result of weakness in one or more layers of the pavement that has experienced movement after construction. The deformation may be accompanied by cracking. Surface distortions can be a traffic hazard (Haas, 1994). The basic types of surface deformation are:

- **Rutting**

Rutting is the displacement of pavement material that creates channels in the wheel path. Very severe rutting will hold water in the rut. Rutting is usually a failure in one or more layers in the pavement. The width of the rut is a sign of which layer has failed. A very narrow rut is usually a surface failure, while a wide one is indicative of a subgrade failure. Inadequate compaction can lead to rutting (Zumrawi, April 2016).

- **Corrugation**

Corrugation is referred to as wash boarding because the pavement surface has become distorted like a washboard. The instability of the asphalt concrete surface course may be caused by too much asphalt cement, too much fine aggregate, or rounded or smooth textured coarse aggregate. Corrugations usually occur at places where vehicles accelerate or decelerate (Sultan Tarawne & Sarireh, 2013).

- **Shoving**

Shoving is also a form of plastic movement in the asphalt concrete surface layer that creates a localized bulging of the pavement. Locations and causes of shoving are similar to those for corrugations (S.S.Adlinge & A.K.Gupta, 2013).

- **Depressions**

Depressions are small, localized bowl-shaped areas that may include cracking. They are typically caused by localized consolidation or movement of the supporting layers beneath the surface course due to instability (Sultan Tarawne & Sarireh, 2013).

- **Swell**

A swell is a localized upward bulge on the pavement surface. Swells are caused by an expansion of the supporting layers beneath the surface course or the subgrade. The expansion is typically caused by frost heaving or by moisture. Subgrades with highly plastic clays can swell in a manner similar to frost heaves (Reem Alaamri et al., 2017).

1.2.3.3 Disintegration

The progressive breaking up of the pavement into small, loose pieces is called disintegration. If the disintegration is not repaired in its early stages, complete reconstruction of the pavement may be needed (Shahin, 2005). The two most common types of disintegration are:

- **Potholes**

Potholes are bowl-shaped holes like depressions. They are a progressive failure. First, small fragments of the top layer are dislodged. Over time, the distress will progress downward into the lower layers of the pavement. Potholes are often located in areas of poor drainage and then are formed when the pavement disintegrates under traffic loading, due to inadequate strength in one or more layers of the pavement, usually accompanied by the presence of water (Shiferaw Wayessa & Abuye, 2019).

- **Patches**

A patch is defined as a portion of the pavement that has been removed and replaced. Patches are usually used to repair defects in a pavement or to cover a utility trench. Patch failure can lead to a more widespread failure of the surrounding pavement (S.S.Adlinge & A.K.Gupta, 2013).

1.2.3.4 Surface Defects

Surface defects are related to problems in the surface layer (Zumrawi, April 2016). The most common types of surface distress are:

- **Raveling**

Raveling is the loss of material from the pavement surface. It is a result of insufficient adhesion between the asphalt cement and the aggregate. Initially, fine aggregate breaks loose and leave small, rough patches in the surface of the pavement. As the disintegration continues, larger aggregate breaks loose, leaving rougher surfaces. Raveling can be accelerated by traffic and freezing weather. Some raveling in chip seals is due to improper construction technique (Sultan Tarawne & Sarireh, 2013).

- **Bleeding**

Bleeding is defined as the presence of excess asphalt on the road surface which creates patches of asphalt cement. Excessive asphalt cement reduces the skid-resistance of a pavement, and it can become very slippery when wet, creating a safety hazard. (Shiferaw Wayessa & Abuye, 2019).

- **Polishing**

Polishing is the wearing of aggregate on the pavement surface due to traffic. It can result in a dangerous low friction surface. A thin wearing course will repair the surface (Kajner, 1995).

1.3 Pavement Performance Models

A road pavement deteriorates under the combined action of traffic loading and environment, thus reducing the quality of the ride (Madanat, Karlaftis, & McCarthy, 1997). Models should be able to quantify the contribution of variables such as strength of pavement materials, traffic, and environmental conditions that are relevant to pavement deterioration (Ortiz-Garcia, Costello, & Snaith, 2006). PMS are commonly used to select maintenance strategies that result in lower project life cycle costs (Haas, 1994; Premkumar & Vavrik).

It is clear from the number of possible deteriorations causes and mechanisms that pavement deterioration is not a straightforward thing to measure and monitor. Modeling the performance of pavements is an important activity in pavement management, and many highway agencies have developed a variety of pavement performance models for use in their pavement management activities (Lethanh, Adey, & Fernando, 2015). This research presents a methodology to develop new models for the various pavement families in the Libyan road network in order to predict the condition of a given area of pavement. The predicted future condition of the pavements is used in estimating its remaining service life to failure, which will consequently be used to help find the best ways to intervene in the maintenance and rehabilitation activities for a given area of the network (N. Li, Haas, & Xie, 1997).

There are mainly two basic kinds of performance models: deterministic and probabilistic (Kobayashi et al., 2012). The deterministic models predict a single number for the life of a pavement or its level of distress or any other measure of its condition. In contrast, the probabilistic models predict a distribution of such events. There are many deterministic models some of them are Mechanistic, Empirical-Mechanistic, Polynomial Constrained Least Squares, and S-Shaped Curve models(T. Hong et al., 2013). In general, probabilistic models include Bayesian and Markov process models. Bayesian modeling assigns a prior probability distribution to pavement condition based on experience; it then mixes it with the experimentation and data collection to predict the future condition (Feng Hong & Prozzi, 2006). Furthermore, Markov models can be used when pavement data is a sequence of conditions in which the probability of each condition depends only on the state attained in the previous condition (Z. Li, 2005; Prozzi & Madanat, 2003).

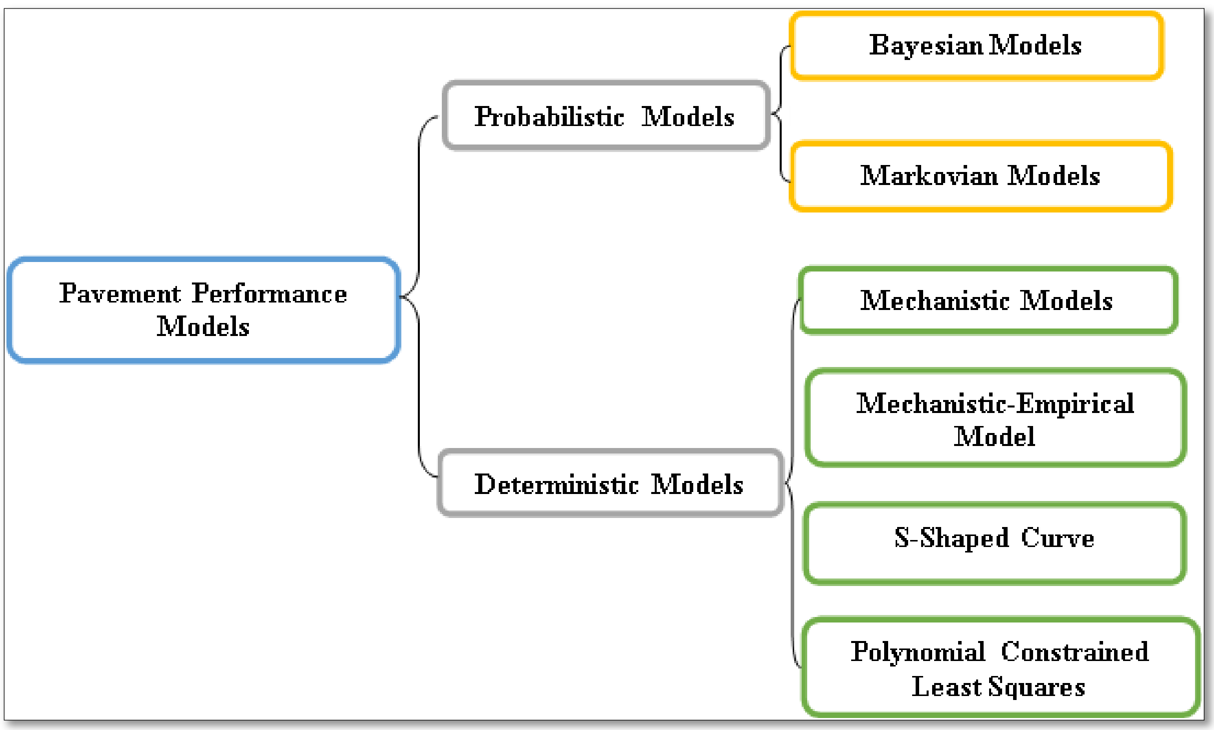

Figure 1.6 General Pavement Performance Models. Taken from Ningyuan (2014, p. 180-193)

1.3.1 Deterministic Models

Deterministic models are the models in which no randomness is involved in the development of future states of the system. A deterministic model will thus always produce the same output from a given starting condition or initial state (Amin, 2015).

1.3.1.1 Mechanistic Models

Mechanistic Model sis also called regression technique because regression analysis is used to establish an empirical relationship between two or more variables. Each variable is described in terms of its mean and variance. Several forms of regression analysis are used, and the simplest form is linear regression between two variables; the model is described in Eq. 1.3.

$$Y_i = \alpha + \beta X_i + \varepsilon_i \tag{1.3}$$

Y is a dependent variable which represents the pavement condition indices. *X* is explanatory or independent variable that represents time since last major rehabilitation.ε is the prediction error. Finally, α and β are the regression parameters. The estimated value (mean) of Y_i which is $E(Y_i)$ for each value of X_i, can be thus determined as shown in Eq 1.4:

$$E(Y_i) = \hat{Y}_i = \hat{\alpha}_i + \hat{\beta} X_i \tag{1.4}$$

Where $\hat{Y}_i$, $\hat{\alpha}_i, and\ \hat{\beta}$ are respectively the estimations of *Y, α, and* β. The values of $\hat{\alpha}$ and $\hat{\beta}$ are determined so as to minimize the sum square of errors of the observed values Y_i from their estimate $\hat{Y}_i$, that is, minimize *s* is given as in Eq. 1.5.

$$S = \sum_{i=1}^{n} \left[Y_i - \hat{Y}_i\right]^2 \ = \sum_{i=1}^{n} \left[Y_i - \hat{\alpha} - \hat{\beta} X_i\right]^2 \tag{1.5}$$

Where n is the number of observed data points. The above method is known as the method of (least squares). The values of $\hat{\alpha}$ and $\hat{\beta}$are determined by setting the partial derivation of *s* with respect to $\hat{\alpha}$ and $\hat{\beta}$ as follows:

$$\hat{\alpha} = \bar{Y} - \hat{\beta}\bar{X} \tag{1.6}$$

$$\widehat{\beta} = \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{\sum(X_i - \bar{X})^2} \tag{1.7}$$

Where $\bar{X}$ and $\bar{Y}$ are the average values of *X* and *Y*, respectively. $\widehat{\alpha}$ is the intercept of the line that measures the estimated value of *Y* corresponding to a value of *X* equal to zero. $\widehat{\beta}$ is the slope of line that measures the estimated value of *Y* corresponding to a unit change in the value of *X* (N. Balakrishnan, Enrique Castillo, & Sarabia, 2006). Fig. 1.7 shows the regression concept and how it is possible to prove mathematically that:

$$\sum_{i=1}^{n}(Y_i - \bar{Y})^2 = \sum_{i=1}^{n}\left(\hat{Y}_i - \bar{Y}\right)^2 + \sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2 \quad (1.8)$$

$$\sum(Y_i - \bar{Y})^2 = total\ sum\ of\ squares\ (SST) \quad (1.9)$$

$$\sum\left(\hat{Y}_i - \bar{Y}\right)^2 = regression\ sum\ of\ squares\ (SSR) \quad (1.10)$$

$$\sum\left(Y_i - \hat{Y}_i\right)^2 = error\ sum\ of\ squares\ (SSE) \quad (1.11)$$

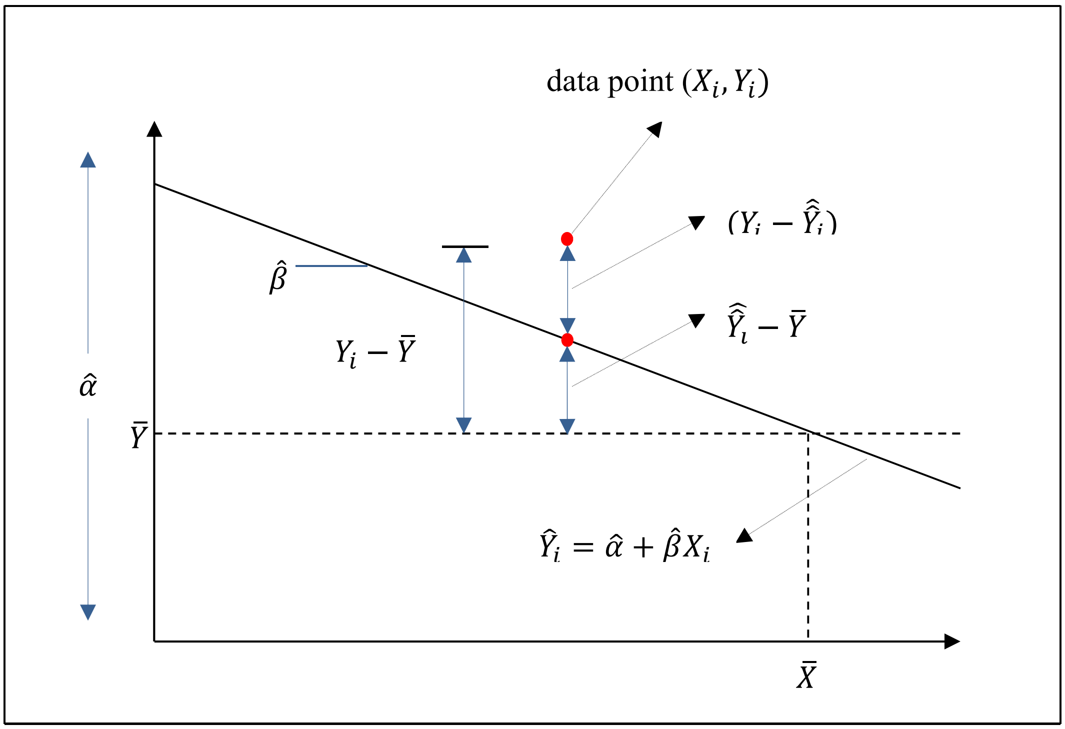

Figure 1.7 Regression Line. Taken from Davison (2003, p. 162)

The advantage of fit the regression line can be measured using the coefficient of determination R^2 , which means the proportion of total variation about the mean $\bar{Y}$, which explained by regression:

$$R^2 = \frac{SSR}{SST} = \frac{\sum(\hat{Y}_i - \bar{Y})^2}{\sum(Y_i - \bar{Y})^2} = \frac{\hat{\beta}^2\sum(X_i - \bar{X})^2}{\sum(Y_i - \bar{Y})^2} \quad (1.12)$$

Another important regression parameter to examine is the error term, which is expressed as,

$$e_i = Y_i - \hat{Y}_i \quad (1.13)$$

Errors *(e_i)* are assumed to be independent normal values with a mean of zero and standard deviation of σ which can be computed as:

$$\sigma(Y_i - \hat{Y}_i) = \sqrt{\frac{\sum(Y_i - \hat{Y}_i)^2}{n-2}} \tag{1.14}$$

It is desirable that the value of $\sigma(Y_i - \hat{Y}_i)$ be small since it has a significant effect on the confidence band for prediction. Linear regression analysis can be performed for more than two variables, and in that case, it is known as (multiple linear regression). It is assumed that the dependent variable, *Y* is a linear function of the independent variable, that is,

$$E(Y) = a + b_1X_1 + b_2X_2 + \cdots + b_nX_n \tag{1.15}$$

The estimation of the regression parameters is calculated in way like that for straight-line regression analysis. Nonlinear regression analysis may be necessary when the relationship between Y and X is not linear. An example is the relationship between condition and time as in Fig. 1.8.

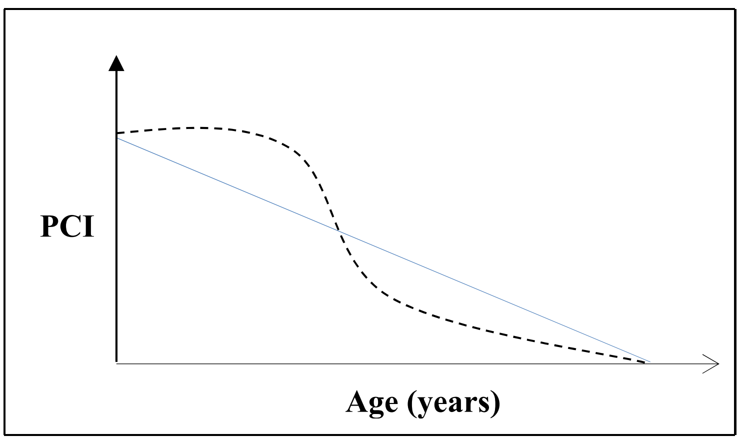

Figure 1.8 Relationship Between Pavement Condition and Age. Taken from (Haas, p. 203)

A linear relationship may be used, but the model will underestimate the condition during the early life of the pavement and will overestimate underestimate the condition during the later part of the pavement life. A nonlinear relationship can be analyzed as a linear model by transforming the *X* variable as in Eq. 1.16.

$$E(Y) = \hat{\alpha} + \hat{\beta}f(x) \tag{1.16}$$

1.3.1.2 Mechanistic-Empirical Model

A pure mechanistic approach to modelling is applicable only to calculating pavement response such as strain, stress, and deflection. This response is normally caused by forces created by traffic, climate, or a combination of the two. Pure mechanistic models for calculating stress and strain cannot be classified as prediction models. However, the calculated stress and strain can be used as input (independent variable) to a regression (empirical) prediction model as presented in the previous section. A prediction model developed using regression technique with pavement response as the dependent variable is called a mechanistic-empirical model. An example of a mechanistic-empirical model is that used for predicting asphalt pavement fatigue life (*N*).

$$N = A * (1/e)^B \tag{1.17}$$

In that prediction model, the strain *e* produced by wheel loading is calculated mechanistically. The coefficients *A* and *B*, however, are determined using regression techniques (Ningyuan Li, 2014).

1.3.1.3 Polynomial Constrained Least Squares

This is one of the most powerful techniques for predicting the change in a variable *Y* (for example roughness) as a function of variable *X* (for example age or traffic). This technique can be applied as following:

First step, given the observations (X_i, Y_i) ; $i = 1,2, \ldots., n$, a polynomial of a degree *n* can be expressed in the following formula.

$$P(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n \tag{1.18}$$

Then, least squares fit is applied on Eq. 1.18, and the desired constraint is met. For example, when fitting PCI vs. age, it is desirable to ensure that the polynomial slope:

$$\acute{P}(x) = a_1 + 2a_2 x + 3a_3 x^2 + \cdots + n a_n x^{n-1} \tag{1.19}$$

which is nonpositive at any age *(x)=0,1,2,…z* when *z* is the highest age (Davison, 2003).

1.3.1.4 S-Shaped Curve

In 1986 Smith used an S-shaped model for relating *PCI* to pavement age. The model had the form:

$$PCI = 100 - \rho/(\ln(\alpha) - \ln(AGE))^{1/\beta} \tag{1.20}$$

where the "Age" is the Remaining Service Life (RSL) of the pavement and α, β and ρ are fixed coefficients that relate to the curve and pavement conditions as shown in Figs 1.9 (a), 1.9 (b) and 1.9 (c). These three constants are determined using regression analysis (Amin, 2015).

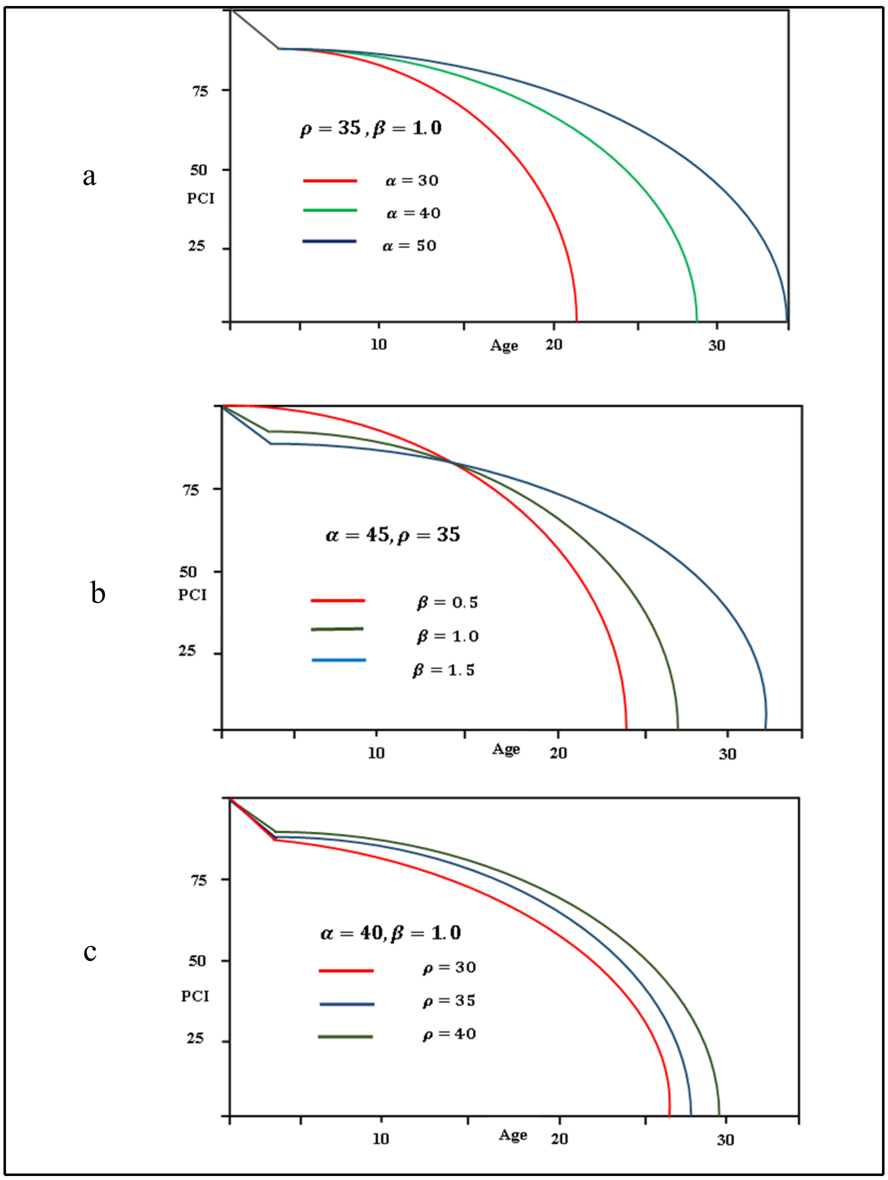

Figure 1.9 S-Sharp Curve Constants Control

1.3.2 Probabilistic Models

Pavement probabilistic models incorporate random variables and probability distributions into the model of pavement performance. While a pavement deterministic model gives a single possible outcome value for a pavement performance, a probabilistic model gives a probability distribution as a solution and then pavement performance can be calculated using the mathematical expectation. For example, according to (Haas, 2005) IRI can be treated as a random variable and therefore it can be described as a probability distribution.

1.3.2.1 Bayes Estimation Concept

For district random variables, let $A_1, \ldots \ldots A_k$ be events that partition a sample space, and let B be an arbitrary event on that space for which $P(B) > 0$. Then Bayes' theorem is

$$P(A_j|B) = \frac{P(B|A_j)P(A_j)}{\sum_{i=1}^{k} P(B|A_i)P(A_i)} \tag{1.21}$$

This reverses the order of conditioning by expressing $P(A_j|B)$ in terms of $P(B|A_j)$ and the marginal probability $P(B)$ in the denominator. For continuous random variables Y and Z,

$$f(z|y) = \frac{f(y|z)f(z)}{\int f(y|z)f(z)dz} \tag{1.22}$$

$f(y) > 0$ is the marginal density, with integration replaced by summation for discrete variables (Davison, 2008).

- **Bayes Inference**

To see how Bayes' theorem is used for inference, suppose that there is a probability model $P(x|\theta)$ for data X. Suppose that we can summarize our beliefs about θ in a prior density, $P(\theta)$, constructed separately from the data X. This implies that we think of the unknown value θ that underlies our data as the outcome of a random variable whose density is $P(\theta)$, just as our probability model is that the data X are the observed value of a random variable X with density P(x |θ). Once the data have been observed, our beliefs about θ are contained in its conditional density given that X = x. Eq. 1.23 is the posterior density for θ given x.

$$P(\theta|x) = \frac{P(x|\theta)P(\theta)}{P(x)} = \frac{P(x|\theta)P(\theta)}{\int P(x|\theta)P(\theta)d\theta} \tag{1.23}$$

Note that $P(x \mid \theta)$ is the likelihood for θ based on y, so that in terms of θ, we have *posterior* ∝ *prior* × *likelihood*. Frequentist inference treats θ as an unknown constant, whereas the Bayesian approach treats it as a random variable (Gatignon, 2010).

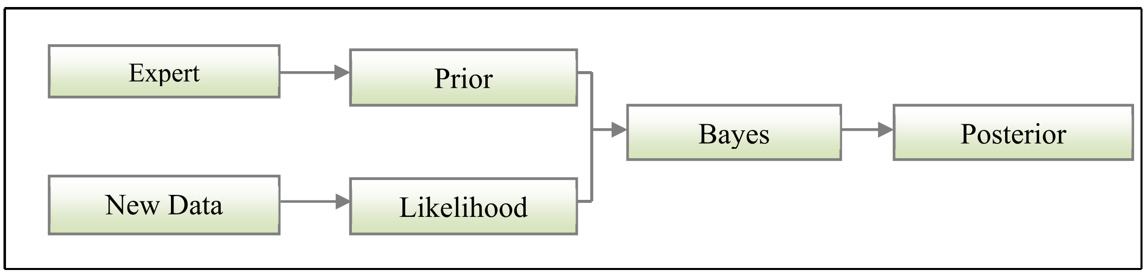

Figure 1.10 Bayes Theorem General Concept. Taken from Gongdon (2003, p. 1-28)

- **Prior Knowledge $P(\boldsymbol{\theta})$**

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

- **Maximum Likelihood Estimation (MLE) $P(\boldsymbol{X}|\boldsymbol{\theta})$**

The maximum likelihood estimation (MLE) approach is one of the most important statistical methodologies for parameter estimation. It is based on the fundamental assumption that the underlying probability distribution of the observations belongs to a family of distributions indexed by unknown parameters. The MLE estimator of the unknown parameters is the maximizer of the likelihood function, corresponding to the probability distribution in the family that gives the observations the highest chance of occurrence. The MLE method starts from the joint probability distribution of the in measured values $x_1, x_2, \dots, x_n$. For independent measurements this is given by the product of the individual densities $p(x|\theta)$ as in Eq.1.24 (N. Y. Li et al., 2011).

$$P(X|\theta) = p(x_1|\theta)p(x_2|\theta) \dots . p(x_n|\theta) = \prod_n^{i=1} p(x_i|\theta) \tag{1.24}$$

- **Posterior Distribution $P(\theta|X)$**

Posterior expresses what is known about a set of parameters based on both the sample data and prior information. Bayes theorem works as a mechanism for generating a posterior of any parameter mixing the prior knowledge with the likelihood. The first iteration production of the prior knowledge and the MLE will then be divided by $P(X)$a normalizing factor, to normalize the distribution. When the posterior distribution $P(\theta|X)$ is in the same family as the prior probability distribution$P(\theta)$, the prior and posterior are then called conjugate distributions. Non-conjugate prior distributions can make interpretations of posterior inferences more difficult.

1.3.2.2 Bayes Linear Regression

Parameter estimation is a very important task for the pavement management systems to be used in prediction models. Nevertheless, historical data to support the developments and updating of these models are in many cases is either not available or not efficient. These databases are often inadequate in sample size, noisy, or incomplete. Conventional statistical modeling tools, such as classical regression analysis, may have limited success in these applications (Kajner, 1995). A promising solution lies in the use of Bayesian regression, which explicitly allows experts to be used to supplement poor quality. Bayesian regression methodology was adopted by the Canadian Strategic Highway Research Program (C-SHRP) for the Canadian Long-Term Pavement Performance (C-LTPP) monitoring program. Nesbit and Sparks (1990) discussed the complete rationale for employing the Bayesian approach for the C-LTPP program in the report "Design of Long-Term Pavement Monitoring System for the Canadian Strategic Highway Research Program."

Bayesian linear regression is an approach to linear regression in which the statistical analysis is undertaken within the context of Bayesian inference. When the regression model has errors that have a normal distribution, and if a particular form of prior distribution is assumed, explicit results are available for the posterior probability distributions of the model's parameters (Gongdon, 2003).

1.3.2.3 Markov Chains Model

Markov chains is a special type discrete-time stochastic process where the state of the system (for example pavement condition) X_{t+1} $at\ time\ t+1$depends on the state of the system X_t at some previous time t but does not depend on how the state of the system X_t wasobtained. In mathematical form this can be expressed as:

$$P(X_{t+1} = j|X_t = i) \tag{1.25}$$

Where P is the probability of the state at time $t + 1$ being j given that the state at time t was i , assuming that the probability is independent of time. This assumption is known as the stationary assumption, and it represents a major limitation for most of the probabilistic models because it implies that the rate of deterioration of pavements is time independent. Few models use so called non-homogenous (time dependent) Markov chains to overcome this limitation (Davison, 2003).

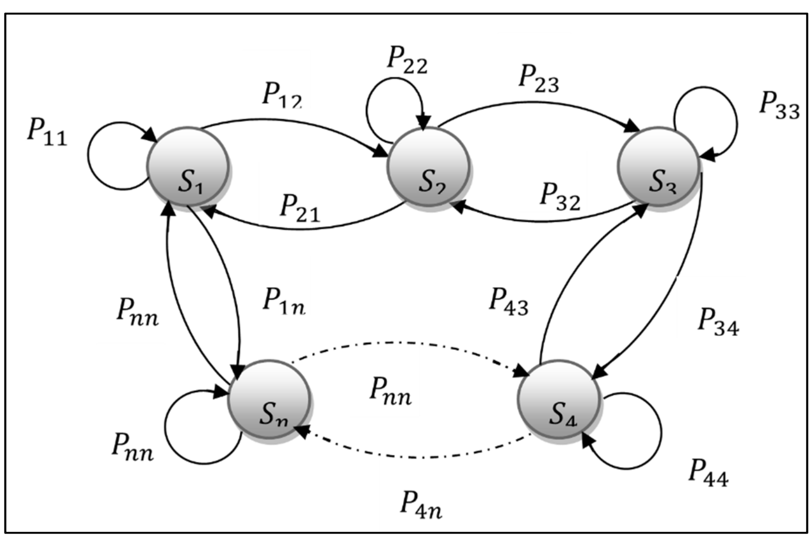

Figure 1.11 Probabilities Transition Between Markov States.
Taken from Wai et al (2006, p. 141-147)

One of the major challenges facing existing probabilistic models is the difficulty in establishing the Transition Probability Matrices (TPMs). A TPM is a square $n \times n$ matrix wheres is the number of possible states in the system. The matrix contains the probabilities of transitioning from state *i* to state *j*. The TPM can be established using historical data or subjective opinions of experienced engineers through individual interviews and questionnaires, which takes considerable time and expenses.

$$TPM = \begin{pmatrix} p_{11} & p_{12} & \cdots\cdots & p_{1n} \\ p_{21} & p_{22} & \cdots\cdots & p_{2n} \\ \vdots & \vdots & \cdots\cdots & \vdots \\ p_{n1} & p_{n2} & \cdots\cdots & p_{nn} \end{pmatrix}_{n\times n} \tag{1.26}$$

The deterioration of pavements is affected by several factors some of which are difficult to observe. Traffic load and environmental conditions change over time and are difficult to predict. This makes the performance or deterioration of pavements to vary greatly showing uncertain or random characteristics. Furthermore, uncertainty can arise from the inspection or measurement process and from inability to quantify the factors that affect the deterioration process, and to model the true deterioration process of the materials. Thus, pavement deterioration process shows stochastic characteristics. Probabilistic models attempt to tackle the stochastic characteristics of the pavement deterioration process. Most of the proposed probabilistic models are based on Markov process modelling. The pavement performance deterioration versus age is modeled as a time-independent Markov process as Eq. 1.27 (Kobayashi et al., 2012).

$$V(t)_{n\times 1} = V(t-1)_{n\times 1} TPM_{n\times n} = V(0)_{n\times 1} TPM^{t}_{n\times n} \tag{1.27}$$

Where :

$V(t)_{n\times 1}$ The predicted condition state matrix at year *t*,

$V(0)_{n\times 1}$ is the initial condition state matrix at year 0,

and *TPM* is the one-step transition probability matrix $TPM_{n\times n}$

1.3.2.4 Time Series Models

A time series is a set of observations y_t, each one being recorded at a specific time t. While time series analysis is a term that includes methods of analyzing time series data to extract meaningful statistics and features of the data. On the other hand, time series forecasting is the use of a model to predict future values based on previously observed values. There are many examples of which are time series and some from them are a population of a country measured at ten-year intervals, stock prices, Sales demand, daily temperatures, and quarterly sales (Peter J. Brockwell, 2002). A time series in general is supposed to be affected by four main components, which can be separated from the observed data. These components are trend, cyclical, seasonal and irregular components (Ratnadip & Agrawal;, 2007).

- **Nonstationary and Stationary Data**

A stationary time series is one whose statistical properties such as mean, variance, autocorrelation, etc. are all constant over time. However, in many studies, time series data is non-stationary data, unpredictable and cannot be modeled or forecasted. The results obtained by using non-stationary time series may be spurious in that they may indicate a relationship between two variables where one does not exist. In order to receive consistent, reliable results, the nonstationary data needs to be transformed into stationary data (Davison, 2003).
A time series is a set of observations y_t, each one being recorded at a specific time t. While time series analysis is a term that includes methods of analyzing time series data to extract meaningful statistics and features of the data. On the other hand, time series forecasting is the use of a model to predict future values based on previously observed values. There are many examples of which are time series and some from them are a population of a country measured at ten-year intervals, stock prices, Sales demand, daily temperatures, and quarterly sales (Peter J. Brockwell, 2002). A time series in general is supposed to be affected by four main components, which can be separated from the observed data. These components are trend, cyclical, seasonal and irregular components (Ratnadip & Agrawal;, 2007).

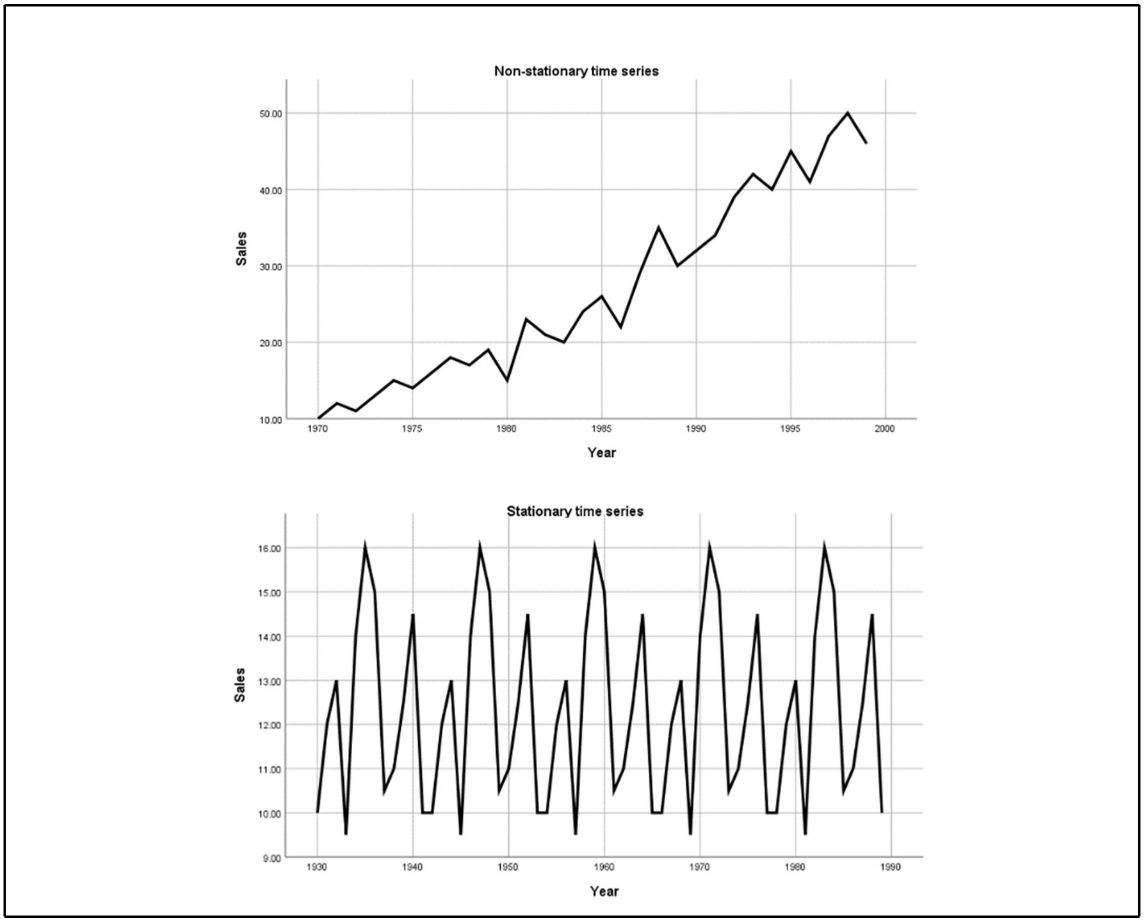

Figure 1.12 Non-stationary and stationary time series data. Taken from Durango (2007, p. 495)

- **Time Series Forecasting**

In general, models for time series data can have many forms and represent different stochastic processes. There are two widely used linear time series models Autoregressive (AR) and Moving Average (MA) models. The combination of these two models together produces the Autoregressive Moving Average (ARMA) and Autoregressive Moving Average (ARIMA). ARMA model is used if the time series data is stationary and If the time series data is non-stationary, then ARIMA is used (Ratnadip & Agrawal;, 2007).

- **Autoregressive Integrated Moving Average (ARIMA)**

ARIMA model is a type of models that expresses a given time series based on its own past values, that is, its own lags and the lagged forecast errors, so that equation can be used to forecast future values. In ARIMA models a non-stationary time series is made stationary by applying finite differencing of the data points (Hunt & Bunker, 2003). An ARIMA model is characterized by three terms: p, d, q and expressed as ARIMA(p,d,q) where, p is the order of the AR term q is the order of the MA term d is the number of differencing required to make the time series stationary. AR model with p lags is expressed as in equation 1.28 where $(\varphi_0, \dots, \varphi_{t-p})$ are the model constants and y_p is the coefficient for the lagged variable in time $t-p$. The moving average models (MA) which represent the possible relationship between the variable and the residuals from past periods. MA with q lags is expressed as in equation 1.29 where θ_q is the coefficient for the lagged error term in time $t-p$. By combining AR part which is shown in equation 1 with the MA part in equation 2, ARIMA model is resulted in equation 1.30.

$$y_t = \varphi_0 + \sum_{i=1}^{p} \varphi_i \, y_{t-i} + \epsilon_i \tag{1.28}$$

$$y_t = \mu + \epsilon_t - \sum_{i=1}^{q} \theta_i \, \epsilon_{t-i} \tag{1.29}$$

$$\hat{y}_t = \mu + \sum_{i=1}^{p} \varphi_i \, y_{t-i} + \epsilon_t - \sum_{j=1}^{q} \theta_j \, \epsilon_{t-j} \tag{1.30}$$

Where: μ is a constant, $\sum_{\mathrm{i=1}}^{\mathrm{p}} \varphi_{\mathrm{i}} \, y_{t-i}$ is the autoregressive terms, and $\sum_{j=1}^{q} \theta_j \, \epsilon_{t-j}$ is the moving average terms (Wei, Ju, Liyang, Susan, & Lili, 2019).

1.4 Previous Work on Probabilistic Modeling of Pavement Performance

There are many studies that have used probabilistic models to model pavement performance and where different probabilistic methods have been used.

1.4.1 Bayesian Models

Bayes linear regression is a powerful tool that combines the classical linear regression with the Bayes concept to estimate model parameters. It investigates probabilistically the value of each parameter inference which then produces an accurate estimation. The basic principle of

Bayesian statistics lies in combining prior probabilities with experimental outcomes to determine a post-experimental or posterior probability. The posterior distribution expresses what is known about a set of parameters based on both the sample data and prior knowledge (Han, Kaito, & Kobayashi, 2014). In pavement management systems, many studies have used the Bayes linear regression in order to estimate the pavement performance. All these studies have generally conducted the same steps to develop the estimation models. However, they were differing in some tiny details in accordance with the road network specifications and data availability.

In 1995, Canadian Strategic Highway Research Program (CSHRP) has developed a Bayesian linear regression model to predict the roughness progression of asphalt overlay which was applicable in central Alberta. Prior knowledge was provided by five pavement engineers from Alberta Road Agency. Riding Comfort Index (RCI) was selected as an independent variable. Five dependent variables were selected to be a function of the RCI. These variables are overlay thickness, age, initial RCI, and cumulative traffic loading (Kajner, 1995).

In a research paper presented by(Feng Hong & Prozzi, 2005), an existing incremental pavement deterioration model which was fundamentally developed based on data from the AASHO Road Test, was analyzed and updated. For the prior knowledge, they have suggested previous research results (Prozzi & Madanat, 2003) to be used as the learned knowledge on the issue of interest. However, researcher judgement and experience could be used when reliable data is not available. The dependent variable was the Present Serviceability Index (PSI) and the independent variables are layer thicknesses, traffic increment, and frost gradient.

(Amador-Jimenez & Mrawira, 2011a, 2011b, 2012) have produced three research papers; the first paper has introduced a rut depth Bayes regression model which could employ expert criteria combined with historical knowledge and current observations. Pavement network was categorized based on road strength under the effect of traffic load accumulations. The model estimates posterior probabilistic distributions for the regression coefficients of the mechanistic equation. In their second paper, an International Roughness Index (IRI) deterministic performance model was initially developed for the Costa Rica Road network. Then, a multilevel Bayesian regression model was obtained to anticipate the pavement performance and calculate the confidence intervals for the model parameters. Eventually, using the Bayesian model, the pavement strength coefficients were calibrated to local observations. In the last paper, authors have developed a pavement performance models which were capable of capturing variability based on adequate time series predictors. The model was designed to estimate the pavement condition even with missing data with an ability to generate many deterioration models with no need of classifying the road network to families. IRI was the independent variable while the pavement age and traffic accumulation load in ESAL are the dependent variables.

1.4.2 Markovian Models

Markov model is one of the most important probability prediction tools. It has been used in many research disciplines such as bridge deterioration, speech recognition, bioinformatics, and asset management.

In pavement management systems, many studies have used Markov model to predict the future pavement condition. In a study conducted by (N. Li et al., 1997), a Markovian model was developed to predict asphalt pavement performance. The study aims to overcome the limitations of the Ontario Pavement Analysis of Costs (OPAC) system which is a deterministic model. The model main limitation is the existence of uncertainties and variations in pavement design variables and parameters. Li et al have noticed that it is not accurate to apply deterministic models to all situations of pavement management. Authors have come out with a new concept that converts the existed deterministic model to a probabilistic model. The main idea is to generate a set of time-related non-homogeneous Markovian transition probability matrices, which is determined by a Monte Carlo simulation. For more accurate results, a Bayesian technique was used to update the Markovian model whenever new pavement condition data is captured (N. Li et al., 1997).

Markov model is recommended when the system states are observable and when it is difficult to observe all system states, Hidden Markov Model is recommended to predict the future condition. (Nagaraja, 2006) have introduced a probabilistic methodology to predict bridge deterioration processes based on Hidden Markov Model (HMM). The transition probabilities to forecast the deterioration of every bridge component was proposed. The transition progress between a set of condition states representing the conditions of each bridge component were defined by using exponential hazard models. Furthermore, based on periodical inspection data, the maximum likelihood method was proposed to determine the parameters of the exponential hazard model. Heterogeneous inspection data related to specific structural characteristics and usage conditions were employed to estimate the deterioration of bridge components in a disaggregate way. The exponential hazard model proposed by this research permits estimating Markov transition probabilities for arbitrary time intervals. In addition, the usefulness of the Markov transition probabilities estimation method proposed by this research was positively verified through an empirical example applied to steel bridges. The methodology proposed by this research can be applied to forecast the deterioration not only of steel bridges but also to other infrastructures if the respective empirical research is accumulated.

(Kobayashi et al., 2012) have adopted Tsuda’s technique to be used in pavement to predict future transition probabilities based on pavement condition data sets. The authors have proposed an innovative analytical methodology to forecast the deterioration process of infrastructure through a hidden Markov model. In the model, selection biases are considered as random variables. Selection biases are eliminated through the assumption of prior and posterior distribution in Bayesian estimation. Furthermore, Markov Chain Monte Carlo simulation is introduced to generate a random sampling population in Bayesian estimation algorithm. Authors have presented an empirical study on the Japanese national road system. Authors have noticed that selection biases were existing in the monitoring data especially in states 3 and 4. As a result, theses selection biases have affected the deterioration prediction model comparing the estimation results of multi-stage exponential Markov model of the proposed hidden Markov model.

To come over the limitation of (Kobayashi et al., 2012) method to predict pavement future condition, (Lethanh & Adey, 2013) have introduced a methodology when data quality and quantity are not enough to validate prediction models. They proposed an Exponential Hidden Markov Model (EHMM) to predict pavement future condition using the pavement condition transition probabilities. To be able to apply this methodology, Authors have pointed to an important consideration which is that the physical deterioration of the road is something that cannot be observed directly, but that can be derived through the evolution over time of one aspect related to the physical deterioration of the road such as roughness. As a result, the physical deterioration of the road is a hidden process, and the aspect related to the physical deterioration of the road is an observable process. To model the deterioration process using a Markov model, authors have estimated the probability of the process passing between condition states in a period using the exponential function as a probability distribution. The model was tested using the data collected on 10237 road sections in the Vietnamese national road network. It was found that the more the texture depth data available, the closer the deterioration prediction between the multi-stage exponential Markov model and the proposed hidden Markov model.

1.4.3 Time Series

Time series methodology has not been efficiently used to predict future pavement performance few studies have been conducted in this context. There are few studies that have applied time series technique in PMS to analyze and predict future pavement condition. Previous studies in pavement deterioration prediction models have mainly focused on analyzing time series data to investigate the pavement condition trend and extract some features from the data. Few studies have addressed the issue of non-stationary time series data, when forecasting pavement future condition using time series, it is important to remove non-stationarity from time series data to get reliable results.

In 2003, Hunt PD and Bunker JM, have introduced a methodology to better understanding of unbound granular pavement performance by checking out roughness progression, and for the purpose of modeling the roughness progression to improve roughness prediction approaches using time series evaluation. The project partner road agency`s PMS and the available data for this study gave way to a methodology allowing roughness progression to be investigated using fund priorities techniques for each 1 km pavement section in a 16,000km road network. The researchers have combined family group data fitting and site-specific data modelling approaches to investigate the roughness progression for all targeted road sections. The study was conducted throughout six steps and started with the data collection. The second step was to establish a roughness progression trend with time for each road section. Third step is to determine the masking effect that could be caused by pavement maintenance and affect the roughness progression. The following step was to determine the best fit over filtered roughness progression data and establish a Linear Roughness Progression Rate (LRPR) for each road section. Then, pavement performance scale was defined and was followed by the investigation of the effect of pavement variables on LRPR. The research results indicated that the prediction of future roughness would be more accurate when more historical data is available. For road sections, use of a LRPR estimated based on Annual Average Daily Traffic AADT categories, would appear to be reliable.

Durango-Cohen has presented in 2007 an integrated framework to address performance prediction and maintenance optimization for transportation infrastructure facilities. Facility deterioration was represented as a time series which provides an approach to determine and estimate performance models. The study illustrated how to use the framework to determine the effect of inspection technologies on facility life-cycle cost. thus, to help decision makers when selecting optimal intervention strategy. Based on numerical showcase examples, the author has explained that merging imprecise technologies can enhance the capabilities of the prediction process and can result in substantial operational cost savings. This is particularly essential because highly precise systems tend to be substantially more expensive to look at, especially when they are very first introduced.

(Amador-Jiménez & Mrawira, 2011) have presented a multi-level Bayesian model to check mechanistic model parameters using archived data to ensure reliability when estimating predictions confidence interval. Authors have faced a missing data issue which can affect the results of the prediction models. Therefore, to overcome this problem, many methods were suggested one of them is a time series methodology which was used to establish general trends and replace them to fill in missing data.

(Wei et al., 2019) have presented an innovative IRI prediction model based on fuzzy-trend time-series forecasting and particle swarm optimization (PSO) techniques. Pavement performance database were used for model performance assessment. IRI data was divided into categories based on granular spaces. additionally, the multifactor interval division method was suggested. To predict the fuzzy trend of each factor, a second-order fuzzy-trend model and fuzzy-trend relationship classification method were proposed. The next step was to generate the fuzzy-trend states for multiple granular spaces depending on various uncertainties. at last, performance model was optimized using the PSO technique simultaneously with the implementation of IRI forecasting. The research results indicated that the IRI prediction model has achieved high accuracy as well as provided detailed explanation of the IRI time series trends.

CHAPTER 2

BAYES LINEAR REGRESSION PERFORMANCE MODEL DEPENDING ON EXPERTS` KNOWLEDGE AND CURRENT ROAD CONDITION

Abdussalam Heba [a] and Gabriel J. Assaf [b]

[a, b] Department of Construction Engineering, École de Technologie Supérieure, 1100 Notre-Dame West, Montreal, Quebec, Canada H3C 1K3

Paper published in *Civil Engineering Materials (ASTM) International journal*, August 2018

2.1 Introduction

This chapter establishes a Bayesian regression method to develop a performance model for cases when archived pavement data are not available, based on expert knowledge as a prior distribution. As such, experts who have worked for a long time with the road and transportation agencies have been interviewed to develop a portion of the input data. Posterior distribution was calculated using the likelihood estimation function based on road condition inspections according to the predefined protocol. The results were prediction models of pavement deterioration based on a mixture of a few onsite inspections interacting with expert knowledge. This chapter was published in Advances in Civil Engineering Materials (ASTM) International journal and then has been rewritten to accommodate the continuity context of the thesis.

2.2 Problem Definition

Libya has paid great attention to road projects and has established a public network with a length of 34,000 kilometers according to the latest engineering standards and specifications, including 15,000 kilometers of main roads and 18,000 km of secondary and agricultural roads. Unfortunately, because regular maintenance is not conducted in the absence of any PMS, the road network is continuously deteriorating, causing direct losses in the road network asset value. It is obvious that the road network needs maintenance intervention, which is expected to take a long time to be completed. To conduct any M&R activities, the road agency needs to develop prediction models to be used in conjunction with the PMS for implementing cost-effective pavement management solutions. Because of the lack of archived data and the absence of any type of pavement performance measurements, a Bayesian probability approach was used. To illustrate the effectiveness of this method, the Libyan road network was chosen as a case study.

2.3 Methodology

Roads deteriorate and their IRI drop gradually over time. This relationship can be represented using linear regression but, practically, road sections having the same zone, age, load, and soil strength conditions could still have a different rate of deterioration. Therefore, Bayesian linear regression is the appropriate technique wherein basic Bayesian philosophy is applied. This is because Bayesian regression is a probabilistic approach that accounts for variability (refer to FIG. 2.1). As such, in Bayesian inference, MLE is considered to be point estimation. However, in Bayesian linear regression, productive probability around each inference of the IRI is probabilistically investigated.

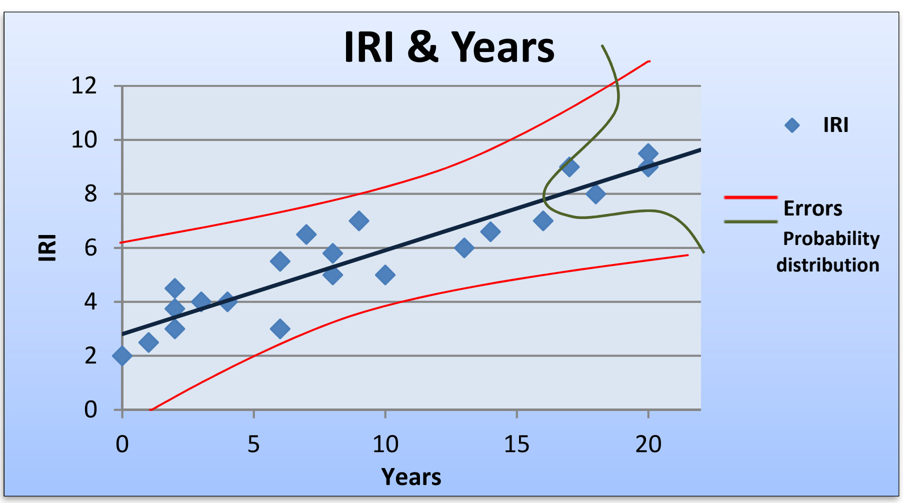

Figure 2.1 The Differences Between Linear Regression and Bayesian Linear Regression Taken from Hong (2006, p. 7)

Bayes linear regression is the technique that used in the research to anticipate the performance of a specific road section. It is a way of integrating engineers' experience with the current available information on road performance. This is done by considering the engineers' experience as the prior probability distribution and the current pavement condition is the MLE function. By Bayesian processing, the whole range of the IRI inferential solutions is recovered, rather than a point estimate and a confidence interval as in classical regression. Moreover, MLE of the parameters of a Non-Bayesian Regression model or simply a linear regression model overfits the data, meaning the unknown value for a certain value of independent variable becomes too precise when calculated. Bayesian Linear Regression overcomes this fact, saying that there is uncertainty involved by incorporating predictive distribution. The research methodology has four main stages (see Fig. 2.2), which are as follows: firstly, the experts` interview, and based on the results of these interviews, the prior probability distribution is configured. Then, to compute the MLE function, IRI is measured for selected sections of the road network. Then, the posterior distribution is defined. Finally, compute the predictive distribution to estimate the IRI.

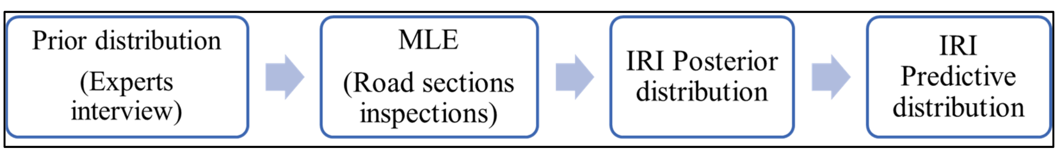

Figure 2.2 Methodology Steps

2.3.1 Pavement Zones and Model Variables

In this study, Libya as a developing country was selected to be a case study, so far Libyan Road and Transportation Agency (LRTA) does not apply any pavement management system. As a result, different types of archived data such as pavement performance data are not available. Libya is in North Africa, so both the Mediterranean Sea and the desert affect its climate. In the winter, the weather is cool with some rain on the coast and in the drier the desert.

In the summer, the weather becomes hot and dry, especially in desert. It is therefore appropriate to divide the pavement areas into two northern and southern zones.

When developing equations that predict pavement conditions, soil and traffic load are often considered as independent variables. In many cases, they are combined with age as an independent variable. In this study, 3 loading levels, 3 soil conditions and 2 climate zones interact with each other and result in 18 pavement families. Each pavement family's information is then used to develop Bayesian linear regression prediction, the classification is as shown in Table 2.1 which provides a summary of the characteristic of 18 families and codification of 18 associated databases.

Table 2.1 Road Network Zones Interacting with Traffic Loads and Soil Strength

Soil Strength	**North Zone**			**South Zone**		
	Load Level			**Load Level**		
	L	**M**	**H**	**L**	**M**	**H**
L	DS1	DS2	DS3	DS4	DS5	DS6
M	DS7	DS8	DS9	DS10	DS11	DS12
H	DS13	DS14	DS15	DS16	DS17	DS18

The soil conditions were divided into three categories (low, medium, high). Soil strength is measured by a penetration test in accordance with the California Bearing Ratio test (CBR) which evaluates the subgrade strength of roads. Traffic loads are used as a factor that affects the rate of deterioration of a road surface. Traffic loads are categorized as follows:

- Low: < 50 vehicles/day
- Medium: 50-500 vehicles/day
- High: 500-2000 vehicles/day

In general, these models are network-level deterioration models and not project-level deterioration models because the characteristics and the properties of the materials are not presently available in Libya. As a result, IRI relating to time in years is used this study case to estimate the pavement performance.

The research data required for IRI estimations has been divided into two main categories. The first category was extracted from the interviews with the experts. The second category is the MLE data; this data has been collected using the IRI as part of the road section inspections process and has been done to measure the road deterioration. The MLE data is extracted and summarized as pairs of (t_i, IRI_j) where IRI represents the road section condition and *t* indicates the time (see Eq. 2.1).

$$D = \left((t_1, IRI_1), \dots\dots, (t_n, IRI_n)\right), 0 \leq t_i \leq 20, 0 \leq IRI_j \leq 16 \tag{2.1}$$

Therefore, the IRI_j is a model to be conditionally independent given the *w* vector which will be the prior distribution (refer to Eqs. 2.2, 2.3).

$$IRI_j \sim N(w^T t_i\ ,\ a^{-1}), a > 0 \tag{2.2}$$

$$w \sim N(0, b^{-1}I), b > 0, w = (w_1, \dots\dots, w_d) \tag{2.3}$$

Where $a = \frac{1}{\sigma^2}$ is the precision factor, *b is* the covariance matrix; *a* and *b* are known, and *w* is a parameter vector with a Gaussian multivariate density.

2.3.2 Experts` Knowledge and Prior Probability Distribution

The Bayesian statistical approach combines prior knowledge (experience) with field data. In highway engineering, new models are continually needed to better predict pavement performance or to run various PMS; however, it takes much time and expense to gather data about pavement performance (N. Y. Li et al., 2011). In such situations, the Bayesian approach is useful in short circuiting the data collection cycle. After gathering some data, which may not be sufficient to support meaningful classical regression, one can collect some expert judgment and combine the two sources of information into a relatively robust regression model. The expert judgment serves to bridge the gaps in field data (Z. Li, 2005)

It is obvious that a lot of valuable information can be obtained from people who have observed pavement performance throughout their careers. This professional and field staffs know what variables are contributing to pavement performance. They understand the functional relations of the variables. Their impressions on these relationships can be encoded and when combined with field data, these impressions can have profound impacts on the resulting posterior models. That is why; initial data has been collected by interviewing Libyan experts who have worked for many years on the development of the Libyan road network. Six engineers were interviewed using a standardized, open-ended interview technique (see APPENDIX I). The experts' interview data was combined by taking the average of all experts' opinions for each pavement family.

An analysis of variance was done before combining the experts' knowledge; this ensures that all experts' priors are consistent. From Fig. 2.3 and Table 2.2, there is no significant evidence to show that there is a difference in group means. As a result, the experts' opinions were considerably compatible; this means that all of the experts' knowledge about the roughness progression was close to each other.

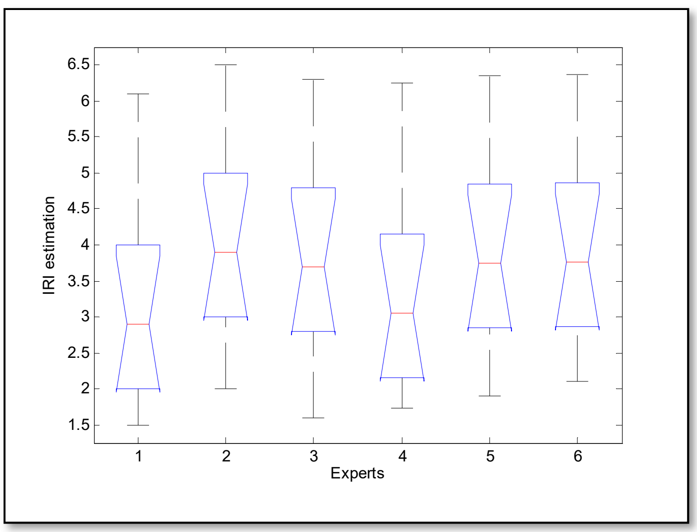

Figure 2.3 ANOVA Experts' Opinions Comparison

Table 2.2 ANOVA Differences Between the Means of Experts' Opinions

Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	42.79	5	8.56	186.57	1.17E-49	2.30
Within Groups	4.68	102	0.05			
Total	47.47	107				

After that, an inspection of representative road sections from each the 18 families were conducted in Libya. The road deterioration was measured by the IRI on a subjective basis.

2.3.3 IRI Maximum Likelihood Estimation (MLE)

Given data $D = (IRI_1, \ldots., IRI_n)$ where D represents a sample from the IRI statistical population that has been collected from road section inspections. Then, the MLE is computed using Eq. 2.4:

$$P(D|w) \propto \exp(-\frac{a}{2}(IRI - Aw)^T(IRI - Aw)) \quad (2.4)$$

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

$$A = \begin{pmatrix} - & t_1^T & - \\ \vdots & \vdots & \vdots \\ - & t_n^T & - \end{pmatrix}, \; IRI = (IRI_1, \ldots\ldots, IRI_n)^T \quad (2.5)$$

2.3.4 IRI Posterior Distribution

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

$$P(w|D) \propto P(D|w)P(w) \quad (2.6)$$

After that, we replace the MLE expression in the posterior; this is shown in equation 2.7.

$$P(w|D) \propto \exp(-\frac{a}{2}(IRI - Aw)^T(IRI - Aw) - \frac{b}{2}w^T w \quad (2.7)$$

With a little calculus we can express *w* in the form of a Gaussian distribution and call it a precision matrix:

$$P(w|D) = N(w|\mu, A^{-1})\; Where\; \mu = a\Lambda^{-1}A^T iri\; ;\;\; \Lambda = aA^T A + bI \quad (2.8)$$

That shows us the MLE estimations of *w,* which are:

$$w_{MLE} = (A^T A)^{-1}A^T iri \quad (2.9)$$

2.3.5 Predictive Distribution

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

$$P(iri|t, D) = \int P(iri|t, w)(w|t, D)dw \quad (2.10)$$

$$= \int N(iri|w^T t, a^{-1})N(w|\mu, A^{-1})dw \quad (2.11)$$

$$\propto \int exp(-\frac{a}{2}(iri - w^T t)^2 exp(-\frac{1}{2}(w-\mu)^T \Lambda(w-\mu)dw \quad (2.12)$$

This formula is then factored and put in a quadratic form as a function of *w* in a formula similar to the following general expression: $\int N(w| \ldots \ldots)g(iri)dw$ and then, since $g(iri)$ is not dependent on *w,* it comes out of the integral and $\int N(w| \ldots \ldots)dw$ integrates to 1. After several algebraic steps, finalization of the predictive distribution is as in Eq. 2.13:

$$P(iri|t, D) = N\left(iri\middle|u, \frac{1}{\lambda}\right) where\ u = \mu^T t \quad and \quad \frac{1}{\lambda} = \frac{1}{a} + t^T \Lambda^{-1} t \quad (2.13)$$

Finally, using mathematical expectation and equation 2.13 in all road section families, IRI will be estimated depending on *iri* which is the expert interview data, *t* which is the time corresponding with road conditions, and *D* which is the data collected from road inspections.

2.3.6 Model Implementation

This section consists of all required steps to apply the Bayes regression on the collected data. The model has 1000 iterations using a combined prior to present all expert knowledge encoded in one model for each pavement family. A combined prior was selected to develop a single model for each pavement family.

WinBUGS was chosen as a programming platform; this is free software available from the Biostatistics Unit of the Medical Research Council in the UK (Jongsawat & Premchaiswadi, 2010). The WinBUGS program consists of three parts, all of which can be placed into a single file or as three separate files. The first part is the main program that is a string of computer code that lets WinBUGS know what the prior and likelihood of the model is. The second part is the data set that can be entered using matrixes in the same program or can be called from a file. The last part is the initial values that are used to start the algorithm. To estimate the parameters in Bayesian analysis, the prior distribution is multiplied by the likelihood; samples are then taken from the posterior distributions via an iterative Markov Chain Monte Carlo (MCMC) algorithm (Davison, 2008).

2.4 Model Results

The model combines data taken from the road condition inspections in accordance with a pre-established protocol and prior knowledge of the six experts who participated in the interviews. Once the model, the data, and the initial values have been specified, the program will be ready to be compiled and to run the MCMC algorithm. WinBUGS offers a Sample Monitor Tool panel which consists of several task icons as shown in Fig 2.4.

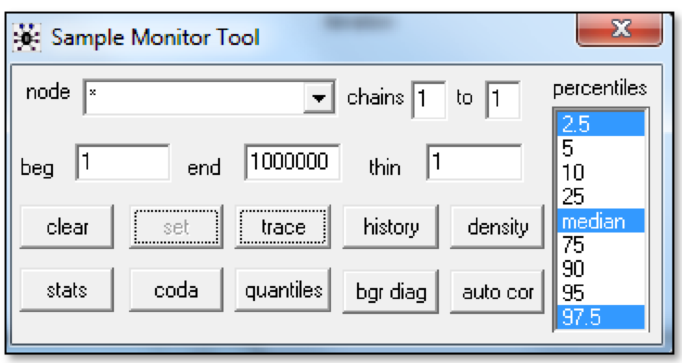

Figure 2.4 Sample Monitor Tool

One of these tools is the stats tool; this gives a zoomed-out view of the entire posterior summary for the Bayes linear regression parameters, as illustrates in Table 2.3 which represents the first pavement family. Complete parameters estimation results are summarized and shown in the Tables (2.4, 2.5).

Table 2.3 1st Pavement Family Posterior Summary

Node	μ_1	sd_1	MC error	2.5%	Median	97.5%	Start	Sample
α_1	0.7039	0.1896	0.01671	0.4934	0.6965	0.885	1	1000
β_1	0.2850	0.01372	0.001211	0.2720	0.2857	0.3014	1	1000
τ_1	29.71	9.722	0.3757	12.81	28.91	51.02	1	1000

Table 2.4 North Zone Parameters Estimations

North Zone Pavement families									
Parameters	1	2	3	4	5	6	7	8	9
α_i	0.7039	0.7505	0.7505	0.5845	0.914	0.7714	0.9021	1.132	0.7431
β_i	0.2850	0.3665	0.3665	0.3719	0.3154	0.3415	0.3167	0.3086	0.3627
τ_i	29.71	9.243	9.243	14.09	16.46	10.29	17.1	16.09	9.543

Table 2.5 South Zone Parameters Estimations

South Zone Pavement families									
Parameters	10	11	12	13	14	15	16	17	18
α_i	0.7489	0.7505	0.7622	0.7658	0.7741	0.7806	0.7851	0.7868	0.7945
β_i	0.3644	0.3665	0.3719	0.3751	0.3783	0.3808	0.3822	0.3853	0.3882
τ_i	9.306	9.243	9.197	8.846	8.653	8.672	8.448	8.568	8.383

As a result of the Bayesian analysis, IRI predictive posterior models for the first pavement family were developed. The models have one independent variable, and the predictive posterior equation is, as shown in equation 2.14 where α, β are the estimated parameters and t represents time in years.

$$IRI = \alpha + \beta t_i \tag{2.14}$$

Fig. 2.5 shows the MCMC behavior, where the chain appears to be moving around readily. This behavior is called the dynamic trace because it will be continuously refreshed in real time if the model is updated.

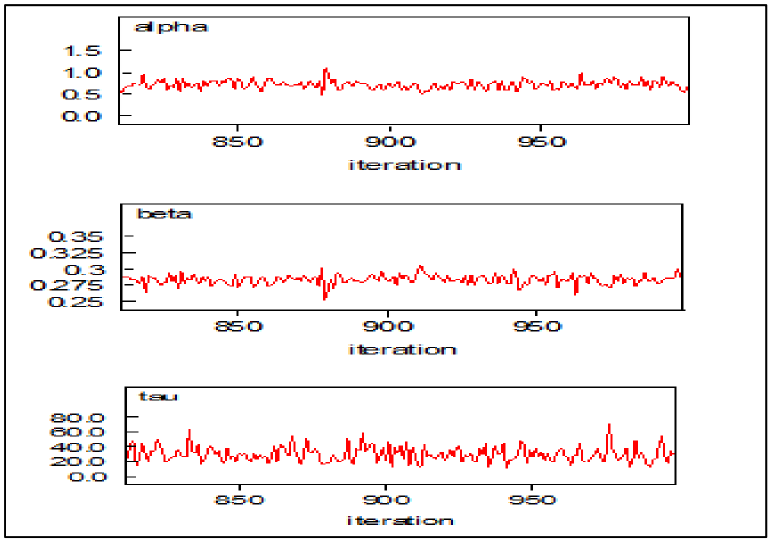

Figure 2.5 Parameter Outputs Dynamic Trace

The probability distribution densities of the parameters are displayed and summarized, as shown in Fig. 2.6.

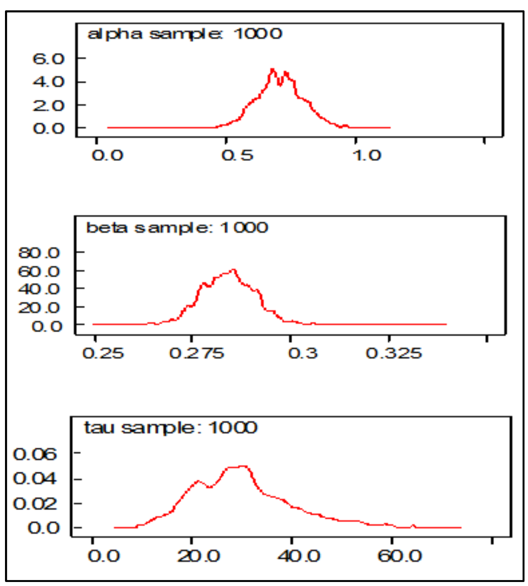

Figure 2.6 Model Parameters Posterior Densities

The basic analysis of the MCMC output is obtained by checking the convergence of the chain and the autocorrelation as respectively show in Fig. 2.7 and 2.8.

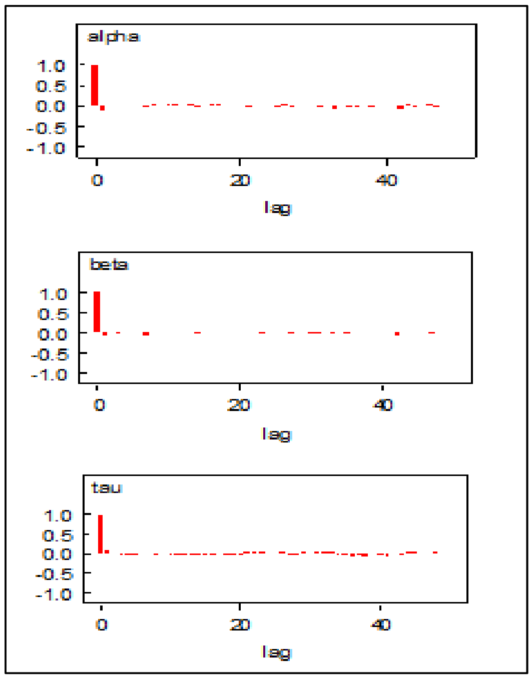

Figure 2.7 Parameters Autocorrelation Functions

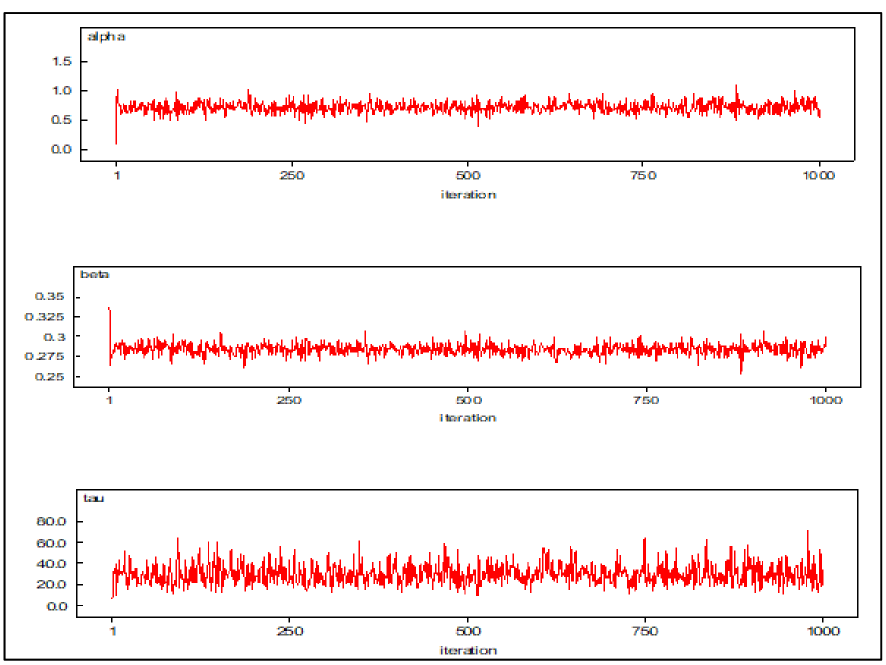

Figure 2.8 Parameters Trace History Output

Fig. 2.9 illustrates the moving averages of the mean and the 95% credibility interval; all parameters appear stable over the course of the run.

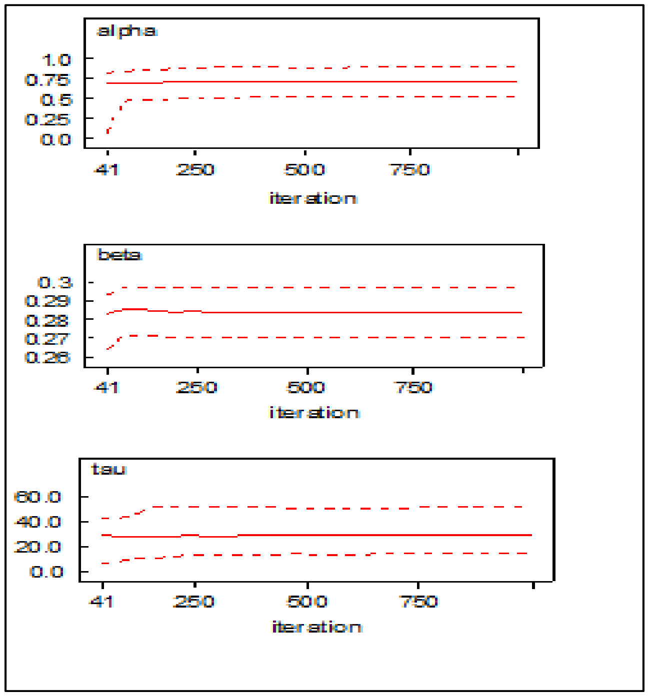

Figure 2.9 Parameters Running Quantiles

The parameters posterior estimation and the 95% parameters credibility intervals are summarized in Tables 2.6, 2.7 and Fig 2.10 illustrate a simple comparison between the estimated parameters for all 18 pave families.

Table 2.6 α Parameters Model Result Summary

α_i	95% Credibility Interval	Posterior Mean	MLE
1	(0.42, 1.10)	0.7505	0.7387
2	(0.32,0.88)	0.6068	0.5937
3	(0.31,0.83)	0.5767	0.5683
4	(0.31,0.83)	0.5845	0.5714
5	(0.36,0.40)	0.9140	0.9043
6	(0.46,1.10)	0.7609	0.7609
7	(0.66,1.10)	0.9021	0.8922
8	(0.88,1.40)	1.1300	1.1200
9	(0.40,1.00)	0.7431	0.7328
10	(0.42,1.10)	0.7489	0.7349
11	(0.42,1.10)	0.7505	0.7387
12	(0.43,1.07)	0.7622	0.7504
13	(0.43,1.10)	0.7658	0.7534
14	(0.44,1.10)	0.7741	0.7638
15	(0.42,1.10)	0.7809	0.7690
16	(0.42,1.10)	0.7851	0.7715
17	(0.45,1.10)	0.7868	0.7728
18	(0.46,1.12)	0.7945	0.7803

Table 2.7 $\boldsymbol{\beta}$ Parameters Model Result Summary

β_i	95% Credibility Interval	Posterior Mean	MLE
1	(0.34,0.30)	0. 3665	0.3673
2	(0.35,0.39)	0. 3676	0.3685
3	(0.35,0.39)	0.3719	0.3728
4	(0.35,0.39)	0.3719	0.3728
5	(0.30,0.34)	0.3154	0.3161
6	(0.32,0.37)	0.3415	0.3423
7	(0.30,0.34)	0.3176	0.3176
8	(0.29,0.33)	0.3086	0.3285
9	(0.31,0.37)	0.3627	0.3635
10	(0.34,0.40)	0.3644	0.3654
11	(0.34,0.39)	0.3665	0.3673
12	(0.35,0.40)	0.3719	0.3983
13	(0.35,0.40)	0.3751	0.376
14	(0.35,0.41)	0.3783	0.3792
15	(0.31,0.41)	0.3808	0.3817
16	(0.36,0.41)	0.3822	0.3834
17	(0.36,0.41)	0.3853	0.3863
18	(0.37,0.41)	0.3882	0.3892

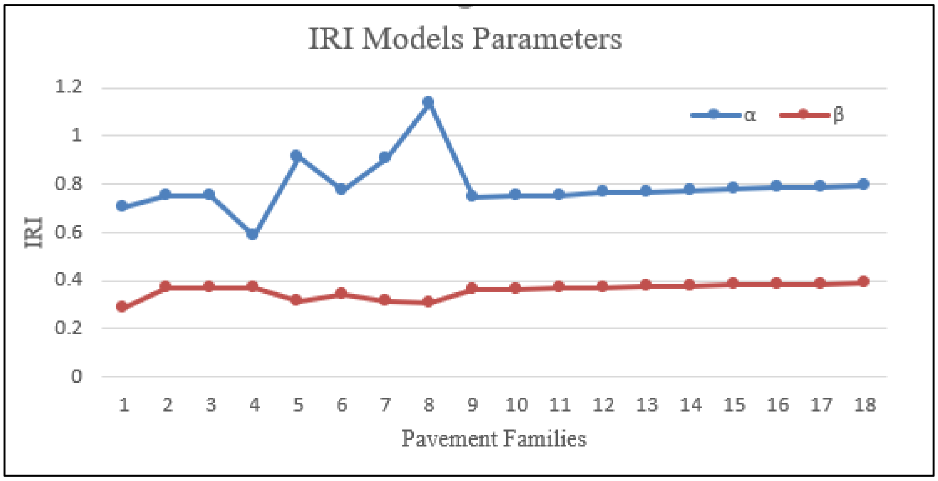

Figure 2.10 Pavement families Model Parameters Estimations

2.5 Conclusions

This chapter has shown how the Bayesian linear regression model provides a more reliable framework for predicting when historical data is not available. The linear regression model is implemented within the constraints of Bayes' inference theory, in order to investigate the parameters estimation errors probabilistically.

The chapter consists of three main steps: interviewing pavement experts to determine the prior probability distribution of the; measuring the current road deterioration using IRI on selected samples from the road network. These first two steps can be done in parallel which save more time in the research. This followed by the posterior probability distribution calculation and thereafter the predictive distribution is calculated. The result is the linear regression of Bayes consisting of two parameters α and β. The model is expressed as $iri_i = \alpha_i + \beta_i t_i$. In addition, the credibility intervals for the model parameters were obtained for all pavement families as shown in the Tables 2.4 and 2.5 with the parameter estimates; this increases the reliability of the estimation range for the posterior probability distribution, as shown in Tables 2.6 and 2.7.

The use of this technique is recommended when developing a pavement performance model in the absence of historical data. Moreover, this method is not limited to a particular road network but applies to any road network when conditions are similar, especially in developing countries. The main limitations of this method are the lack of archived data, which results in relying on a methodology that adapts interviews with experts to overcome this issue. The formulation of the interview questions is considered a difficult and time consumption process. In addition, because of the lack of archived data, the authors divided the road network into 18 pavement families based on geographical location, traffic load and soil strength; this meant that 18 models were developed, which were sometimes impractical and often much more work is needed.

The recommendation of future work is proposing new variables for the model such as those related to the characteristics and the properties of the materials. The characteristics and properties of the materials are now not available. Once these variables are available, the predictive model will be more accurate and informative.

CHAPTER 3

FORECASTING PAVEMENT CONDITION TRENDS USING AUTOREGRESSIVE INTEGRATED MOVING AVERAGE

Abdussalam Heba [a] and Gabriel J. Assaf [b]

[a, b] Department of Construction Engineering, École de Technologie Supérieure, 1100 Notre-Dame West, Montreal, Quebec, Canada H3C 1K3

Paper submitted for publication, Nov 2020

3.1 Introduction

PMS consists of many essential network and project levels activities. Modeling of pavement deterioration or performance is one of the most important PMS network level activities that must be coordinately executed with other activities to have a proper PMS. Obviously, preventive maintenance of pavement is less expensive than reconstructing it after being deteriorated. Pavement deterioration models enable decision makers to optimally intervene in the pavement preventive maintenance tasks. However, analyzing the progression of pavement deterioration over time enables better understanding for the pavement functional behavior to efficiently support PMS. This chapter aims to investigate and forecast the trends of the pavement deterioration using Autoregressive Integrated Moving Average (ARIMA) time series analysis. This chapter was submitted to Advances in Civil Engineering Materials (ASTM) International journal and then has been rewritten to accommodate the continuity context of the thesis.

3.2 Problem Definition

The problem of this research is the lack of historical data that represents pavement condition and to overcome this problem, a method has been proposed to visually assess the pavement condition. IRI for all selected road sections were visually measured based on information extracted from Google Street view. Analyzing pavement condition general trend and understanding its behavior would support decision makers to determine when and how to intervein to improve the pavement condition with the consideration of allocated budgets. In this chapter, 18 pavement families road sections in two geographical zones (north and zone) were selected from the Libyan National Road Network. Pavement roughness indices are equally spaced time-correlated values so their general trend can be investigated and analyzed. Consequently, prediction model can be developed to forecast future using time series analysis and forecasting (Lethanh & Adey, 2013).

3.3 Methodology

Time series is an important prediction tool when data points indexed in time order. As mentioned above*, IRI* is a sequence taken at successive equally spaced points in time as shown below.

$$IRI = \{IRI_t : t \in T\} \tag{3.1}$$

Therefore, time series analysis can be used to investigate the *IRI* general trend and forecast its future changes (Hunt & Bunker, 2003). However, there are few technical conditions that must be available in the *IRI* data. *IRI* data must be stationary which means the *IRI* mean must be constant regarding to time and the variance and covariance should be equal at different time intervals from the mean as shown in Eqs. 3.1 and 3.2 (Wei et al., 2019).

$$VAR(IRI) = E(IRI_i^2) - E(IRI_i)^{\,2} = \sigma^2, \forall i\,; i = 1,2, \dots \dots ., n \tag{3.2}$$

$$COV(IRI, t) = E(IRI_i * t_i) - E(IRI_i)E(t_i) = \sigma_{IRI,t}^2\,, \forall i\,; i = 1,2, \dots \dots \dots, n \tag{3.3}$$

If the *IRI* data is nonstationary, a differencing of the *IRI* data can be applied to convert the data to stationary case. The aim of this study is to investigate the general trend of the *IRI* and develop a prediction model to forecast the pavement future condition.

3.3.1 Autoregressive Integrated Moving Average (ARIMA)

When data is nonstationary, it is preferable to use Autoregressive Integrated Moving Average (ARIMA) time series model. This method combines two models integrated together. The first part is the autoregressive (AR) models where a lag property must be available. Lag means each variable value in specific period is related to its variable values in the past periods. AR model with p lags is expressed as in Eq. 3.4 where $(\varphi_0, \ldots., \varphi_{t-p})$ are the model constants and IRI_p is the coefficient for the lagged variable in time $t-p$ (Ratnadip & Agrawal;, 2007).

$$IRI_t = \varphi_0 + \sum_{i=1}^{p} \varphi_i \, IRI_{t-i} + \epsilon_i \tag{3.4}$$

The following part is the moving average models (MA) which represent the possible relationship between the variable and the residuals from past periods. MA with q lags is expressed as in Eq. 3.5 where θ_q is the coefficient for the lagged error term in time $t-p$.

$$IRI_t = \mu + \epsilon_t - \sum_{j=1}^{q} \theta_i \, \epsilon_{t-\mathrm{j}} \tag{3.5}$$

A differencing process is applied to transform the data to stationary status when IRI data is nonstationary. A differenced variable, ΔIRI_t is a common method used to avoid this situation. First order difference is as in Eq. 3.6. When the first difference produces a stationary process, the variable IRI_t is integrated of order one.

$$\Delta IRI_t = IRI_t - IRI_{t-1} \tag{3.6}$$

3.3.2 ARIMA Model Order

An ARIMA model is defined by 3 terms: *p, d, q* where, *p* is the order of the AR term, *q* is the order of the MA term, and *d* is the number of differences needed for stationary time series(Peter J. Brockwell, 2002). By combining AR part which is shown in Eq. 3.4 with the MA part in Eq. 3.5, ARIMA model is resulted in Eq. 3.7.

$$\widehat{IRI}_t = \mu + \varphi_0 + \sum_{i=1}^{p} \varphi_i \, IRI_{t-i} + \epsilon_t - \sum_{j=1}^{q} \theta_j \, \epsilon_{t-j} \tag{3.7}$$

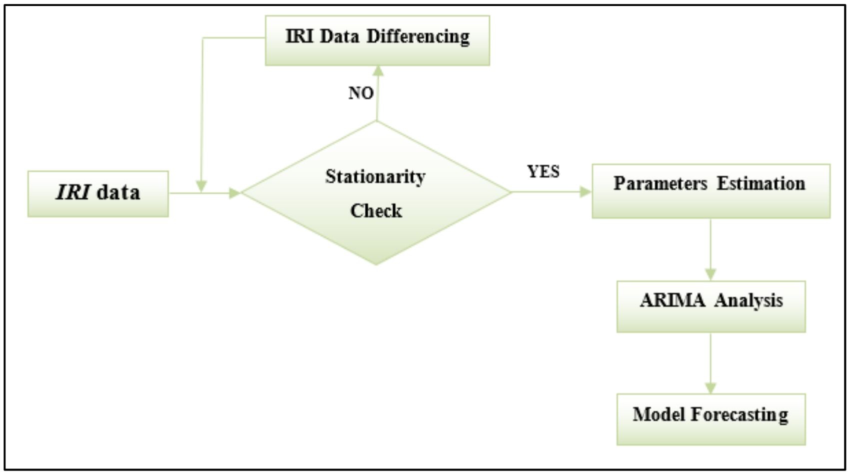

Figure 3.1 ARIMA Model Procedure. Taken from Peter (2002, p. 179-219).

Analyzing pavement condition general trend and understanding its behavior would support decision makers to determine when and how to intervein to improve the pavement condition with the consideration of allocated budgets. In this study, 18 Pavement families road sections in two geographical zones (north and zone) were selected from the Libyan National Road Network.

For the period from June 2010 to June 2020 on a semi-annual basis, IRI were recorded each six months for all selected pavement families. The data gathered were used to predict future IRIs before any road maintenance intervention was conducted. The IRI data is interpreted as time series data so that some important principles of analysis of time series can be applied. In this chapter, ARIMA model was applied for all pavement families. IRI time series data for three selected pavement families are presented in Fig. 3.2 below. From these figures, all pavement families data sets of IRI time series data were nonstationary. As a result, all IRI data sets were differenced in one order to make all time series data stationary. In this study all models order, and parameters were estimated and reported using SPSS ARIMA procedure.

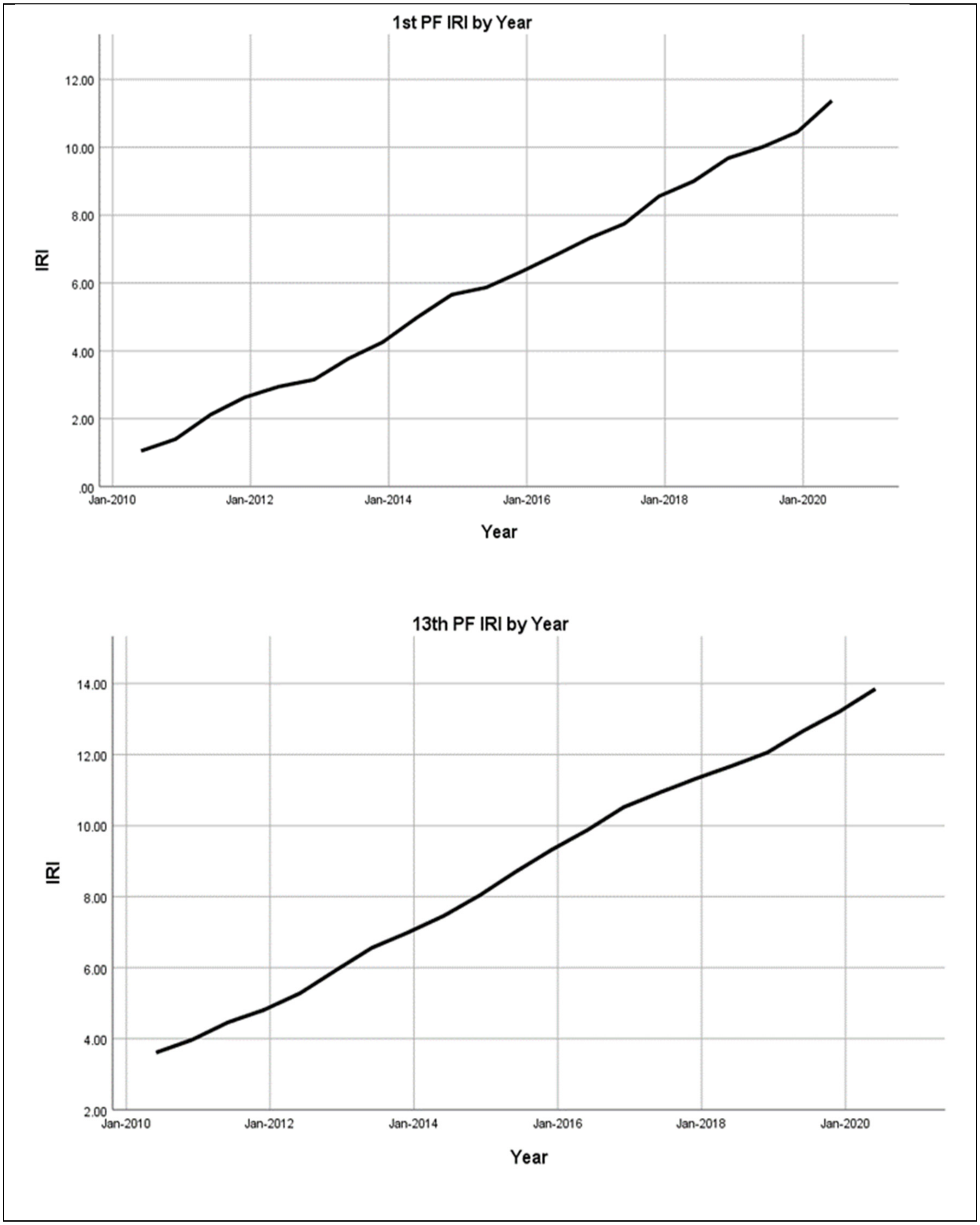

Figure 3.2 1st and 13th Pavement families IRI Correlating with Time (Jun 2010 - Jun 2020)

3.3.2.1 Autocorrelation Function (ACF) and Partial Autocorrelation (PACF)

The model *p* and *q* values were determined based on the Autocorrelation Function (ACF) and Partial Autocorrelation (PACF) coefficients. ACF refers to how time series data is compared with its previous values by investigating the indirect effect. However, PACF investigates the direct effect. ACF and PACF theoretical models show how to know model`s *p* order. In all pavement families IRI values, ACFs were timing out to zero and PACFs had cut off at value 1. Consequently, in this situation, it is recommended to assign 1 for the AR model which means $p=1$. Fig. 3.3 shows that ACF is timing out to zero quickly which indicates that IRI data for all pavement families were significantly correlated (Durango-Cohen, 2007).

PACF gives the partial association of a time series with its own lagged observations, adjusting time series observations at all shorter lags. It is unlike the role of autocorrelation (no power for other lags). Fig. 3.4 displays the plots of residual PACF for each estimated model for the selected pavement families, the number of delays exceeding 0.5 was 1, and this indicates that *p* arrangement can be 1.

There were many ARIMA models which might be suggested to represent the data depending on the normalized Bayesian Information Criterion (BIC) values. The most fitted model was with the smallest normalized BIC value. Then the most suitable ARIMA models for all pavement families were with the order (1,1,0) since all their BIC were small.

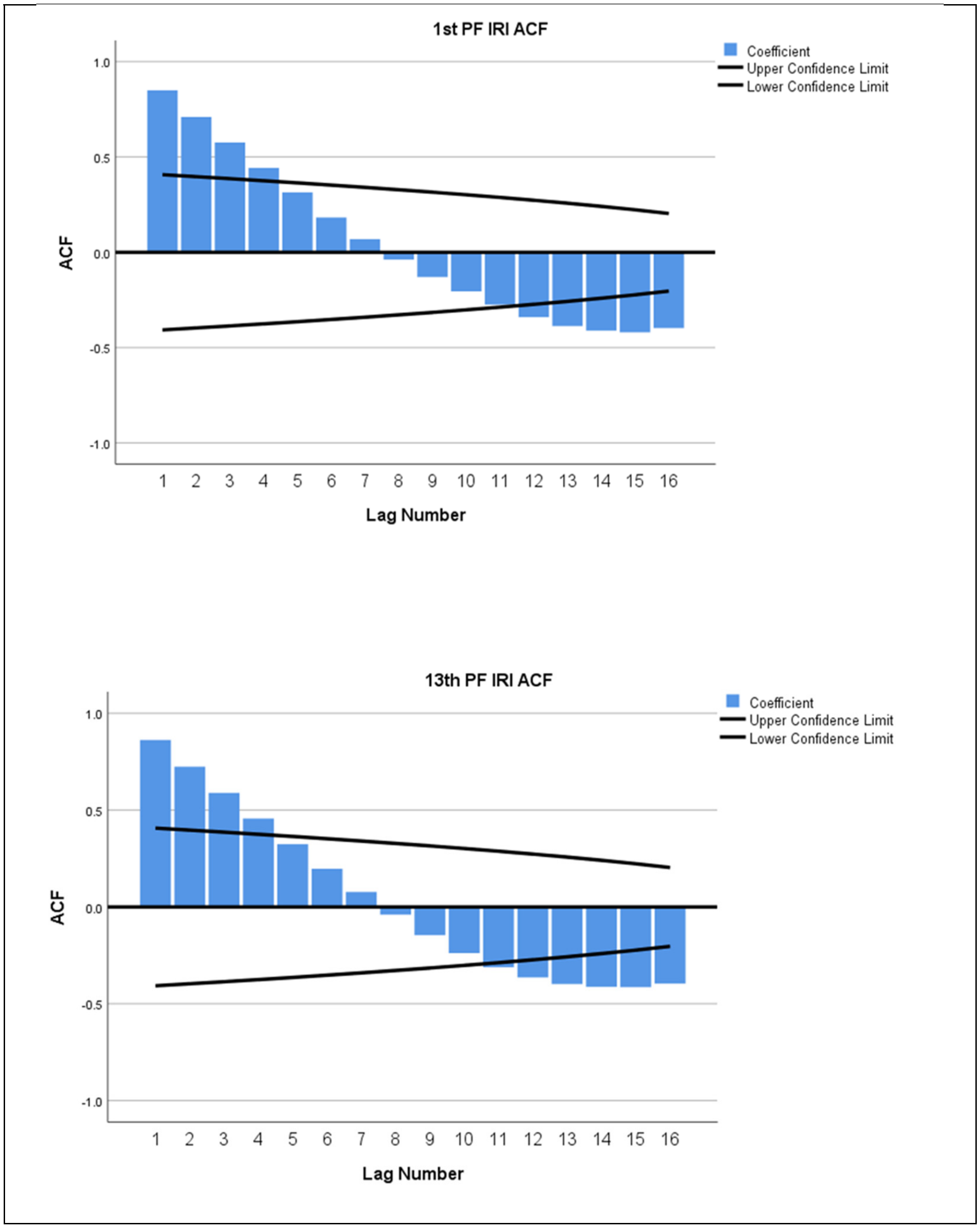

Figure 3.3 1st and 13th Pavement families IRI Autocorrelation Functions

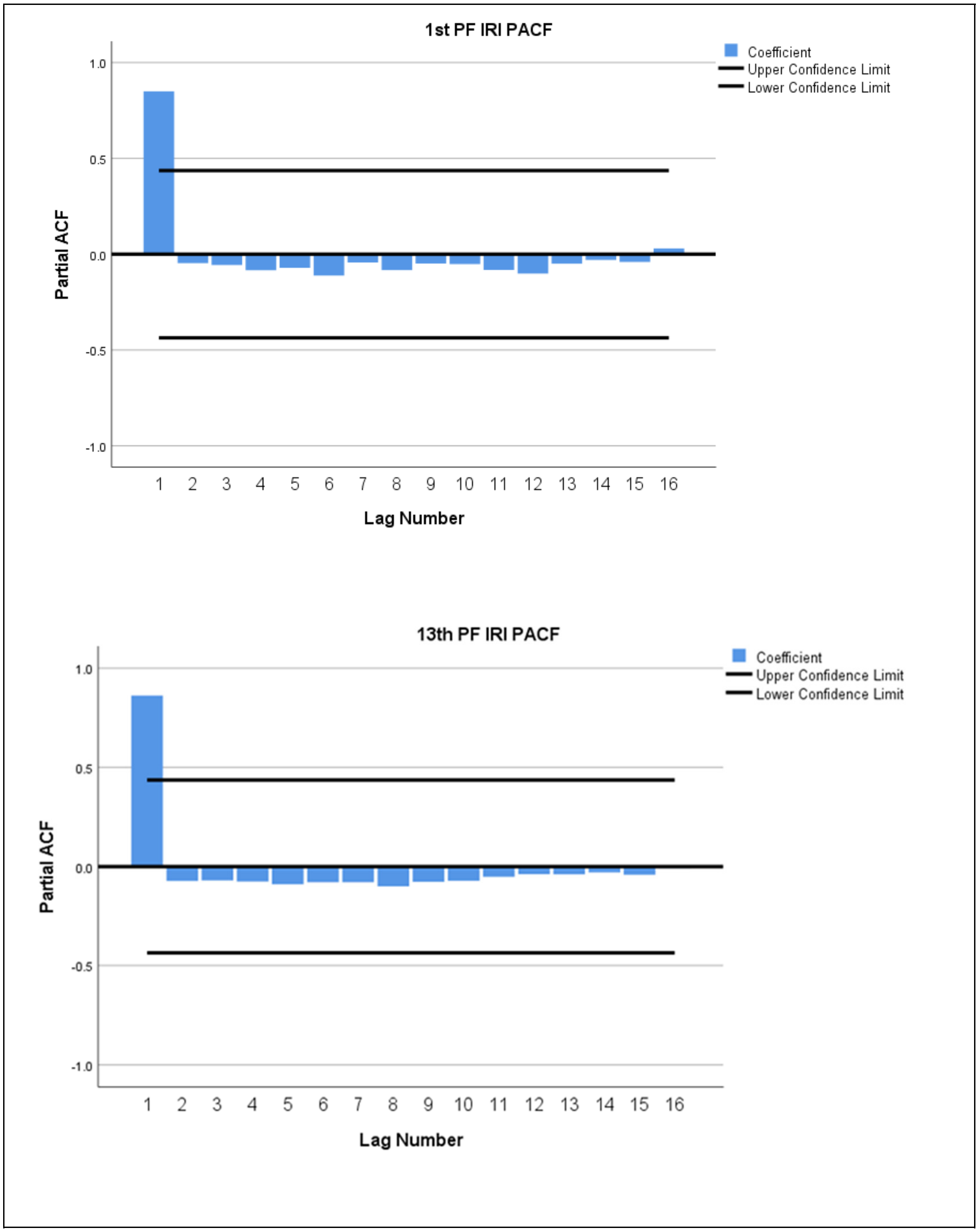

Figure 3.4 1st and 13th Pavement families IRI Partial Autocorrelation Functions

3.4 Models Results

Having tentatively established what would seem acceptable models, the next step was to obtain the least square estimation of the model parameters using SPSS to fit the IRI data to the models. That imply the following equation of evolution for the 18th pavement families models:

$$IRI_{t,f} = \alpha_f + \beta_f IRI_{t-1,f} + e_{t,f}\ ; \quad f = 1,2,\dots,18 \tag{3.8}$$

Equation 3.8 is the *IRI* estimation model that represents the 18 pavement families. $IRI_{t,f}$ can be defined as the estimated value of *IRI* in year *t* for the pavement family *f*. Each pavement family has two parameters which are α_f and β_f with $e_{t,f}$ error estimation. The order of all ARIMA(p,d,q) models were shown in Table 3.1. Model fit statistics: stationary R squared, and R squared were shown in Table 3.1. R squared is the *IRI* model regression coefficient determination which is the proportion of variance described by the model in the target variable. The value of R squared is ranging between 0 and 1. Small values indicate that the model does not fit the data well. To illustrate, first pavement model R squared value is 0.996 therefore, the model is suitable for representing the first pavement family data.

Table 3.1 shows also the Ljung-Box test which investigates the lack of fit of the IRI time series model. The test was applied to the residuals of the *IRI* time series after fitting the ARIMA model to the *IRI* data. The test examines m residuals autocorrelations, if the autocorrelations are very small, this indicates the model does not exhibit significant lack of fit.

The Ljung-Box statistics value is 19.38 for 17 d.f and P-value equal 0.31 which means the *IRI* model is significantly different from zero as shown in Table 3.1. Thus, the hypothesis was accepted, and this meant that the model is adequate as it absorbed all information about the IRI data. Table 3.1 shows that all models are suitable for representation the other pavement families models depending on Ljung-Box test.

Table 3.1 pavement families IRI Models Orders and Statistics

Pavements	Model Fit statistics			Ljung–Box Q(18)		
	BIC	Stationary R-S	R-squared	Statistics	DF	Sig.
1	-2.962	0.035	0.996	19.379	17	0.307
2	-2.462	0.161	0.994	12.898	17	0.743
3	-1.930	0.022	0.991	15.426	17	0.565
4	-2.131	4.39E-5	0.995	11.081	17	0.852
5	-3.087	0.187	0.997	9.192	17	0.934
6	-2.564	0.272	0.998	18.549	17	0.355
7	-2.837	0.004	0.997	11.403	17	0.835
8	-1.965	0.003	0.994	21.249	17	0.215
9	-3.025	0.017	0.997	9.969	17	0.905
10	-2.317	0.083	0.994	18.444	17	0.361
11	-2.292	0.202	0.995	10.007	17	0.903
12	-3.484	0.051	0.998	25.492	17	0.084
13	-4.145	0.128	0.999	15.311	17	0.573
14	-3.222	0.05	0.997	25.282	17	0.089
15	-3.398	0.058	0.997	21.778	17	0.193
16	-3.511	0.284	0.998	16.228	17	0.508
17	-3.652	0.024	0.998	11.566	17	0.826
18	-3.126	0.000	0.997	8.008	17	0.966

Table 3.2 illustrates the models' parameters estimations, In the 1st pavement families model, p-value of the model constant was $\alpha_1 = 0.514$ and the coefficient of AR was β_1 =-0.209. The values of the Standard Errors (SE) of the model constant and AR parameter were 0.037 and 0.265, respectively. The t-value for the parameter α_1 was 14.016 and -0.789 for the β_1. They were significantly different as their p-values were (0.000 and 0.440) for the model constant and AR at lag 1, respectively. Then, the 1st pavement family IRI model will be as follows:

$$\widehat{IRI}_{t,1} = 0.514 - 0.209 IRI_{t-1,1} + e_{t,1} \quad (3.9)$$

Where $\widehat{IRI}_{t,1}$ is the predicted value of IRI at time *t,* IRI_{t-1} is the value of *IRI* at the previous year, and e_t is the estimation error at year *t*.

Table 3.2 Pavement families IRI ARIMA Models Parameters

Pavements	Constants				AR Lag1			
	Estimation	SE	t	Sig.	Estimation	SE	t	Sig.
1	0.514	0.037	14.016	0.000	-0.209	0.265	-0.789	0.440
2	0.546	0.041	13.419	0.000	-0.398	0.215	-1.850	0.081
3	0.580	0.064	9.067	0.000	-0.153	0.234	-0.653	0.522
4	0.128	0.014	9.067	0.000	-0.007	0.248	-0.028	0.978
5	0.622	0.090	6.932	0.000	0.579	0.197	2.939	0.009
6	0.684	0.036	19.024	0.000	-0.480	0.204	-2.353	0.030
7	0.638	0.044	14.557	0.000	-0.067	0.238	-0.282	0.781
8	0.670	0.076	8.778	0.000	0.059	0.236	0.251	0.805
9	0.580	0.038	15.441	0.000	-0.135	0.233	-0.579	0.570
10	0.531	0.048	11.154	0.000	-0.282	0.226	-1.251	0.227
11	0.626	0.042	15.080	0.000	-0.485	0.219	-2.214	0.040
12	0.566	0.024	24.098	0.000	-0.413	0.228	-1.811	0.087
13	0.511	0.038	13.331	0.000	0.390	0.225	1.730	0.101
14	0.542	0.039	14.028	0.000	0.004	0.243	0.015	0.988
15	0.509	0.046	11.077	0.000	0.247	0.231	1.067	0.300
16	0.498	0.068	7.328	0.000	0.539	.199	2.701	0.015
17	0.472	0.026	17.988	0.000	-0.198	0.276	-0.715	0.484
18	0.598	0.040	15.033	0.000	-0.014	0.238	-0.061	0.952

3.4.1 Diagnostic Checking of the Model

To evaluate the appropriateness of all models, the residuals ACF and PACF were plotted. If the residuals are not within the bands, then there are some correlations, and the model should be improved. Fig. 3.5 shows ACF and PACF residuals for IRIs in the selected pavement families. Generally, if those plots are in the UCL and LCL band, then the residuals are random, and the model is suitable. All plots are clearly within the UCL and LCL bands as shown in Fig. 3.6. This means that the residuals are random and not correlated, which is an indication that the parameters of the model are appropriate for that data. Moreover, all models Ljung – Box (Q) statistics for testing of the residual are greater than their P-values which means that all models were not significantly different from zero as in Table 3.1. As a result, the null hypothesis of white noise was accepted, and this meant that all models fitted are adequate as they absorbed all information of the data.

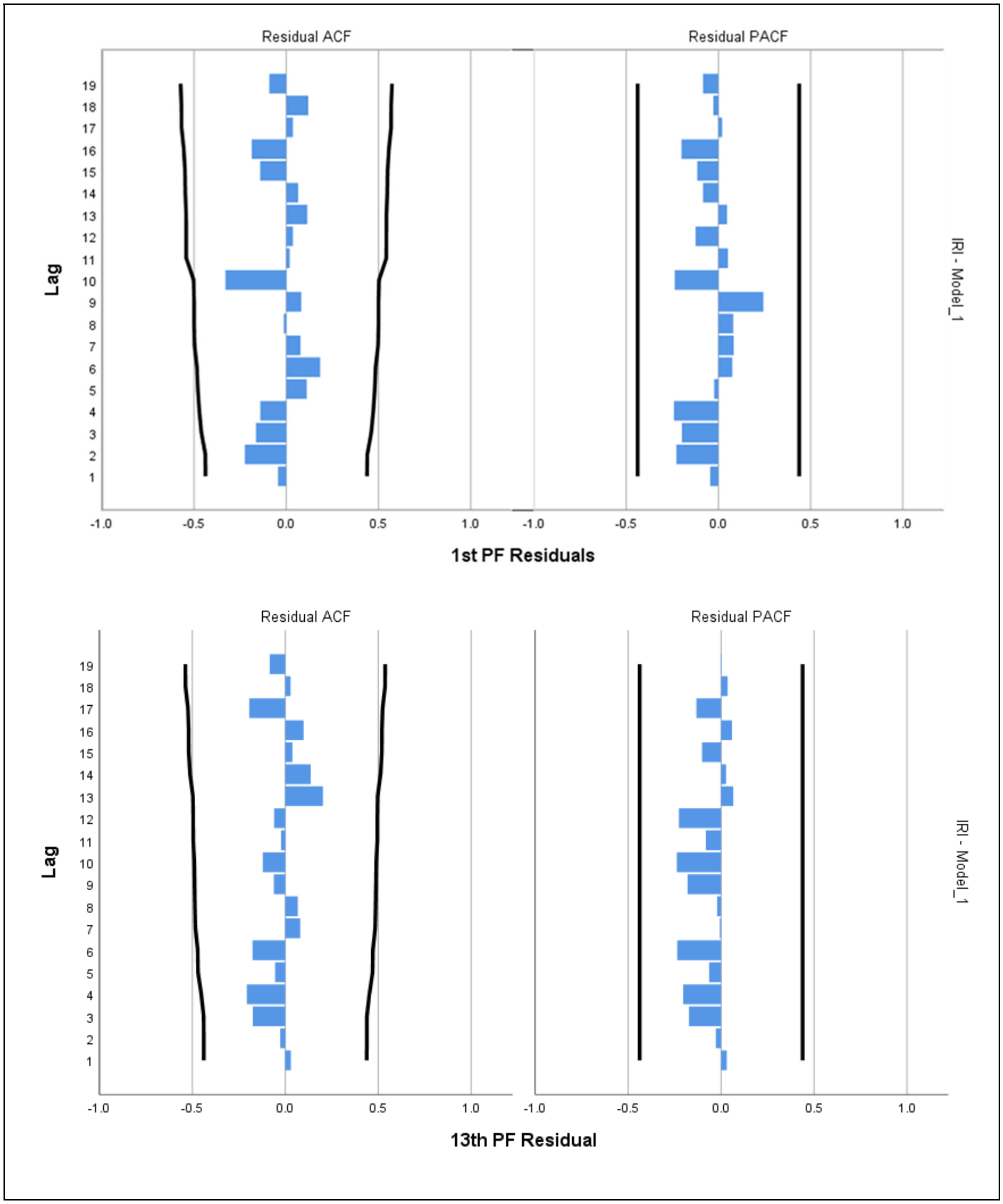

Figure 3.5 1st and 13th Pavement families IRI ACF and PACF Residuals

3.4.2 IRI Forecasting

After the most appropriate *IRI* models were identified, forecasting the *IRI* and investigating its trend were the following steps. Table 3.3 and Fig. 3.6 illustrate the results of the *IRI* forecasts using ARIMA (1,1,0) model for years from December 2021 to December 2023. All estimated *IRI* values were located between their 95% Upper Confident Level (UCL) and Lower Confident Level (LCL). For instance, and form Table 3.3, Dec 2021 *IRI 95%* estimation confident interval for 1st pavement family can be expressed as in inequality 3.10.

$$11.39 \leq \widehat{IRI}_{2021,1} \leq 12.21 \tag{3.10}$$

This means the first pavement family IRI forecasted value was between $[11.39, 12.21]$ with a probability of 95%. The forecasted values of all pavement families and their associated 95 % confidence limits are shown below in Table 3.3 for the IRI time series data.

Table 3.3 1st to 18th Pavement families IRI Forecast Results from June 2021-June 2023

Pavement families	IRI in 5 years estimated values														
	Dec 2021			Jun 2021			Dec 2022			Jun 2022			Dec 2023		
	UCL	E-value	LUL	UCL	E-value	LUL	UCL	E-value	LUL	UCL	E-value	LUL	UCL	E-value	LUL
1	12.21	11.80	11.39	12.85	12.33	11.81	13.47	12.84	12.21	14.07	13.36	12.64	14.66	13.87	13.08
2	13.25	12.72	12.20	13.90	13.28	12.67	14.55	13.82	13.09	14.19	14.37	13.55	15.82	14.91	14.01
3	14.40	1371	13.02	15.19	14.29	13.39	15.59	14.87	13.79	16.69	15.45	14.21	17.40	16.03	14.66
4	17.54	16.92	16.30	18.56	17.61	16.66	19.50	18.30	17.10	20.41	19.00	17.59	21.29	19.70	18.11
5	15.79	15.42	15.04	16.82	16.11	15.41	17.78	16.78	15.77	18.70	17.42	16.14	19.59	18.06	16.53
6	16.97	16.48	15.98	17.71	17.16	16.60	18.51	17.84	17.18	19.26	18.53	17.79	20.02	19.21	18.40
7	15.95	15.52	15.08	16.75	16.15	15.56	17.52	16.79	16.07	18.26	17.43	16.60	19.00	18.07	17.14
8	17.01	16.34	15.66	17.99	17.01	16.02	18.90	17.68	16.46	19.76	18.35	16.93	20.61	19.02	17.43
9	14.92	14.53	14.13	15.63	15.11	14.58	16.32	15.69	15.05	16.99	16.27	15.54	17.65	16.85	16.04
10	14.29	13.72	13.15	14.94	14.25	13.55	15.61	14.78	13.95	16.25	15.31	14.37	16.88	15.84	14.80
11	16.10	15.53	14.96	16.89	16.25	15.61	17.60	16.83	16.06	18.33	17.48	16.63	19.03	18.10	17.16
12	14.75	14.44	14.14	15.40	15.05	14.69	16.02	15.60	15.17	16.65	16.17	15.70	17.26	16.74	16.21
13	14.64	14.41	14.19	15.33	14.94	14.56	15.98	15.46	14.94	16.61	15.98	15.34	17.22	16.49	15.76
14	15.14	14.78	14.42	15.84	15.32	14.81	16.49	15.87	15.24	17.13	16.41	15.68	17.76	16.95	16.14
15	14.60	14.27	13.94	15.31	14.78	14.25	15.97	15.29	14.61	16.61	15.80	14.99	17.23	16.31	15.39
16	14.22	13.91	13.60	14.94	14.37	13.80	15.65	14.85	14.04	16.35	15.34	14.32	17.03	15.83	14.63
17	13.77	13.48	13.19	14.34	13.96	13.59	14.88	14.43	13.99	15.41	14.91	14.40	15.94	15.38	14.81
18	15.73	15.35	14.97	16.48	15.95	15.41	17.19	16.54	15.89	17.89	17.14	16.39	18.58	17.74	16.90

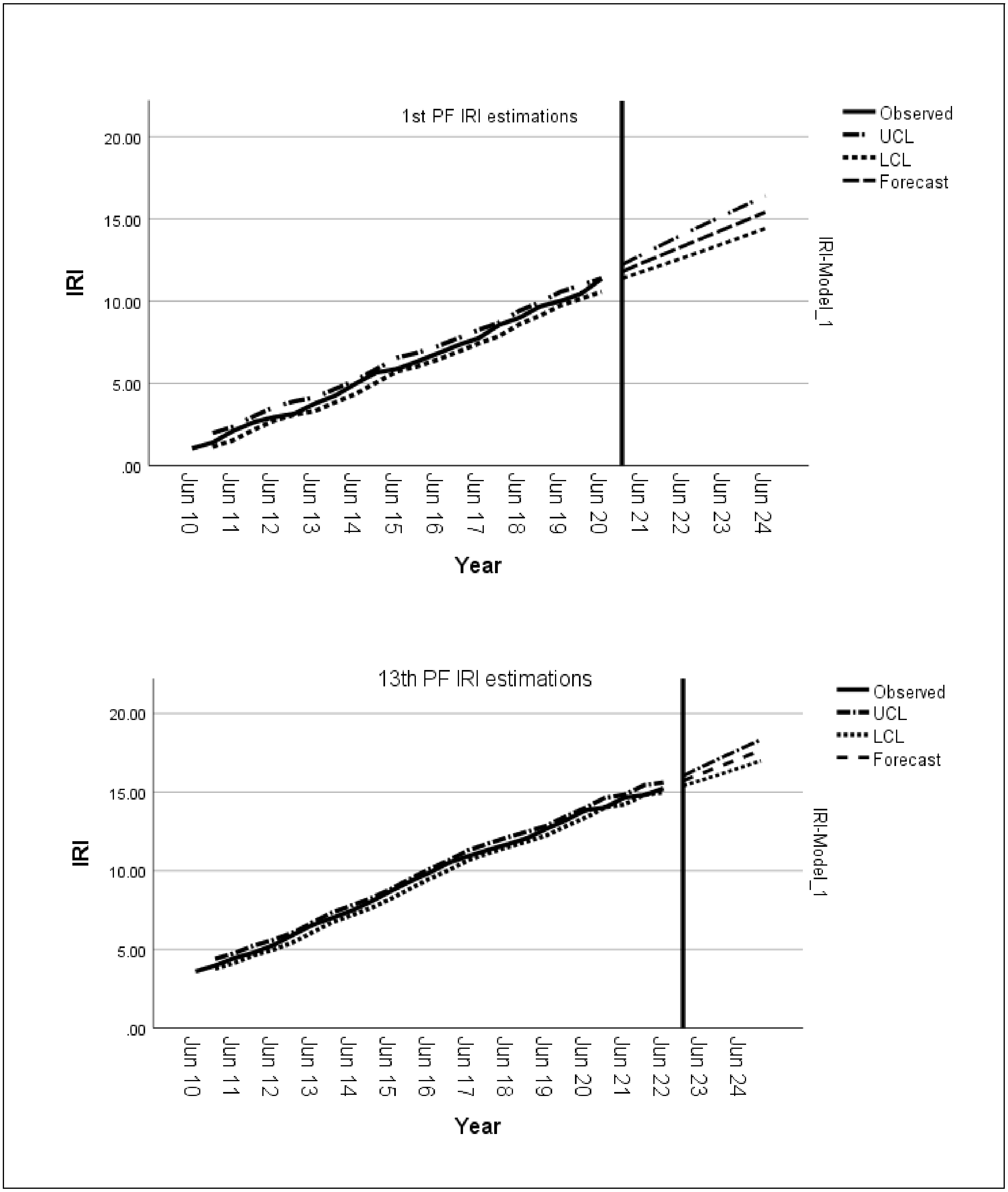

Figure 3.6 1st and 13th Pavement families IRI Estimations, UCL and LUL

The fitted ARIMA (1,1,0) model was used to forecast values of the IRI for years from June 2021 to June 2023. The original data and forecasted values for the selected pavement families with 95 % confidence intervals are shown in Fig. 3.6.

3.5 Conclusion

Pavement condition forecasting is an important activity in PMS. Its integration with other network level activities makes it an essential task that support top management decisions related to fund allocation priorities for pavement maintenance and rehabilitation. In this context, a time series models were developed for all pavement families to forecast their performance using ARIMA time series approach. In this study, 18 pavement families *IRI* half-year-based data between December 2021 to December 2023 were used to forecast the pavement condition trend. The sample ACF and PACF of the actual data were computed using SPSS modeler and their plots were graphed. These were used in identifying appropriate models. Pavement families data series exhibited non-stationary behavior following the inability to die or even at high lags. Each series was transformed by differencing once, and stationary was attained. The plot of each differenced series indicated that the series was evenly distributed around the mean. Following the of the ACF and PACF differenced series, an ARIMA (1,1,0) model for all pavement families given by $IRI_{t,f} = \alpha_f + \beta_f IRI_{t-1} + \varepsilon_t$ was identified for all pavement families. The parameters of fitted models were estimated. Models were subjected to statistical diagnostic check at 0.05 α using Ljung-Box tests for all pavement families with Q (18) and 17 *d.f.* and the BIC. Analysis showed no autocorrelation between residuals at different lag times and proved that the models were statistically significant, appropriate, and adequate. Fitted models were used to forecast values of the *IRI* from December 2021 to December 2023 on a semi-annual basis. The forecasts were good representation of the original data which neither decrease nor increase. Models of this study were built using half year-based data, so the seasonal analysis of the data was involved. A more comprehensive ARIMA model can be built in the future based on more accurate historical data.

CHAPTER 4

MODELING PAVEMENT DETERIORATION CONDITION USING GAUSSIAN HIDDEN MARKOV MODEL

Abdussalam Heba [a] and Gabriel J. Assaf [b]

[a, b] Department of Construction Engineering, École de Technologie Supérieure, 1100 Notre-Dame West, Montreal, Quebec, Canada H3C 1K3

Paper submitted for publication, March 2021

4.1 Introduction

Under normal circumstances, pavement deterioration modelling based on historical data correlating with its age, requires regular measurements of the condition rating of pavement sections over time. However, in the absence of such information and records in many cases such as in developing countries, such method cannot be used, and alternative is to use probabilistic modelling. Thus, Markovian approach is one of the appropriate technique that can be used in such cases. The hidden Markov model is a special case of Markov chain model, which has been widely used in several research areas such as image processing and speech recognition. One of the great advantages of the hidden Markov model is that it allows the unobserved condition state to be captured, eliminating the noise and bias associated with monitoring data. This method has been applied in this research to develop a Gaussian Hidden Markov deterioration prediction model to anticipate future pavement conditions. An initial database representing the pavement condition for an available period of times used in the development process. In addition to transition probabilities, emission probabilities are also calculated. This chapter was submitted to Advances in Civil Engineering Materials (ASTM) International journal and then has been rewritten to accommodate the continuity context of the thesis.

4.2 Problem Definition

It is very important in pavement management systems to provide a clear future vision of the pavement condition to support decision-makers. The aim of this study is to apply a hidden Markov modeling method to a road network that does not have archived data to predict future pavement condition. Under these difficulties, and to overcome the problem of lack of archived data a subjective data extracted from a previous study conducted on the same road network was used to develop the hidden Markov model.

4.3 Methodology

HMM is the model in which a series of emissions can be noticed, however it is difficult to know the path the model went through to generate these emissions. HMM is a tool to probabilistically estimate the sequence of cases from observed data. The hidden Markov model is a special case of Markov chain model, which has been widely used in several research areas such as image processing, speech recognition, and applied statistics (Davison, 2008).

Suppose that $Z_1, Z_2, \dots \dots, Z_n$; $X_1, X_2, \dots, X_n$ are random variables where $Z_1, Z_2, \dots \dots, Z_m$ is the pavement deterioration hidden process, and $X_1, X_2, \dots, X_n$ is the observed deterioration hidden process.

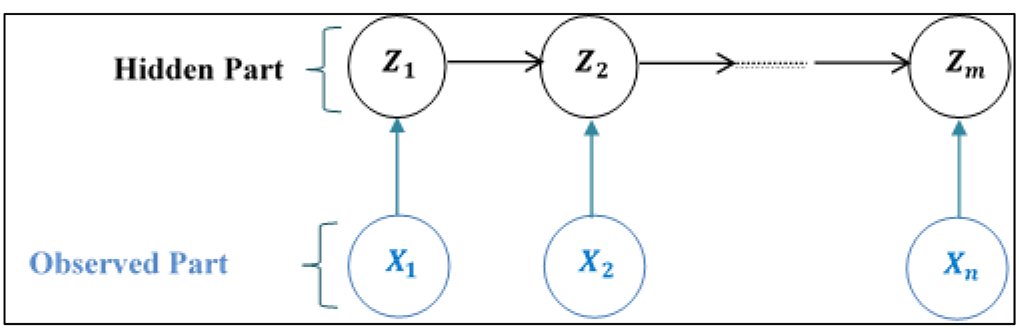

Figure 4.1 HMM Basic Diagram. Taken from Kobayashi et al (2012, p. 243)

The Markov hidden model is written as following:

$$P(x_1 \ldots, x_n, z_1, \ldots, z_m) = P(z_1)P(x_1|z_1)\prod_{k=2}^{n} P(z_k|z_{k-1})\,P(x_k|z_k) \qquad (4.1)$$

The goal is to compute, for all i, the posterior distribution $P(Z_m|X_1, X_2, \ldots, X_n)$ using a process of two steps which are prediction and then Bayesian update (see Fig. 4.2)

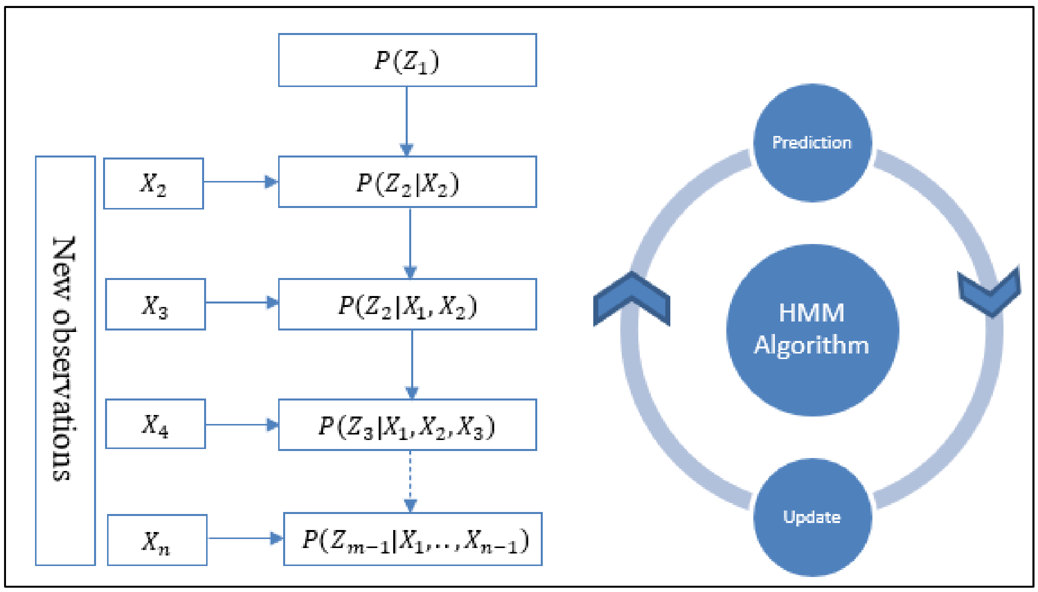

Figure 4.2 HMM Algorithm

4.3.1 HMM Parameters

The model mainly consists of three parts which are the transition probabilities, the emission probabilities, and the initial distribution.

4.3.1.1 Transition Probabilities

The first set of parameters is called the transition probabilities which represent the transition from a state to another. These set would be calculated using the following formula.

$$\pi_{ij} = P(z_{k+1} = j|z_k = i)\,;\,(i, j \ \in \{1,2, \ldots., r\})\,;\,\sum_{.j=1}^{J} \pi_{ij} = 1\,;\ \ 0 \geq \pi_{ij} \geq 1 \qquad (4.2)$$

An r*r matrix would be formed to hold all transition probabilities as following:

$$TPM = \begin{bmatrix} \pi_{11} & \cdots & \pi_{1J} \\ \vdots & \ddots & \vdots \\ \pi_{I1} & \cdots & \pi_{IJ} \end{bmatrix} \tag{4.3}$$

4.3.1.2 Emission probabilities (EM)

In order to specify HMM emission probabilities would be needed. These probabilities are calculated using the following formula.

$$\varepsilon_i(x) = P(x|z_k = i)\ ; i \in \{1,2, \dots, r\} \tag{4.4}$$

4.3.1.3 Initial Distribution

The initial probability distribution over states t_i is the probability that the Markov chain starts in state i. Some states j may have $t_j = 0$, meaning that they cannot be initial states.

$$t_i = P(Z_1 = i)\ ; i = \{1,2, \dots r\} \tag{4.5}$$

Then, the HMM formula is written in the following way:

$$P(x_1 \dots, x_n, z_1, \dots, z_n) = t(z_1)\varepsilon_{z_1}(x_1) \prod_{k=2}^{n} \pi_{z_{k-1}, z_k}\, \varepsilon_{z_k}(x_k) \tag{4.6}$$

In this chapter, Libyan road network was chosen as a case study and the data used was taken from a previous study in which the International Roughness Index (IRI) was estimated for the road network (Heba & Assaf, 2017).

A road is normally considered to be in perfect condition immediately after it is built and to deteriorate over time. The condition of the road can be described in different ways, for example discrete condition states may be used, $(i = 1, \dots\dots, I)$ where condition state 1 is considered perfect or like perfect and I alarming condition. Pavement performance indices represent the pavement physical or real condition (see Fig. 4.4). Unlike pavement condition states, pavement indices increase over time as in Fig.4.5.

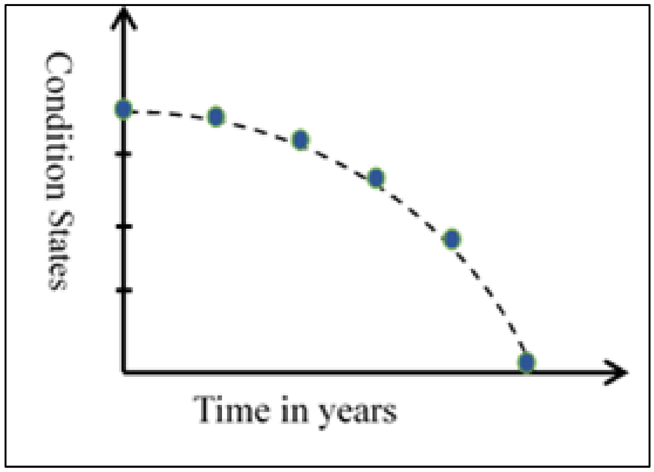

Figure 4.3 Evolution of pavement condition over time

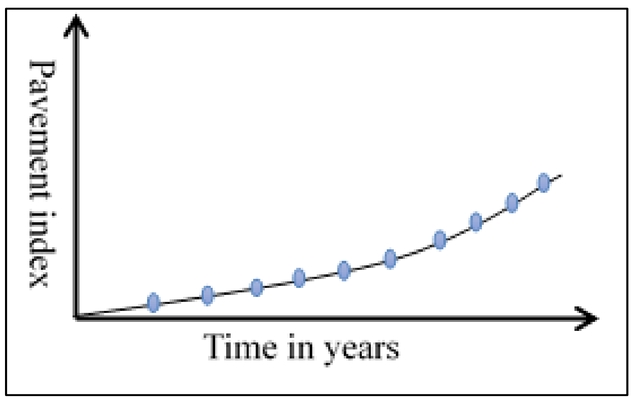

Figure 4.4 Evolution of pavement indices over time

To model the deterioration process using HMM, the probabilities of the process passing between condition states must be estimated. Available condition indicator values from expert pool were used to estimate the transition probabilities of the deterioration process. Accurate estimation of these transition probabilities requires adequate data on past conditions. In this paper, pavement condition data was obtained from the author previous paper (Heba & Assaf, 2018).

To use the HMM, it is important to consider that pavement physical deterioration cannot be directly observed. While it can be predicted through the evolution over time of one of the aspects those related to physical deterioration (e.g. roughness). As a result, pavement physical deterioration was considered a hidden process and pavement physical deterioration aspect considered as an observed process.

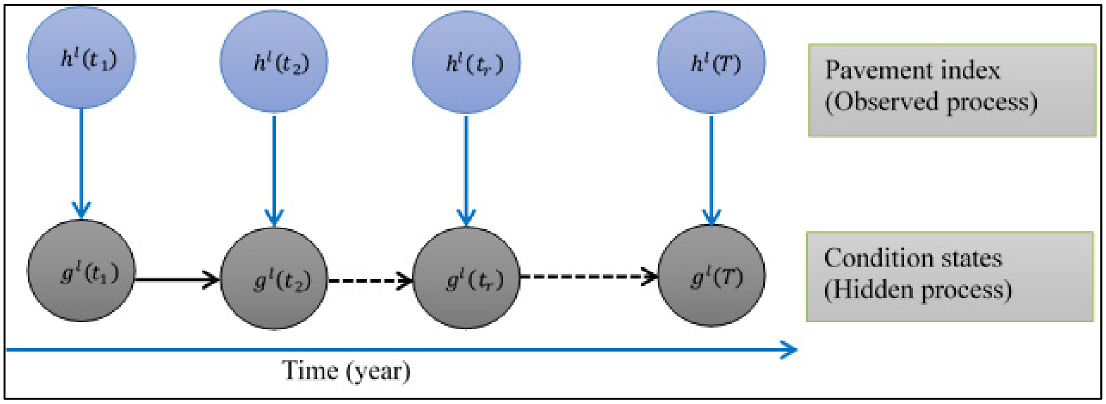

Figure 4.5 Pavement Index and Condition States Relationship

In this study, one pavement family was chosen to explain how to apply the HMM using forward algorithm. To apply HMM, prior or initial states probability distribution, states transition probability matrix, states observed conditions matrix (emission matrix) were calculated to develop the model.

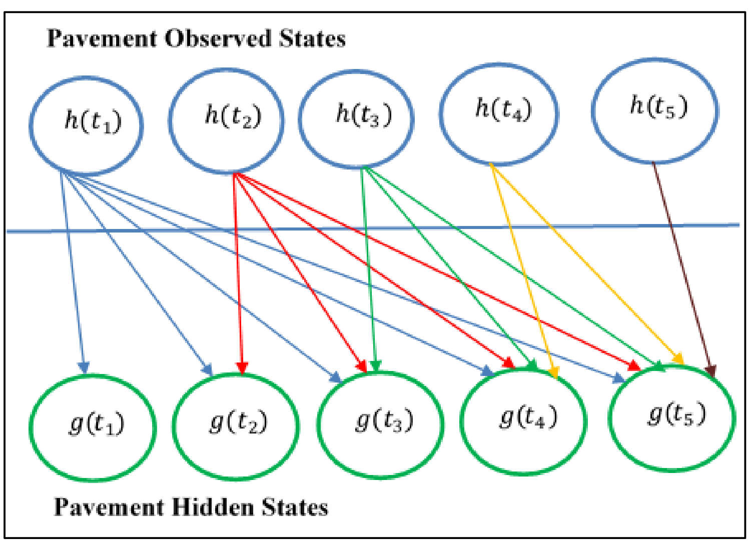

Figure 4.6 Pavement observed and hidden states

Fig. 4.4 shows the hidden and observable state in the pavement deterioration process. It is assumed that the hidden states (the true pavement condition) were modelled by simple first order Markov process. The TPM is in a forward move form, in which pavement condition states could move only forward to deteriorate to a lower pavement condition state. The connections between the pavement hidden states and the pavement observable states represents the probability of generating a particular observed state given that the Markov process is in a particular hidden state. It must be clear that all probabilities entering an observable state sum to 1, $\sum_{ij}^{n} g(t_{ij}) = 1$. In addition, each probability in the pavement TPM and EM is time independent.

Prior or initial pavement state probability distribution is the first HMM component was determined. Which is the initial probability of transitioning to a hidden state. This can also be looked at as the prior probability. The pavement condition was given the following condition states see Eq. 4.7.

$$Initial\ state\ distirbution = [IS]_{5\times1} = \begin{bmatrix} Ecxellent \\ Good \\ Medium \\ Poor \\ Fail \end{bmatrix} \tag{4.7}$$

Figure 4.7 represents 1st pavement family probability density functions (PDF) for experts' knowledge and pavement current condition data. Both functions are probabilistically distributed according to normal distribution see Table 4.1.

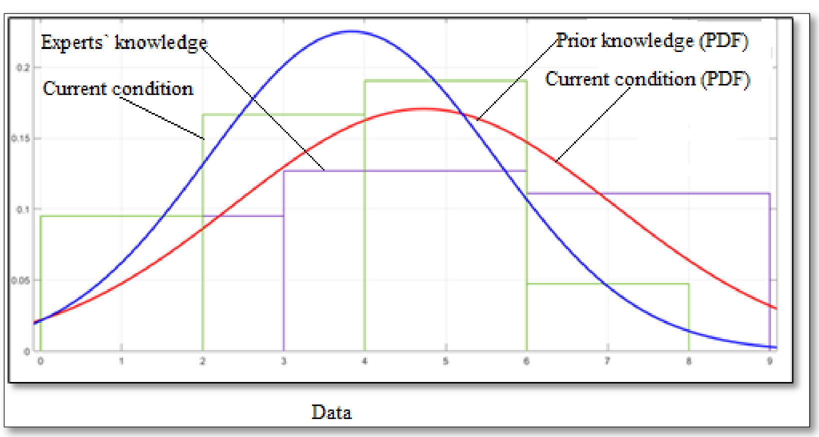

Figure 4.7 Pavement family prior and current condition PDFs

Table 4.1 1st Pavement family prior and current condition PDFs

	Pavement Experts Knowledge		Pavement Current Condition	
Probability Distribution	$N\sim(4.73{,}2.34)$		$N\sim(3.84{,}1.77)$	
Parameter Estimation	$\mu = 4.73$	$\mu_{std.Err} = 0.51$	$\mu = 3.84$	$\mu_{std.Err} = 0.39$
	$\sigma = 2.34$	$\sigma_{std.Err} = 0.38$	$\sigma = 1.77$	$\mu\sigma_{std.Err} = 0.28$
Parameter Covariance	$cov(\mu,\mu) = 0.29$	$cov(\mu,\sigma) = 2.66e^{-17}$	$cov(\mu,\mu) = 0.15$	$cov(\mu,\sigma) = -1.36e^{-17}$
	$cov(\sigma,\mu) = 2.66e^{-17}$	$cov(\sigma,\sigma)$ =0.14	$cov(\sigma,\mu) - 1.36e^{-17}$	$cov(\sigma,\sigma)$ =0.08

Pavement condition deteriorates by one or more condition state in one duty cycle. Which means condition states could move only one forward step or remain in the same condition. Total TPM steps in form are *n-1*.

Pavement states transition probability matrix is the second computed component of the HMM. TPM represents the probability of transitioning to a new state conditioned on a present state (see Eqs. 4.2 and 4.3). Eq. 4.8 and Eq. 4.9 show the TPM for the HMM.

$$Transition\ Probability\ Matrix = [TPM]_{5\times5} = \begin{bmatrix} g(t_{11}) & g(t_{12}) & g(t_{13}) & g(t_{14}) & g(t_{15}) \\ 0 & g(t_{22}) & g(t_{23}) & g(t_{24}) & g(t_{25}) \\ 0 & 0 & g(t_{33}) & g(t_{34}) & g(t_{35}) \\ 0 & 0 & 0 & g(t_{44}) & g(t_{45}) \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \quad (4.8)$$

$$= \begin{bmatrix} 0.73 & 0.11 & 0.10 & 0.04 & 0.02 \\ 0 & 0.65 & 0.15 & 0.12 & 0.08 \\ 0 & 0 & 0.75 & 0.13 & 0.12 \\ 0 & 0 & 0 & 0.97 & 0.03 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \quad (4.9)$$

Pavement observed states conditions matrix (Emission Matrix) which is the probability of transitioning to an observed state conditioned on a hidden state. This matrix was determined from the pavement observed condition (see Eq. 4.4). EM is result is shown in Eqs. 4.10 and 4.11.

$$Emission\ Matrix = [EM]_{5\times5} = \begin{bmatrix} h(t_{11}) & h(t_{12}) & h(t_{13}) & h(t_{14}) & h(t_{15}) \\ 0 & h(t_{22}) & h(t_{23}) & h(t_{24}) & h(t_{25}) \\ 0 & 0 & h(t_{33}) & h(t_{34}) & h(t_{35}) \\ 0 & 0 & 0 & h(t_{44}) & h(t_{45}) \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \quad (4.10)$$

$$= \begin{bmatrix} 0.83 & 0.09 & 0.05 & 0.02 & 0.01 \\ 0 & 0.70 & 0.16 & 0.09 & 0.05 \\ 0 & 0 & 0.61 & 0.25 & 0.14 \\ 0 & 0 & 0 & 0.56 & 0.44 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \quad (4.11)$$

4.3.2 Pavement Deterioration HMM Analysis

In this study, pavement deterioration process was considered an HMM case in which a sequence of emissions is observed, but the sequence of states the model has passed through to produce the emissions is unknown. Analysis of pavement HMM aims to capture the sequence of pavement condition states depending on the pavement observed condition.

4.3.2.1 Generating Random Sequence of Sates and Emissions

First step was to generate a random sequence of states and emissions from the HMM pavement model. This was beginning in state 1 at step 0, transitioning to $state_{1+i}$; $i \geq 1$.

4.3.2.2 Estimating States Sequence

The second step was the estimation of the most likely sequence of states the model would go through given the transition and emission matrices to generate a given sequence of emission. In this step Viterbi algorithm was used to compute this sequence. See Fig. 4.8.

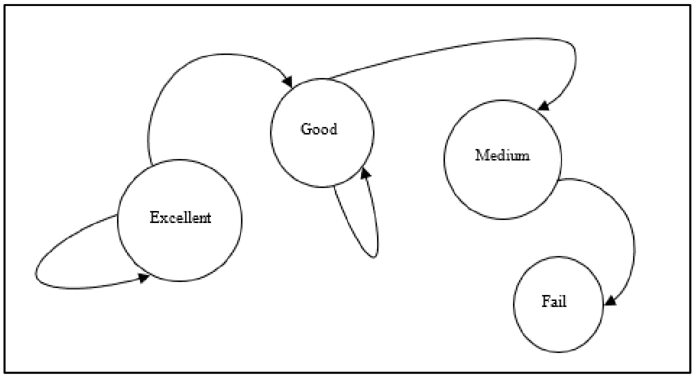

Figure 4.8 Viterbi algorithm 5 steps pavement condition states sequence

4.3.2.3 Estimating Posterior State Probabilities

After that, from the produced sequence of states which were generated using the Viterbi algorithm, transition and emission matrices were estimated. The posterior state probabilities are the conditional probabilities that the model was in a particular state when it generated a sequence of states, given that sequence was emitted. These probabilities were computed and produced the following pavement condition states sequence.

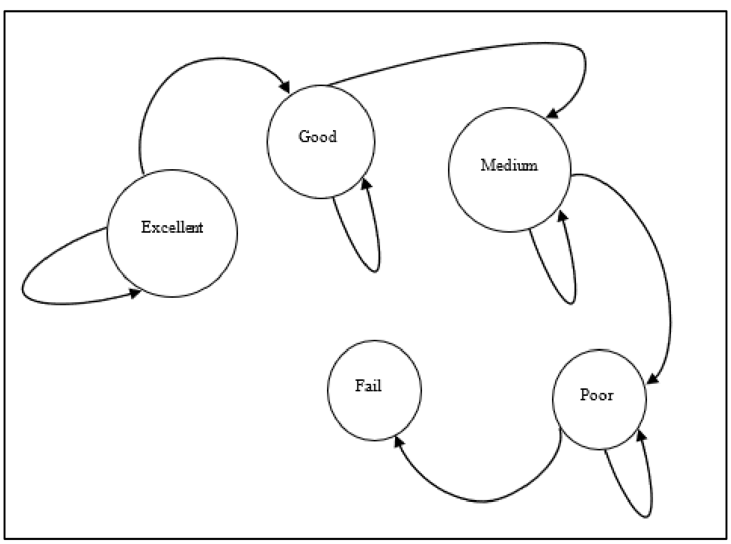

Figure 4.9 The Posterior pavement condition states sequence

Table 4.2 shows the posterior pavement condition state probabilities. It consists of the 10 steps that HMM has went through and 5 pavement conditions. For instance, in the 2nd step, there was a strong probability that the pavement condition will remain in the same condition which Excellent. While the pavement condition in this step could transit to next pavement condition which is Good with a probability of 0.002.

Table 4.2 Posterior pavement condition state probabilities

Step	Excellent	Good	Medium	Poor	Fail
1st	1.000	0.000	0.000	0.000	0.000
2nd	0.982	0.002	0.000	0.000	0.000
3rd	0.080	0.776	0.144	0.000	0.000
4th	0.000	0.680	0.200	0.093	0.027
5th	0.000	0.000	0.974	0.016	0.010
6th	0.000	0.000	0.870	0.009	0.121
7th	0.000	0.000	0.000	0.855	0.145
8th	0.000	0.000	0.000	0.703	0.297
9th	0.000	0.000	0.000	0.186	0.814
10th	0.000	0.000	0.000	0.003	0.997

4.4 Conclusion

In this study, 10 transitioning steps were taken to run the pavement HMM. However, it is recommended to use several transitioning steps that are equal or close to the pavement life spin. In pavement deterioration process, pavements go through multiple condition state until they reach the failure state. It is easy to approximately know when pavements fail and become unusable from their life span. However, it is difficult to know the condition states that a specific pavement went through to fail. This study explained how to estimate the most likely sequence of pavement condition states that a specific pavement has went through to fail. Depending on Figure 4.9 and table 4.2, the results of the HMM model can be easily interpreted. For example, in 1st transitioning step, the posterior probability was 1 which indicates to Excellent state and rest of other posterior probabilities are equal to 0. In the 7th transitioning step, pavement condition state is transferred to Poor state with a strong posterior probability of 0.855. While it could move to Fail state with a low posterior probability of 0.145. As a result, decision makers can decide when it is important to intervein to maintain or rehabilitate that pavement in accordance with the fund availability and maintenance priorities.

CONCLUSION

This thesis has shown the benefits of probabilistic models when data about road condition is insufficient in quality or quantity as well as if it is not available. In many deterministic models, there are uncertainties and variations in pavement design variables and parameters. As a result, it is not accurate to apply deterministic models to all situations of pavement management. In contrast, in probabilistic models, selection biases are considered as random variables, so these biases are eliminated through the assumption of the related probability distributions.

To probabilistically develop a pavement performance model, taking into consideration the quality of the available data, Bayes linear regression is an appropriate estimation solution. Bayes linear regression is the linear regression model applied under the constraints of Bayes' inference theory. In Bayes estimations, experts' opinion is used as prior knowledge instead of using archived data if it does not exist. The prior knowledge is then mixed with current road condition to produce the predictive distribution. The experts' knowledge data was exposed to ANOVA analysis and the results showed that there was no significant difference in group means. As a result, the experts' opinions were considerably compatible; this means that all the experts' knowledge about the roughness progression were close to each other. The current pavement condition was then measured, and the results used to derive the MLE function. The prior distribution and the MLE were used to produce the predictive distribution to estimate the model parameters. 95% confidence intervals for the model parameters were established around the estimated parameters; this increases the reliability of the estimation range for the posterior probability distribution.

For all pavement families, prediction models were built to forecast the pavement performance using ARIMA time series method. *IRI* half-year-based data between June 2021 to June 2023 were used to forecast the pavement condition trend. The sample ACF and PACF of the actual data were computed using SPSS modeler and their plots were graphed. These were used in identifying appropriate models. Pavement families data series exhibited non-stationary behavior following the inability to reach an unacceptable condition. Each series was transformed by differencing once, and stationary was attained. The plot of each differenced series indicated that the series was evenly distributed around the mean. an ARIMA (1,1,0) model for all pavement families was identified for all pavement families. The parameters of fitted models were estimated. Models were subjected to statistical diagnostic check at 0.05 α using Ljung-Box tests for all pavement families with Q (18) and 17 *d.f* and the BIC. Analysis showed no autocorrelation between residuals at different lag times and proved that the models were statistically significant, appropriate, and adequate. Fitted models were used to forecast values of the *IRI* from June 2021 to June 2023 on a semi-annual basis. The forecasts were good representation of the original data which neither decrease nor increase.

The third method used in this thesis to investigate the pavement performance is the HMM. The model consists of three parts which are the transition probabilities, the emission probabilities, and the initial distribution. To run the pavement HMM,10 transitioning steps were taken. The study explained how to estimate the most likely sequence of pavement condition states that a specific pavement has went through to fail using Viterbi algorithm. from the produced sequence of states which were generated using the Viterbi algorithm, transition and emission matrices were estimated. The posterior state probabilities are the conditional probabilities that the model was in a particular state when it generated a sequence of states, given that sequence was emitted.

RECOMMENDATIONS

The main purpose of this thesis was to develop pavement prediction models when archived or historical pavement condition data is not available using different statistical models, the following recommendations and future work are suggested to use and increase models' estimation accuracy:

1. The developed models could be used to predict pavement condition in countries with road networks where historical data is not available.

2. It is noticeable, in all developed models, that the estimated IRI value increases with time. Here, decision makers should make sure that proper maintenance treatment is placed at the right time. It is recommended to use preventive maintenance, because it is performed before the pavement shows significant distress. Furthermore, it is used to stop minor deterioration, delay progressive failures, and reduce the need for corrective maintenance.

3. In Bayes regression model, a more accurate prediction model can be obtained by constantly updating the prior distribution. I emphasize here that using Bayes linear regression is better than MLE. The main reason is that Maximum Likelihood Estimation returns a single fixed value (point estimation), but Bayesian linear regression inference returns probability density function.

4. Using longer time series and more accurate pavement condition visual estimator can increase ARIMA model estimation accuracy. However, there are some limitations when using ARIMA model. Some of them are, determining (p,d,q) order model is a little bit subjective and it is computationally expensive because it runs many different models.

APPENDIX I PAVEMENT FAMILIES' DATA

	Family 1		Family 2		Family 3		Family 4		Family 5		Family 6		Family 7		Family 8		Family 9	
Years	**Prior**	**LH**	**Prior**	**LH**	**Prior**	**LH**	**Prior**	**LH**	**Prior**	**LH**	**Prior**	**LH**	**Prior**	**LH**	**Prior**	**LH**	**Prior**	**LH**
1	1.05	1.15	1.22	1.32	1.25	1.35	1.24	1.36	1.44	1.54	1.47	1.54	1.53	1.53	1.56	1.65	1.62	1.59
2	1.29	1.35	1.46	1.52	1.49	1.55	1.49	1.57	1.68	1.74	1.71	1.74	1.77	1.73	1.81	1.87	1.86	1.80
3	1.75	1.65	1.92	1.82	1.95	1.85	1.96	1.87	2.14	2.04	2.18	2.04	2.23	2.03	2.31	2.18	2.32	2.11
4	2.10	1.90	2.27	2.07	2.30	2.10	2.32	2.13	2.49	2.29	2.54	2.30	2.58	2.28	2.69	2.45	2.67	2.36
5	2.45	2.15	2.62	3.32	2.65	2.35	2.67	2.39	2.84	2.54	2.90	2.55	2.93	2.53	3.07	2.72	3.02	2.62
6	2.80	2.40	2.97	2.57	3.00	2.60	3.03	2.65	3.19	2.79	3.25	2.80	3.28	2.78	3.45	2.96	3.37	2.88
7	3.15	2.65	3.32	2.82	3.35	2.85	3.39	2.90	3.54	3.04	3.61	3.05	3.63	3.03	3.82	3.25	3.72	3.14
8	3.50	2.90	3.67	3.07	3.70	3.10	3.74	3.16	3.89	3.29	3.97	3.30	3.98	3.28	4.2	3.52	4.07	3.4
9	3.85	3.15	4.02	3.32	4.05	3.35	4.10	3.42	4.24	3.54	4.32	3.55	4.33	3.55	4.58	3.79	4.42	3.65
10	4.20	3.40	4.37	3.57	4.40	3.75	4.46	3.68	4.59	3.79	4.68	3.80	4.68	3.80	4.96	4.06	4.77	3.91
11	4.90	3.90	5.07	4.07	5.10	4.95	5.17	4.88	5.29	4.29	5.40	4.30	5.38	4.30	5.71	4.59	5.77	4.43
12	5.25	4.15	5.42	4.32	5.45	5.15	5.53	5.18	5.64	4.54	5.75	4.55	5.73	4.55	6.09	4.86	6.15	4.78
13	5.60	4.40	5.77	4.57	5.80	5.60	5.89	5.48	5.99	4.79	6.11	4.80	6.07	4.80	6.46	4.88	6.53	5.04
14	5.95	4.65	6.12	4.82	6.15	6.05	6.24	5.78	6.34	5.04	6.53	5.34	6.42	5.05	6.84	5.15	6.91	5.61
15	6.30	4.90	6.47	5.26	6.50	6.35	6.60	6.31	6.69	5.29	6.89	5.61	6.77	5.30	7.22	5.42	7.29	5.89
16	6.65	5.15	6.86	5.65	6.85	6.75	6.96	6.78	7.04	5.64	7.25	5.8	7.12	5.65	7.6	5.69	7.67	6.28
17	7.00	5.40	7.17	5.78	7.20	7.10	7.31	7.04	7.39	6.19	7.61	6.56	7.47	6.20	7.97	6.06	8.06	6.89
18	7.35	5.65	7.52	6.12	7.55	7.25	7.67	7.34	7.74	6.84	7.97	7.25	7.82	6.85	8.35	6.66	8.44	7.61
19	7.70	5.90	7.87	6.47	7.90	7.56	8.03	7.76	8.09	7.19	8.33	7.62	8.17	7.20	8.73	7.35	8.82	8.00
20	8.05	6.45	8.22	6.82	8.25	7.83	8.38	8.18	8.44	7.54	8.69	7.99	8.52	7.55	9.11	7.73	9.20	8.39
21	8.40	7.33	8.57	7.27	8.60	8.10	8.74	8.20	8.79	8.00	9.05	8.48	8.87	8.01	9.48	8.11	9.58	8.90

1st to 9th pavement families' data including the experts' knowledge (Prior) and the likelihood estimations (LH)

	Family 10		Family 11		Family 12		Family 13		Family 14		Family 15		Family 16		Family 17		Family 18	
Years	**Prior**	**LH**	**Prior**	**LH**	**Prior**	**LH**	**Prior**	**LH**	**Prior**	**LH**	**Prior**	**LH**	**Prior**	**LH**	**Prior**	**LH**	**Prior**	**LH**
1	1.63	1.6	1.64	1.61	1.65	1.63	1.66	1.64	1.68	1.66	1.68	1.67	1.69	1.68	1.7	1.69	1.71	1.7
2	1.88	1.81	1.89	1.82	1.9	1.85	1.91	1.86	1.92	1.88	1.93	1.89	1.94	1.9	1.95	1.91	1.97	1.93
3	2.34	2.12	2.35	2.13	2.36	2.16	2.38	2.18	2.4	2.2	2.41	2.22	2.42	2.22	2.43	2.24	2.46	2.26
4	2.69	2.37	2.71	2.38	2.72	2.42	2.74	2.44	2.76	2.46	2.78	2.48	2.78	2.49	2.8	2.51	2.83	2.53
5	3.05	2.63	3.06	2.65	3.08	2.69	3.1	2.71	3.12	2.73	3.14	2.75	3.15	2.76	3.17	2.78	3.2	2.8
6	3.4	2.89	3.42	2.91	3.43	2.95	3.46	2.98	3.49	3	3.5	3.02	3.51	3.04	3.54	3.06	3.57	3.08
7	3.75	3.16	3.77	3.17	3.79	3.22	3.82	3.24	3.85	3.27	3.87	3.3	3.88	3.31	3.9	3.33	3.94	3.36
8	4.11	3.42	4.13	3.43	4.15	3.49	4.18	3.51	4.21	3.55	4.23	3.57	4.24	3.58	4.27	3.61	4.31	3.64
9	4.46	3.67	4.48	3.69	4.5	3.74	4.54	3.77	4.57	3.81	4.6	3.83	4.61	3.85	4.64	3.87	4.68	3.91
10	4.81	3.93	4.84	3.95	4.86	4.01	4.9	4.04	4.93	4.08	4.96	4.11	4.97	4.12	5	4.15	5.05	4.18
11	5.82	4.45	5.85	4.47	5.88	4.54	5.93	4.58	5.97	4.62	6	4.65	6.02	4.67	6.05	4.7	6.11	4.74
12	6.21	4.8	6.24	4.83	6.27	4.9	6.32	4.94	6.36	4.98	6.39	5.02	6.41	5.04	6.45	5.07	6.51	5.11
13	6.59	5.07	6.62	5.09	6.65	5.17	6.71	5.21	6.76	5.26	6.79	5.29	6.81	5.31	6.85	5.35	6.91	5.39
14	6.97	5.64	7.01	5.67	7.04	5.75	7.1	5.8	7.15	5.85	7.18	5.89	7.21	5.91	7.25	5.96	7.31	6
15	7.36	5.92	7.39	5.95	7.43	6.04	7.49	6.09	7.54	6.14	7.58	6.18	7.6	6.21	7.65	6.25	7.72	6.3
16	7.74	6.31	7.78	6.34	7.82	6.44	7.88	6.49	7.93	6.55	7.97	6.59	8	6.62	8.05	6.67	8.12	6.72
17	8.13	6.92	8.17	6.96	8.21	7.06	8.28	7.12	8.34	7.18	8.38	7.23	8.4	7.26	8.45	7.31	8.53	7.37
18	8.52	7.65	8.56	7.69	8.6	7.8	8.67	7.86	8.73	7.93	8.77	7.99	8.8	8.02	8.85	8.08	8.93	8.14
19	8.9	8.04	8.94	8.08	8.99	8.2	9.06	8.27	9.12	8.34	9.17	8.4	9.2	8.43	9.25	8.49	9.34	8.56
20	9.28	8.43	9.33	8.47	9.38	8.6	9.45	8.67	9.52	8.75	9.56	8.81	9.59	8.84	9.65	8.91	9.74	8.98
21	9.67	8.94	9.71	8.99	9.76	9.12	9.84	9.2	9.91	9.28	9.96	9.34	9.99	9.38	10.05	9.45	10.14	9.52

10th to 18th pavement families' data including the experts' knowledge (Prior) and the likelihood estimations (LH)

LIST OF REFERENCES

AASHO. (1962). *AASHO Road Test Report*. Retrieved from Washington, D.C.:

Amador-Jiménez, L., & Mrawira, D. (2011). Capturing variability in pavement performance models from sufficient time-series predictors: a case study of the New Brunswick road network. *Canadian Journal of Civil Engineering, 38*, 210-220. doi:10.1139/L10-127

Amador-Jimenez, L. E., & Mrawira, D. (2011a). Capturing Variability in Pavement Performance Models from Sufficient Time-Series Predictors: A Case Study of the New Brunswick Road Network. *Canadian Journal of Civil Engineering, Vol. 38, 2011,*, pp. 210–220.

Amador-Jimenez, L. E., & Mrawira, D. (2011b). Reliability-based initial pavement performance deterioration modelling.

Amador-Jimenez, L. E., & Mrawira, D. (2012). Bayesian Regression in Pavement Deterioration Modeling: Revisiting the AASHO Road Test Rut Depth Model. *Infraestructura Vial, 25*, 28-35.

Amin, M. S. R. (2015). The Pavement Performance Modeling: Deterministic vs. Stochastic Approaches. In *Numerical Methods for Reliability and Safety Assessment* (pp. pp 179-196): Springer.

Davison, A. C. (Cartographer). (2003). Statistical models

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

Du, Y., Liu, C., Wu, D., & Jiang, S. (2014). Measurement of International Roughness Index by Using Z-Axis Accelerometers and GPS. *Mathematical Problems in Engineering, 2014*(1), 10 pages.

Durango-Cohen, P. L. (2007). A time series analysis framework for transportation infrastructure management. *Elsevier, Transportation Research Part B 41 (2007)*, 13. doi:10.1016/j.trb.2006.08.002

Gatignon, H. (2010). *Statistical Analysis of Management Data*. In. doi:10.1007/978-1-4614-8594-0

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

Haas, R. (1994). *Modern pavement management*. Malabar Florida: Krieger Pub Co.

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

Heba, A., & Assaf, G. (2017). *Road Performance Prediction Model for the Libyan Road Network Depending on Experts' Knowledge and Current Road Condition Using Bayes Linear Regression*. Paper presented at the First GeoMEast International Congress and Exhibition, Egypt. https://link.springer.com/chapter/10.1007/978-3-319-61627-8_12

Heba, A., & Assaf, G. (2018). Bayes Linear Regression Performance Model Depending on Experts' Knowledge and Current Road Condition. *Advances in Civil Engineering Materials, 7*(3), 18. doi:10.1520/ACEM20170115

Hong, F., & Prozzi, J. (2005). Updating Pavement Deterioration Models Using the Bayesian Principles and Simulation Techniques. *First Annual Inter-University Symposium on Infrastructure Management(AISIM)*.

Hong, F., & Prozzi, J. A. (2006). Estimation of pavement performance deterioration using Bayesian approach. *Journal of Infrastructure Systems, 12*(2), 77-86. doi:10.1061/(ASCE)1076-0342(2006)12:2(77)

Hong, T., Chae, M. J., Kim, D., Koo, C., Lee, K. S., & Chin, K. H. (2013). Infrastructure asset management system for bridge projects in South Korea. *KSCE Journal of Civil Engineering, 17*(7), 1551-1561. doi:10.1007/s12205-013-0408-8

Hunt, P. D., & Bunker, J. M. (2003). *Time Series Analysis of Pavement Roughness Condition Data for Use in Asset Management.* Paper presented at the Proceedings rd Asia-Pacific Conference on Systems Integrity and Maintenance.

Jongsawat, N., & Premchaiswadi, W. (2010). Bayesian Network Inference with Qualitative Expert Knowledge for Decision Support Systems. 3-8. doi:10.1109/snpd.2010.10

Kajner, M. H. K. L. (1995). Predicting Roughness Progression of Asphalt Overlays. Retrieved from http://www.cshrp.org/products/bayesian/AB/ABTOC.HTM

Kobayashi, K., Kaito, K., & Lethanh, N. (2012). A statistical deterioration forecasting method using hidden Markov model for infrastructure management. *Transportation Research Part B: Methodological, 46*(4), 544-561. doi:10.1016/j.trb.2011.11.008

Lethanh, N., & Adey, B. T. (2013). Use of exponential hidden Markov models for modelling pavement deterioration. *International Journal of Pavement Engineering, Vol. 14*(No. 7), 645–654.

Lethanh, N., Adey, B. T., & Fernando, D. N. (2015). Optimal intervention strategies for multiple objects affected by manifest and latent deterioration processes. *Structure and Infrastructure Engineering, 11*(3), 389-401. doi:10.1080/15732479.2014.889178

Li, N., Haas, R., & Xie, W.-C. (1997). Development of a new asphalt pavement performance prediction model. *Canadian Journal of Civil Engineering, 24*(4), 547-559.

Li, N. Y., Haas, R., & C Xie, W. (2011). *Development of a new asphalt pavement performance prediction model* (Vol. 24).

Li, Z. (2005). *A Probabilistic and Adaptive Approach to Modeling Performance of Pavement Infrastructure.* (Ph.d), University of Texas at Austin,

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

Madanat, S. M., Karlaftis, M. G., & McCarthy, P. S. (1997). Probabilistic infrastructure deterioration models with panel data. *Journal of Infrastructure Systems, 3*(1), 4-9. doi:10.1061/(ASCE)1076-0342(1997)3:1(4)

Mašović, S., & Hajdin, R. (2013). Modelling of bridge elements deterioration for Serbian bridge inventory. *Structure and Infrastructure Engineering, 10*(8), 976-987. doi:10.1080/15732479.2013.774426

Michael W. Sayers, Thomas D. Gillespie, & Queiroz, C. A. V. (1986a). *The International Road Roughness Experiment*. Retrieved from https://perma.cc/7CRK-RFF4

Michael W. Sayers, Thomas D. Gillespie, & Queiroz, C. A. V. (1986b). *The International Road Roughness Experiment*

Establishing Correlation and a Calibration Standard for Measurements. Retrieved from Washington: https://perma.cc/7CRK-RFF4

N. Balakrishnan, Enrique Castillo, & Sarabia, J. M. (2006). *Advances in Distirbution Theroy, Order statistics, and Inference*. In.

Nagaraja, H. N. (2006). Inference in Hidden Markov Models. *Technometrics, 48*(4), 574-575. doi:10.1198/tech.2006.s440

Ningyuan Li, R. H., ANDWEI-CHAU XIE. (2014). Investigation of Relationship Between Deterministic and Probabilistic Prediction Models in Pavement Management. *Transportation Research Board, 1592*, 70-79.

Ortiz-Garcia, J. J., Costello, S. B., & Snaith, M. S. (2006). Derivation of transition probability matrices for pavement deterioration modeling. *Journal of Transportation Engineering, 132*(2), 141-161. doi:10.1061/(ASCE)0733-947X(2006)132:2(141)

Peter J. Brockwell, R. A. D. (Ed.) (2002). *Introduction to Time Series and Forecasting* (Second ed.): Springer.

Premkumar, L., & Vavrik, W. R. (2016). Enhancing pavement performance prediction models for the Illinois Tollway System. *International Journal of Pavement Research and Technology, 9*(1), 14–19.

Prozzi, J., & Madanat, S. (2003). Incremental Nonlinear Model for Predicting Pavement Serviceability. *JOURNAL OF TRANSPORTATION ENGINEERING, 129*(6), 635-641. doi:10.1061/(ASCE)0733-947X(2003)129:6(635)

Ralph, H., W. Ronald, H., & Lynne, C. F. (2015). *Pavement Asset Management*: John Wiley & Sons, Inc.

Ratnadip, A., & Agrawal;, R. K. (2007). *An Introductory Study on Time Series Modeling and Forecasting*.

Reem Alaamri, Rafeeq Kattiparuthi, & Koya, A. (2017). Evaluation of Flexible Pavement Failures-A Case Study on Izki Road. *International Journal of Advanced Engineering, Management and Science (IJAEMS), 3*(7), 741-749. doi:https://dx.doi.org/10.24001/ijaems.3.7.6

S.S.Adlinge, & A.K.Gupta. (2013). Pavement Deterioration and its Causes. *International journal of innovative research and development*, 09-15.

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

Shiferaw Wayessa, & Abuye, D. (2019). The Major Causes of Flexible Pavement Deterioration and Propose Its Remedial Measures: A Case Study Bako to Gedo Road, Oromia Region, Ethiopia. *American Journal of Engineering and Technology Management, 4*(1), 10-24. doi:10.11648/j.ajetm.20190401.13

Sultan Tarawne, & Sarireh, M. (2013). Causes of Cracks and Deterioration of Pavement on Highways in Jordan from Contractors' Perspective. *iiste, Civil and Environmental Research, 3*(10).

Wei, L., Ju, H., Liyang, X., Susan, T., & Lili, P. (2019). International roughness index prediction based on multigranularity fuzzy time series and particle swarm optimization. *ELSEVIER, 2*. doi:https://doi.org/10.1016/j.eswax.2019.100006.

Zumrawi, M. M. E. (2016). Investigating causes of pavement deterioration in Khartoum State. *International Journal of Civil Engineering and Technology, 7*(2), 203–214.

Zumrawi, M. M. E. (April 2016). Investigating causes of pavement deterioration in Khartoum State. *International Journal of Civil Engineering and Technology, 7*(2), 203–214.