CLUSTER-BASED COMPOSITE PAVEMENT DETERIORATION MODELING: A FRAMEWORK FOR INCORPORATING FLOODING

by

Ishan Neema

A Thesis submitted to the faculty of
The University of North Carolina at Charlotte
in partial fulfillment of the requirements
for the degree of Master of Science in
Construction and Facilities Management

Charlotte

2020

Approved by:

____
Dr. Omidreza Shoghli

____
Dr. Don Chen

____
Dr. Jacelyn Rice-Boayue

____
Dr. Stephanie Pilkington

ABSTRACT

ISHAN NEEMA. Cluster-based composite pavement deterioration modeling: A framework for incorporating flooding. (Under the direction of DR. OMIDREZA SHOGHLI)

Pavement networks are among the most valuable highway assets for a nation as billions of dollars have been invested every year in construction, maintenance, and rehabilitation. These networks undergo deterioration over time due to traffic loading, material characteristics, and environmental factors. Various prediction models were developed to predict pavement performance for several purposes, such as preparing an asset management plan, budget, and investment strategy. However, limited studies were found that were conducted on developing a probabilistic deterioration model for composite pavement networks. Also, most of the pavement management system's prediction models did not integrate flooding, and very few studies were found in this regard. Therefore, this research aims to develop a cluster-based pavement deterioration model through the Markov Chain and the Monte Carlo simulation analysis for composite pavements and propose a framework for incorporating flooding in the model. To this end, a case study was conducted on 102 pavement sections located in the United States' eastern region. For this purpose, the roughness, traffic loading, temperature, and precipitation characteristics from 2015 to 2019 was collected from the LTPP database. These pavement sections are grouped into three different clusters using the *K*-means clustering algorithm. Then, with the application of Markov chain analysis and Monte Carlo simulation, the pavement deterioration model for each cluster was developed. This deterioration model is utilized to predict the family-based

deterioration trend. The proposed framework for incorporating flooding is utilized to predict the pre-and-post flood IRI values of flood-affected pavement sections.

ACKNOWLEDGEMENT

I want to express my sincere and heartfelt gratitude towards everyone who helped me through my endeavor. Without your active guidance, help, cooperation, and management, I would not have come this far. I want to express my deep gratitude to my research advisor, Dr. Omidreza Shoghli, to allow me to do research and provide me with invaluable guidance through this research. His vision, sincerity, and motivation have deeply inspired me. It was a great privilege to work and study under his guidance. I would also like to thank my dissertation committee members: Dr. Don Chen, Dr. Jacelyn Rice-Boayue, and Dr. Stephanie Pilkington, for adding value to my journey in this research and giving me valuable feedback. I would like to thank my research team members, Arash Karimzadeh and Sepehr Sabeti, for providing me valuable assistance throughout this process. I am extremely grateful to my family and friends for their love, motivation, and care. They have played a vital role in preparing me for the future and supporting me through my ups and downs.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AC	Asphalt Concrete
AEP	Annual Exceeding Probability
CBR	California Bearing Ratio
CL_0	Cluster number 0
CL_1	Cluster number 1
CL_2	Cluster number 2
DOT	Department of Transportation
HMA	Hot Mix Asphalt
IRI	International Roughness Index
LCA	Life Cycle Analysis
LCC	Life Cycle Cost
LTPP	Long-term Pavement Performance program
M&R	Maintenance & Rehabilitation
MC	Markov Chain
MEPD	Mechanistic-Empirical Pavement Design
NCDOT	North Carolina Department of Transportation
NHS	National Highway System
PCI	Pavement Condition Index
PCR	Pavement Condition Rating
PMS	Pavement Management System
TAMP	Transportation Asset Management Plan

TPM	Transition Probability Matrix
WIM	Weigh-in-Motion

CHAPTER 1: INTRODUCTION

1.1 Overview

Roads and highway networks are the bloodlines for any nation as it contributes to the nation's economic development by providing employment, ease of doing business, and it plays a vital role in social, educational and health development by connecting people and products. Highway asset is the most valuable asset for any transportation department (DOT), as billions of dollars have been invested every year in its construction, maintenance, and rehabilitation. The budget for highway construction and maintenance is prepared based on the state's revenue generated by the DOT. For example, the North Carolina Department of Transportation (NCDOT) forecasts the generation of revenue for the next ten years and, based on these forecasts, prepares the transportation system's maintenance and rehabilitation budget.

The revenue generated by the NCDOT is grouped into three major categories: highway fund, highway trust fund, and federal funds (NCDOT, 2019). The highway fund is used to construct, maintain, and rehabilitate 80,000 miles of highway network of North Carolina (NCDOT, 2019). The fund estimated by the NCDOT for the investment for highway maintenance, pavement program, and bridge program for the fiscal year 2019-20 is USD 1.63 billion (NCDOT, 2019). The preparation of the budget for pavement maintenance and rehabilitation depends on the pavement's predictive performance in future years. The pavement performance is predicted by evaluating the amount of change in pavements' characteristics such as roughness, rutting, structural strength, and more. The change in the value of these characteristics of the pavement is termed as deterioration. The pavement characteristics are dependent on the traffic loading, environmental condition, quality of materials, and geographical condition.

Pavements tend to deteriorate throughout their lifespan, which can be predicted using a deterioration model. A deterioration model is developed by performing mathematical analysis on the pavements' characteristics such as variable traffic loading, deterioration in material quality, and surrounding geographical & environmental conditions. In some studies, the deterioration model developed was common for all the pavement sections; however, it is not appropriate to have a single deterioration model for a variety of pavements. Sunitha et al. (2012) has compared the deterioration model developed by, with, and without clustering pavement sections. Their comparison illustrates that a single prediction model developed for all pavement sections will either underestimate or overestimate pavement conditions. Also, when large pavement stretches need to be maintained, prioritizing a particular pavement section's maintenance work becomes complicated. In such a situation, pavement sections' clustering is a useful tool for developing a section-wise maintenance strategy (Sandra & Sarkar, 2015). The deterioration trends of pavement sections differ as the characteristics of sections differ. Therefore, all pavement sections cannot be considered a single entity; instead, they should be grouped based on their characteristics' similarity, and the deterioration model for each group can be generated separately. Therefore, the clustering of pavement sections is preferred for developing effective deterioration models and maintenance strategies. The emission of greenhouse gases has increased substantially every decade, which accounts for the rise in temperature and, ultimately, climate change. Various research suggests that climate change leads to weather extremes, such as massive floods, snow, and temperatures. The research done by Paerl et al. (2019) analyzed the rainfall pattern of the coastal parts of the state of North Carolina and found that 36 tropical cyclones occurred in the state, which includes Hurricane Florence (2018) and Hurricane Matthew (2016), since the late 1990s. They estimated that the precipitation and flooding would increase in North Carolina as tropical cyclones' frequency will increase. Previous studies show that climate change is evident. Therefore, this subject cannot be neglected while developing a budget for the construction and maintenance of pavement as its characteristics are dependent on climatic conditions, and billions of dollars are associated with it. The pavement maintenance budget is developed based on predicted pavement performance using deterioration models. However, the pavement deterioration model, which accounts for extreme weather events such as frequent heavy rainfall, floods, and snow, was rarely found in the literature.

1.2 Background

Natural disasters and extreme weather events such as flooding, frequent heavy rainfall, and snow contribute to deterioration in pavement more quickly than normal weather conditions. Various studies are conducted in the past by different researchers for understanding the impact of flooding on the pavement network. In 2005, two hurricanes, Katrina & Rita, hit New Orleans and the southeastern part of Louisiana, United States. Approximately 2,000 miles of road length was submerged in flood runoff for five weeks (Sultana et al., 2018). Zhang et al. (2008) assessed the effect of the hurricane, occurred in New Orleans, and found that there was a significant difference in the structural strength of pavement between the submerged and non-submerged pavement sections. Helali et al. (2008) studied the distress data of these hurricane-affected pavements for pre-and post-hurricane time. They showed that the deterioration rate of flooded sections was 2.5 to 6.5

times more than the non-flooded pavement sections. They also estimated that 90 to 190 mm of the AC overlay is required for the rehabilitation work. In 2012, a tsunami struck Japan, which damaged thousands of miles of pavement sections. After the event of flooding, it was found that the pavement sections are substantially damaged and require reconstruction, or these pavements sections are serviceable only after rehabilitation work (Tokuyama, 2012). In 2010-11, extreme flooding events occurred in Queensland and New South Wales, Australia. This flooding caused damage to more than 21,120 miles of pavement network, and approximately AU$ 6.4 billion was spent in the rehabilitation and reconstruction of flood-affected pavements (Kenley & Harfield, 2014). A recent study was conducted on these flood-affected pavements by Chowdhury et al. (2016) to understand pavement's structural and surface conditions during pre-flood and post-flood time. They developed a deterministic road deterioration model and found that pavement tends to lose its strength more rapidly due to flooding. Khan et al. (2014a) also conducted a study on these flood-affected pavements and developed a probabilistic road deterioration model by incorporating flooding effects. All the previous research suggests that natural disasters, especially flooding and frequent heavy rainfall, affect pavement performance and budget.

The DOTs have developed a wide-ranging pavement management system (PMS) for effectively managing their road assets (Hudson et al., 1998). A good PMS is consists of five main components which are (i) data collection, (ii) quality database, (iii) accurate pavement deterioration model for decision-making strategies, (iv) procedures for implementing these strategies, and (v) feedback (Battiato et al., 1994). For example, the NCDOT collects detailed information of all their highway assets in the pavement management system (PMS) to analyze and evaluate the transportation system. They used an automated data collecting process to extract pavement information and derive the pavement condition rating (PCR) of each segment. Each highway segment's summary score is calculated using PCR, known as the pavement condition index (PCI), reflecting the highway's overall condition. The PCI is measured on the scale of 0 to 100, where 0 is failed, and 100 is excellent (NCDOT, 2019). The PCI predicts pavement performance by developing deterioration trends and determining the pavement's life cycle cost (LCC). Most of the DOTs do not integrate the effects of flooding in their prediction models. This directly results in inaccurate pavement performance prediction and determining the pavement's life cycle cost (LCC).

1.3 Problem Statement

In this research, a comprehensive literature review was conducted to determine gaps in the literature. The gaps identified in the literature are (not limited to):

1. Limited research was conducted on developing probabilistic pavement deterioration models for composite pavement networks,
2. Very few research were found that utilized the clustering algorithm and incorporated flooding in the deterioration model, and
3. Most of the DOT’s PMS does not incorporate the effect of flooding in their prediction models.

This research work is focused on filling the gaps that are identified above.

1.4 Research Objective

The clustering algorithm is a sophisticated technique used for grouping pavement sections into different groups based on their similarity. This technique can improve the accuracy of the deterioration model. The pavement's roughness is stochastic; therefore, a probabilistic method can be utilized to predict pavements' roughness more precisely. Extreme weather events such as frequent heavy rainfall and flooding adversely affect pavement performance and budget. Inundation in pavements due to flooding and heavy rainfall introduces new parameters in the traditional road deterioration models. Therefore, it is essential to understand and document the effect of flooding in the deterioration model. Based on these requirements, the primary research objective for this thesis work are as following:

1. Development of a cluster-based composite pavement deterioration model through Markov Chain and Monte Carlo simulation analysis,
2. Proposing a framework for incorporating flooding in a pavement deterioration model and generating deterioration trends at the various probability of flooding, and
3. Utilization of the model in predicting the pavement performance based on roughness for a short duration period.

This research is focused on developing a cluster-based pavement deterioration model using the Markov chain analysis and incorporating the effects of flooding in the model. The states' PMS can be improved by integrating the effect of flooding in their deterioration model using the framework proposed in this research. This model can be used for developing a predictive maintenance strategy for flood-affected and non-flood affected pavement precisely. It can also improve the Life Cycle Cost Analysis (LCCA) of the pavement network and help prepare effective budgeting and investment strategy.

CHAPTER 2: LITERATURE REVIEW

A systematic literature review was conducted to identify the gaps in the literature and develop the research objective. The literature review is divided into five major sections and which are further divided into different categories. Figure 1 shows an overview of the topics covered in the literature review. These topics are explained in detail in this section.

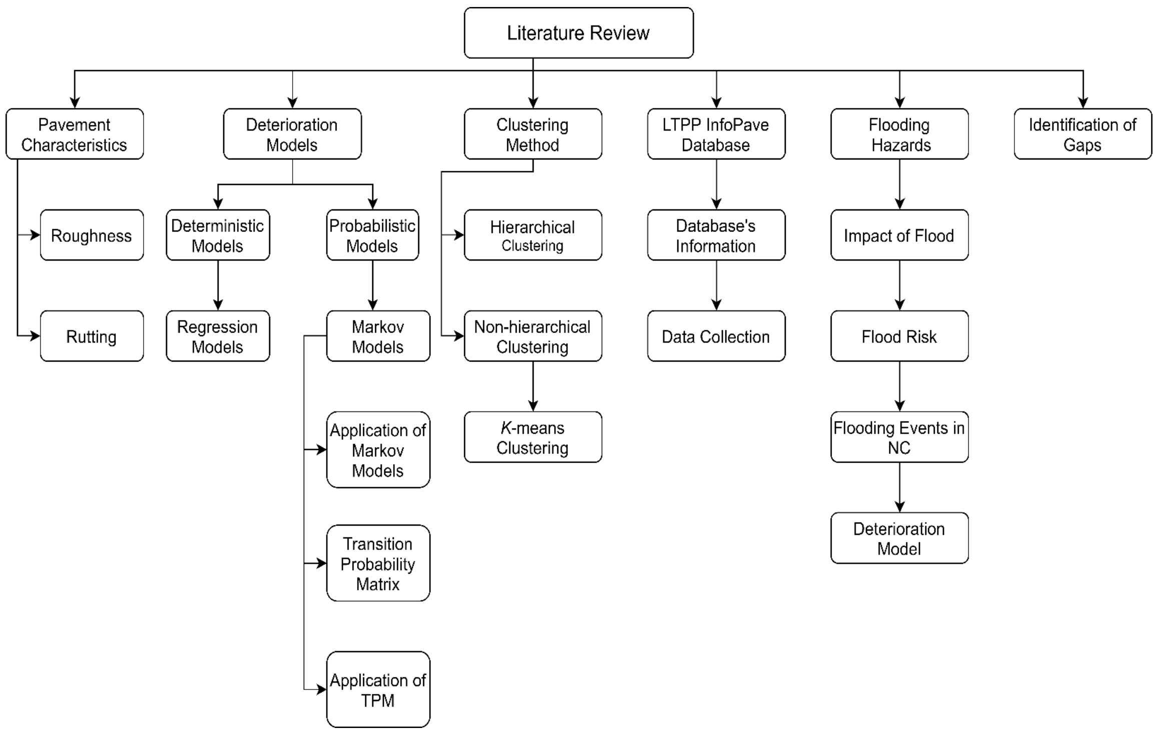


FIGURE 1: Topics covered in the literature review

2.1 Pavement Characteristics

The performance of the pavement tends to decrease gradually throughout its lifespan. The evaluation of pavements' performance is based on its characteristics such as roughness, rutting, and structural strength. The change in the value of these characteristics of pavement is termed as deterioration, and the ability of pavement to be serviceable over time by satisfying its designed requirements is termed as pavement performance.

Developing pavement performance models is essential in pavement engineering because it is used to design and perform maintenance analysis. The first believed pavement performance model was developed by the American Association of State Highway Officials (AASHO) in the 1960s in the form of equation $y = y_0 + bx^c$; where y_0 represents the pavement's initial performance, x represents accumulated traffic level, and b and c are the parameters to be estimated (Chen et al., 2019). Various parameters influence pavement design, such as the surrounding climate, traffic loading, structural capacity, and material requirements. Various researchers had analyzed the impact of climate and traffic on pavement performance for developing a better pavement design.

Change in climatic conditions such as hefty rainfall and floods influences the pavement's structural and surface condition. These parameters are dependent on various factors, such as environmental factors, traffic loading, and the surrounding geography. The environmental conditions and traffic loading are stochastic variables. Sultana et al. (2015) evaluated the impact of flooding on flexible pavements in Queensland, Australia. They found that the flood-affected pavement shows a 67% reduction in CBR (California Bearing Ratio) and a 50% reduction in SN (structural strength number) of the pavements. Also, the flood-affected pavement sections deteriorated more quickly than the non-flood affected pavement sections. Extreme weather events affect the pavement management system as it directly increases the maintenance cost and, hence, the pavement's life-cycle cost. Therefore, it is crucial to account for these uncertainties in pavements' performance prediction. The pavement's performance is measured using various parameters such as roughness, rutting, and resilience modulus. Previous research showed that roughness and rutting are the most critical parameters representing the structural and functional performance of the pavements. These pavement characteristics are explained below.

2.1.1 Roughness

The distortion in the pavement surface, which leads to the rough or objectionable driving experience, can be called roughness (AASHTO, 1993). The roughness is characterized by the international roughness index (IRI) and generally measured in m/km or in/mi. The IRI is measured by a road profiling system, which includes software for calculating the IRI statistic. The IRI is a slope statistic given by the ratio of the average vehicle suspension's motion by the distance traveled by the profiling vehicle (Sayers, 1998). The IRI of different states and countries is identical and can be compared directly. The IRI 0 m/km of a pavement section represents a perfectly flat and smooth profile, while the IRI greater than 8 m/km represents a nearly impassable pavement profile. Generally, the IRI of a newly constructed pavement falls in the range of 0.40 to 2.0 m/km, while 2.5 to 6.0 m/km for the older pavements. The pavement generally deteriorates with the IRI value increment within the range of 0.10 to 0.25 m/km every year (Sayers, 1998). Chen et al. (2014) had determined the IRI threshold values of the pavement network located in North Carolina and suggested the acceptable and unacceptable IRI ranges. The IRI values in the range of 0.78 to 0.95 m/km is perfect, whereas the IRI in the range of 0.95 to 1.15 m/km is good, and the IRI greater than 2.8 m/km is unacceptable for driving. Figure 2 shows the general IRI values of different pavement classes and the average vehicle speed on these pavements.

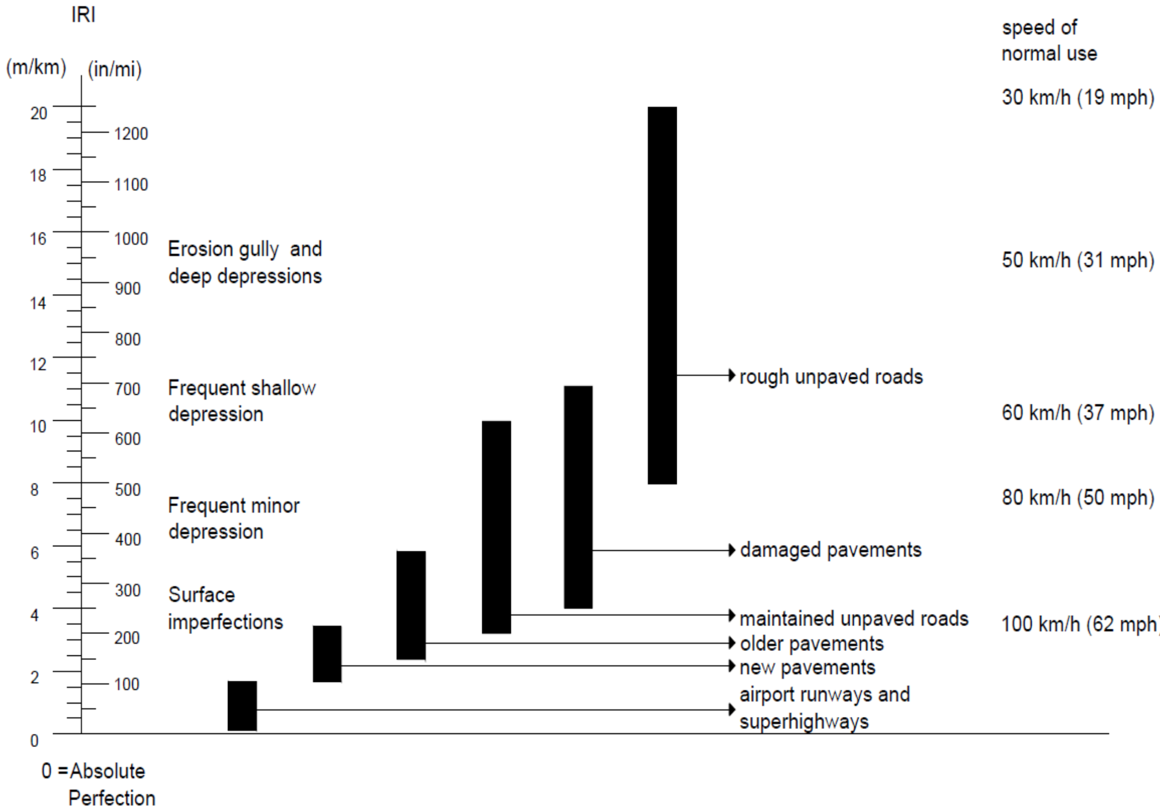


FIGURE 2: IRI ranges of different classes of the road (Adapted from Sayers, 1998)

According to Odoki and Kerali (2000) and Sayers (1998), roughness directly affects the vehicle's operating cost, road accidents, and driving comfort. Cracking, potholes, improper drainage, etc. can cause the pavement's roughness. The IRI is the most used parameter for deriving the pavement's ride quality (Shamsabadi et al., 2014). Therefore, roughness is an important factor directly related to the pavement's performance and serviceability index (Odoki & Kerali, 2000). The roughness of pavement was thoroughly analyzed in this research for developing a pavement deterioration model. The roughness is dependent on external uncertainties such as surrounding climatic conditions, material properties, and traffic loading. Therefore, the analysis of roughness was done using probabilistic methods for deriving the pavement deterioration model. The IRI data of the pavement sections was collected from the LTPP InfoPave database for developing a deterioration model. The details about the LTPP database are explained in section 2.4 of this document.

2.1.2 Rutting

The rutting in any pavement can be caused due to the failure in the pavement's subgrade soil. The accumulation of permanent deformation in pavement layers is called rutting. Rutting significantly affects the pavement's performance, leading to substantial structural failures (Xu & Huang, 2012). Hence, rutting is the critical factor that relates perfectly to pavement deterioration. The unit of measurement of rutting is generally mm or inch. In this research, the intended pavement deterioration model was planned to be developed based on the pavement's roughness and rutting characteristics. However, the LTPP database does not contain sufficient rutting data of the selected pavement sections. Therefore, only the roughness-based deterioration model has been developed in this research.

2.2 Pavement Deterioration Models

The pavement network deteriorates over time due to various factors. Therefore, it is crucial to predict pavement deterioration ahead of time to understand its future performance and develop a comprehensive pavement management system (PMS). The PMS consists of pavement design, financial planning and budgeting, and lifecycle economic analysis. The deterioration model predicts rehabilitation works' required time and simplifies the budget estimation of this works. It also derives the relationship between various pavements' parameters with their serviceability, which helps in pavement designing (George et al., 1989). An accurate pavement deterioration model can be used to precisely measure the pavement performance and maintenance budget associated with it. Two types of pavement deterioration models are found in the literature review: deterministic and probabilistic deterioration models. Deterministic pavement deterioration models provide a specific set of fixed values, whereas the probabilistic models rely on the probability of change in condition.

2.2.1 Deterministic Pavement Deterioration Model

Deterministic pavement deterioration models are developed using pavement's primary response towards the various surface and structural parameters. A prediction model is developed by analyzing pavements' structural and surface responses through regression analysis. The regression analysis is of two types, i.e., empirical regression analysis and mechanistic-empirical analysis. The model generated by empirical regression analysis includes the time-series data of pavement condition with respect to the environmental and traffic loading conditions. The model generated by the mechanistic regression analysis considers the effect of traffic loading, pavement strength, and pavement deflection (George et al., 1989). Different researchers had developed a pavement deterioration model and suggested the methodology for the same. Li et al. (1995) discussed deterministic pavement deterioration models, including regression, mechanistic and mechanistic-empirical models.

Regression analysis requires historical data and independent variables that help in obtaining dependent variables. Regression analysis requires less time and a large amount of data. Madanat et al. (1995) studied the pavement deterioration models' effectiveness developed by statistical regression analysis. Their research recommends that these models do not accurately predict the pavement condition because the reliable data is challenging and time-consuming to obtain. Also, faults present in the data inadvertently impact the accuracy of the prediction. Zhang and Durango-Cohen (2014) have presented a cluster-wise linear regression model (CLR), which shows that different panels or sections of pavement show heterogeneity in deterioration and require developing resource allocation strategies tailored to the specific section of the pavement.

Mechanistic models require many variables such as material properties, geometric design, environmental factors, and loading characteristics to analyze the pavement's stress, strain, and deflection properties. Based on this analysis, the deterioration in the pavement is predicted. Panthi (2009) found that mechanistic models are complex and do not show actual pavement deterioration.

The mechanistic-empirical model gives the final model, which requires less computer power and time. Also, the prediction from this model is better than the mechanistic model. Like the mechanistic model, it requires many variables to analyze stress, strain, and deflection of the pavement. A large number of variables increases the complexity of the model (Panthi, 2009). George et al. (1989) have developed an empirical-mechanistic road deterioration model for predicting pavement serviceability, which helps estimate the budget for operation and maintenance of pavement. The prediction of pavement deterioration done using a deterministic model induces error as it does not consider the uncertainties of environmental and traffic factors.

2.2.2 Probabilistic Pavement Deterioration Model

The probabilistic pavement deterioration model incorporates the uncertain behavior of traffic, environmental factors, and surface characteristics of the pavement in the model.

Li (1997) and Madanat et al. (1995) show the importance of incorporating these uncertainties in pavement deterioration models. Mills (2010) studies two types of probabilistic models, i.e., the Markov chain model and the survival curves model. Both of these probabilistic models are explained later in this document.

2.2.3 Markov Chain (MC) Model

The Markov chain model is a stochastic model that describes the sequence of possible events in which the probability of the next event is dependent on the current event and not on the event before it (Gagniuc, 2017). The Markov chain model can be applied where the variables are stochastic. The Markov chain prediction model is governed by three boundaries, that are:

1. The process must be discrete or continuous in time,
2. The process must have countable outcomes or finite state space, and
3. The process must satisfy the "Markov Property" (Ortiz-García et al., 2006)

The Markov property is satisfied when the next variable's value depends on the current variable's value and not on the past variables.

2.2.3.1 Application of Markov chain theory

The variables such as traffic loading, environmental aspects, and surface characteristics of the pavements are stochastic. Therefore, the Markov theory can be applied to these variables for developing a pavement deterioration model. The Markov Chain process can be applied for deriving pavement deterioration because:

- The deterioration of pavement is a continuous process in time,
- The state-space of the deterioration process is finite in number, and

- The deterioration of pavement is stochastic; hence, it is assumed to hold the Markov property (Kerali & Snaith, 1992)

Various researchers applied the Markov chain theory to the constructed facilities such as pavements and bridges to predict its deterioration. Khan et al. (2014a, 2014b); Madanat et al. (1995); Panthi (2009); Saha et al. (2017) have used the Markov Chain model for predicting pavement deterioration, while Fu and Devaraj (2008); Ranjith et al. (2011) used it for predicting bridge deterioration. The Markov chain model can be used to integrate the pavement deterioration rates with the variables resulting in the change in M&R (Khan et al., 2014b). The Markov chain model focuses on the transition probabilities and the factors responsible for this transition instead of the factors responsible for condition degradation (Karimzadeh & Shoghli, 2020). Two types of Markov chains were found in the literature review: time-independent and time-dependent Markov chains. Both of these models can be utilized for determining pavement deterioration. In the time-independent Markov chain models, the transition probability matrix of the pavement is constant throughout the time, and the pavement is assumed to deteriorate according to this single transition probability matrix; while in the time-dependent Markov chain models, the transition probability matrix of the pavement changes with time and the pavement will deteriorate based on these transition probability matrix (Ortiz-García et al., 2006). The details of the transition probability matrix (TPM) are explained below.

2.2.3.2 Markov chain transition probability matrix

The probability of a pavement section to change its state is termed as transition probability. The pavement sections' transition probability is the probability that the pavement section is currently in condition i at time t and will change to condition j at time

t+1. This transition of pavement sections from one condition to another is combined in a matrix called the transition probability matrix (TPM). The TPM is associated with time-independent and time-dependent Markov chain models that are explained in the previous section.

The classic example of the Markov Chain transition probability matrix is shown below, adapted from Kostuk (2003). Kostuk (2003) shows an example in their research for explaining the Markov Chain process. The example assumes that a bird is sitting on the lily pads in a lake, and there is a finite number of lily pads present in the lake. Lily pads in this example are considered as a state, and since the lily pads are finite in number, this kind of system is described as a finite state system. The location of the bird is captured every five minutes. The probability that the bird will change its position from lily pad i to j is denoted by p_{ij}. Figure 3 shows the simple example of a transition from one state to another.

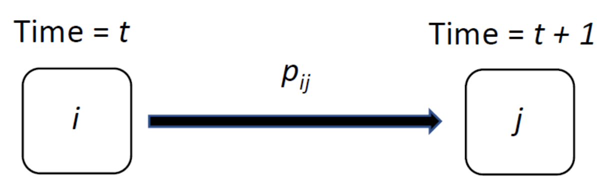


FIGURE 3: The Transition between Two-State (Adapted from Kostuk, 2003)

The likelihood of transition from one state to another state is represented in a matrix, where the rows show the present state, and the columns show the future states. Table 1 shows that the probability of transition from one state to another. For example, the probability of transition from state 1 to state 2 is 0.3.

TABLE 1: Transition Probability Matrix (Adapted from Kostuk, 2003)

From Existing State	To Future State			Total
	1	2	3	
1	0.7	0.3	0	1
2	0	0.6	0.4	1
3	0	0	1	1

The transition in states is combined in a matrix called transition probability matrix, which is denoted by P, and the probability of state change is denoted by p_{ij}, where i denotes the row, and j denotes the column of the matrix. Figure 4 illustrates the transition matrix.

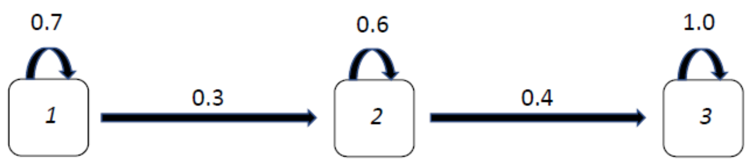


FIGURE 4: Schematic diagram of a three-stage transition model (Adapted from Kostuk, 2003)

Figure 5 represents the complicated situation in which the bird can change its state from *i* to two other states and represent its potential location after one- and two-time epochs. The bird assumed to start with lily pad *i*.

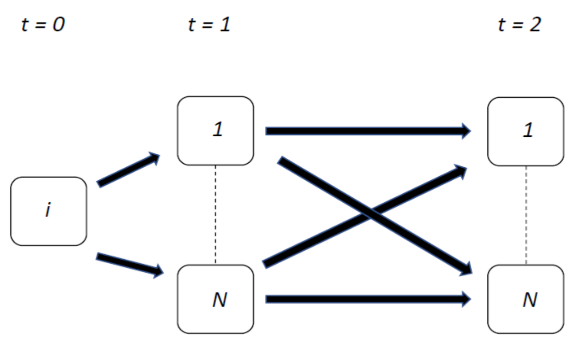


FIGURE 5: The Possible transitions over two epochs (Adapted from Kostuk, 2003)

2.2.3.3 Application of TPM in developing deterioration model

The pavement network's current condition is termed as the initial state and is described in terms of the initial state vector. The initial state vector of the pavement network is given by (Ortiz-García et al., 2006):

$$a_0 = [\alpha_1, \alpha_2, \dots \dots \alpha_n]$$

Initial state vectors assume that all the α_i must be non-negative numbers, and their sum must be equal to one. The pavement deterioration model is generated by utilizing the transition probability matrix (TPM). The transition probability matrix is the probability that the pavement section is currently in condition *i* at time *t* and will change its state to condition *j* at time *t*+1. The TPM is denoted by P and given by (Ortiz-García et al., 2006)

$$P = \begin{bmatrix} p_{11} & p_{12} & \dots. & p_{1n} \\ p_{21} & p_{22} & \dots. & p_{2n} \\ . & . & \dots. & . \\ . & . & \dots. & . \\ p_{n1} & p_{n2} & \dots. & p_{nn} \end{bmatrix}$$

Where, p_{ij} indicates the probability that a road is currently in state *i* and will be in state *j* next year. Like the initial state vector, all the TPM numbers must be non-negative, and the sum of each row must be equal to one. The probability distribution of the states at a future time, say $t = 1$ and at time *t,* may be calculated from the TPM generated and the initial state vector and is shown in equations 1 and 2 (Ortiz-García et al., 2006).

$$a_1 = a_0 \, P^1 \tag{1}$$

$$a_t = a_0 \, P^t \tag{2}$$

Where, a_1 = probability distribution at time t = 1, a_t = probability distribution at time t, a_0 = initial state vector at time t = 0, and P^t = TPM raised to the power of t. In equations (1) and (2), it is assumed that the transition probability matrix (P) of the pavement is constant throughout the time, and the pavement is assumed to deteriorate according to this single transition probability matrix P throughout its lifespan. This equation is used for performing time-independent Markov chain analysis. In this research, a time-independent Markov chain analysis was performed for developing a pavement deterioration model.

The TPM used for representing pavement deterioration were generated based on three assumptions:

1. The condition of the pavement sections cannot be improved without receiving any maintenance treatment, i.e., $p_{ij} = 0$ for $i > j$,
2. The pavement sections which reached their worst condition cannot deteriorate further, i.e., $p_{nn} = 1$, and
3. The pavement section cannot deteriorate by more than one state in a duty cycle.

Based on this assumption, the ideal TPM of the pavement section is denoted by:

$$P = \begin{bmatrix} p_{11} & p_{12} & 0 & \dots. & 0 \\ 0 & p_{22} & p_{23} & \dots. & 0 \\ 0 & 0 & p_{33} & \dots. & 0 \\ . & . & . & . & . \\ . & . & . & . & . \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

The Markov chain prediction model specified above is used by various pavement management systems for predicting pavement deterioration. Some of these are NOS (Kulkarni, 1984), MicroPAVER (Butt et al., 1994), NETCOM (Kerali & Snaith, 1992), and HIPS (Thompson et al., 1987).

Several research studies were conducted to derive TPMs using various mathematical methods in the past. Such as, the simplest proportion method (Wang et al., 1994), the expected value method (Jiang et al., 1988), minimum error method, percentage transition method, ordered probit model, Bayesian technique, and conversion from the deterministic models. All these methods are explained and evaluated later in this section.

2.2.4 Methods for deriving TPM

1. Minimum-error method

The Minimum-error method uses historical data, regression analysis, and historical distribution data for deriving TPMs. Ortiz-García et al. (2006) had used the minimum-error method for deriving time-dependent TPMs. The typical equation of the minimum error method is given by Ranjith et al. (2011).

Objective function $Z = \sum_t C(t) - E(t)$

Subject to $0 \le p_{ij} \le 1$ i,j = 1,2…….n

$\sum_j pij = 1$ i= 1,2.....n

Where,

C(t)= system condition rating at time t based on regression

E(t)= expected rating at time t

According to Madanat et al. (1995), the minimum error method does not consider the effect of vehicle loading, environmental factors, material properties, loading scenarios, and underlying continuous deterioration.

2. Percentage transition method

The percentage transition method is used to derive the change in road condition state with respect to the previous state. Pierce (2003) has used the percentage transition method for deriving TPMs using five years of historical IRI data of rigid pavements. These data are used to develop TPMs and then used in Monte Carlo simulation to derive the road deterioration model. For predicting the bridge deterioration model, Ranjith et al. (2011) had statistically compared the minimum-error and percentage transition method and found that the minimum-error method sometimes gives better results than the percentage transition method. Ranjith et al. (2011) compared this method to predict bridge deterioration only. The percentage transition method addresses different explanatory variables used to develop a pavement deterioration model (Khan et al., 2014b). The transition probability of each pavement section can be calculated using this equation:

$$p_{ij} = \frac{N_{ij}}{N_i} \tag{3}$$

Where, p_{ij} = transition probability from state i to j, N_{ij} = number of sections transition from state i at time t to state j at time t+1. When reliable data is insufficient, the panel of expert engineers' engineering judgment is used to prepare the transition probability matrix. All the transition probabilities are compiled in a matrix called the transition probability matrix, which is given by

$$P = \begin{bmatrix} p_{11} & p_{12} & 0 & \dots. & 0 \\ 0 & p_{22} & p_{23} & \dots. & 0 \\ 0 & 0 & p_{33} & \dots. & 0 \\ . & . & . & . & . \\ . & . & . & . & . \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

3. Probit model

The Probit model is a type of model where the dependent variable can take only two values, and the model is estimated using the maximum likelihood procedure (Khan et al., 2014b). Madanat et al. (1995) used the ordered Probit model for developing time dependent TPMs, and the results show that the proposed method considered explanatory variables and the hidden nature of the pavement performance. They assumed that ordered states were independent and showed similar distribution. However, Fu and Devaraj (2008) study show that the ordered Probit model requires a large amount of data to generate suitable TPMs. Li (2005) derived a method for converting the deterministic model into a probabilistic model, which is easy compared to developing a new model for deriving TPMs. However, this method is not commonly used.

4. Bayesian Technique

The Bayesian technique is used to derive heterogeneity of different individual parameters of the pavement, such as roughness and rutting, and it is used to obtain realistic parameter distributions through data and knowledge (Hong & Prozzi, 2006). The study of Li (1997) shows that the Bayesian technique was feasible for validating the developed pavement deterioration model, rather than developing a new model. However, Panthi (2009) stated that this technique does not consider the pavement's mechanistic behavior and depends heavily on data; therefore, improper data can lead to erroneous models.

5. Conversion from the deterministic model

Li (1997), Li et al. (1995), has used the deterministic equation and converted it using Monte-Carlo simulation to derive TPMs. The deterministic equation is dependent on the pavement design equation, which considers the effect of traffic growth rate, subgrade deflection, and pavement thickness in deriving the TPMs. Historical data was used to derive the TPMs. The conversion from the deterministic model to the probabilistic model is straightforward (Li et al., 1995). However, this method used the existing deterioration model rather than developing and a new one. The researchers rarely use this method.

The qualitative analysis was conducted on all these methods for determining the most feasible method for developing a transition probability matrix. The qualitative evaluation was done based on the results derived from previous studies and is shown in table 2. Based on this evaluation, the percentage transition method was found feasible for deriving a transition probability matrix because it uses real data and considers the effects of the underlying variable affecting the deterioration process.

TABLE 2: Different methods for deriving transition probability matrix

Method Used	**Functions**	**Limitations**	**Sources**
Minimum-Error method	Based on historical data and engineering experience	Does not contain effects of explanatory variables on underlying deterioration	Ortiz-García et al. (2006), Madanat et al. (1995), Khan et al. (2014b)
Percentage Transition Method	A TPM is derived from the probability of transition from one state to other	Uses real data to derive TPMs and considers the effect of explanatory variables	Ranjith et al. (2011), Pierce (2003), Khan et al. (2014b),
Ordered Probit Method	Considers explanatory variable in pavement deterioration: the ordered state is independent of the previous state	It does not provide a transition probability matrix and requires a large amount of data	Fu and Devaraj (2008), Li (2005), Madanat et al. (1995), Khan et al. (2014b)
Bayesian Technique	Address heterogeneity of individual parameter can be used for model calibration	Heavily depends on the data and improper data leads to an erroneous model	Li (1997), Panthi (2009), Hong and Prozzi (2006), Khan et al. (2014b)
Conversion from the deterministic model	Easier and quick process	The method is similar to the statistical method; no new deterioration model is developed	Li (1997), Li et al. (1995), Khan et al. (2014b)

2.2.5 Survival Curves

The insurance companies generally use survival curves to derive the probable life of the product's units and decide the premium values of these products and services. The survival curves show the number of units of the given group surviving at a particular age (Winfrey, 1935). Wang et al. (1994) used long-term pavement performance (LTPP) program data for deriving survival curves patterns of fatigue cracking of flexible pavement. Panthi (2009) found that survival curves are easy to develop and give the only probability of pavement failure corresponding to the age, but this probability induces considerable error if the group of units used is small. The schematic trend of a survival curve is shown in figure 6.

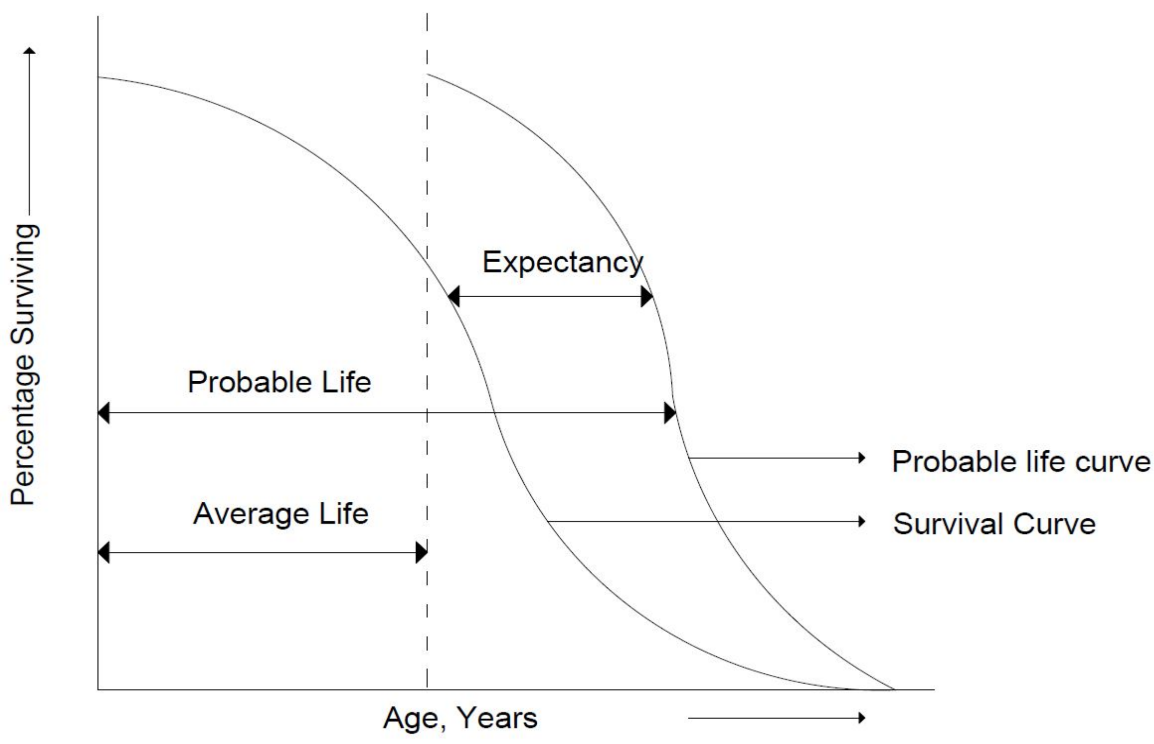


FIGURE 6: Survival Curve (Adapted from Panthi, 2009)

2.3 Clustering Technique

Cluster analysis or clustering is a statistical algorithm used to group a set of data objects into clusters so that the properties of data objects in a cluster are similar to each other. The similarity between the data objects is calculated based on the Euclidean distance ‘*e*’ between the objects' attributes. This algorithm is an unsupervised learning algorithm, where no prior assumptions are made for the likely relationships within the data objects. The clustering algorithm is classified into two types, i.e., hierarchical and non-hierarchical clustering. The classification of the clustering algorithm is shown in figure 7 and explained in this section.

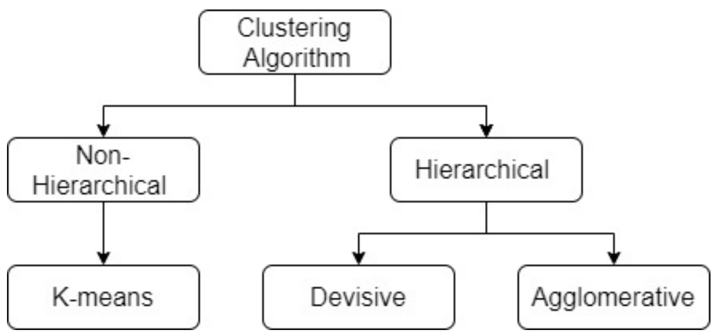


FIGURE 7: Classification of clustering algorithm

2.3.1 Hierarchical Clustering

Hierarchical algorithms form clusters in two methods, i.e., agglomerative methods and divisive methods. In agglomerative methods, the clusters are merged in stages, while in the divisive method, the clusters are divided (Kim et al., 2004). Generally, a dendrogram is used to represent a hierarchical clustering algorithm. This method assumes that the number of clusters is equal to the number of the data objects, and then the dissimilarity matrix is developed for the computation of clusters (Kumar & Swamy, 2015). In this process, the grouping of data points with maximum similarity is done and developing a dissimilarity matrix to obtain the relationship between the newly formed cluster and the remaining entity (Gong & Richman, 1995).

2.3.2 Non-hierarchical Clustering

In the non-hierarchical clustering algorithm, a dataset is divided into groups based on the dataset's similar properties so that these groups are non-overlapping and have no hierarchical relationship between them (Barnard, 1995). The *K*-means clustering algorithm is a widely used non-hierarchical clustering method. This method is explained below.

2.3.2.1 *K*-means Clustering

In *K*-means clustering, the '*K*' is the centroid of the cluster and represents the number of groups in a dataset. The fundamental calculation in this algorithm includes selecting '*K*' points or centroids and then assigning the datasets closest to the centroids and repeating this iteration until there is no change in the centroids (Karypis et al., 2000). The *K*-means clustering algorithm aims to minimize the squared Euclidian distance between the data points and the centroid on the cluster. The advantage of the *K*-means algorithm over the agglomerative algorithm is that it can be performed quickly. This algorithm is the most extensively used clustering algorithm. A scatterplot can be plotted to visualize different clusters.

2.3.3 Application of Clustering analysis

Clustering analysis has been used in diverse work fields such as marketing, land use, insurance, city planning, earth-quake studies, and many more fields. In the transportation engineering field, clustering is primarily applied to group the variable used in Mechanistic-Empirical Pavement Design (MEPD) (Kumar & Swamy, 2015). Wang et al. (2011) collected the weigh-in-motion (WIM) data recorded in the state of Arkansas, USA. They used a clustering algorithm to identify the truck loading patterns and estimate the full axle-load spectrum. Lu and Harvey (2006) used the clustering method for classifying truck composition, volume, speed, and axle load spectrum in California to improve MEPD. Papagiannakis et al. (2006) applied the hierarchical clustering method on the WIM data collected from the long-term pavement performance (LTPP InfoPave) database. They identify the groups of sites with decreasing similarities based on the vehicle class and axle loading. Yan et al. (2011) used the *K*-means clustering algorithm for modeling the flow of traffic at the port of Tianjin. Sunitha et al. (2012) developed two types of deterioration models: by using the *K*-means clustering algorithm and without using any clustering algorithm. They compared the results derived from both the models and found that pavement sections' clustering is preferred for efficient pavement performance prediction. Developing a deterioration prediction model requires data-extraction of the historical condition and maintenance activities of pavement sections. Inadequate and insufficient data extraction may lead to generate inaccurate prediction models. Therefore, the family of assets was developed, and the deterioration of these families was then investigated. Karimzadeh et al. (2020a) used the *K*-means and agglomerative hierarchical algorithms to extract similarities over a broad set of road segments. Their study provides more comprehensive insight for forming family groups, which could result in improving deterioration models.

All previous research suggests that clustering is an effective mathematical tool that can be used for efficient data analysis when less data is available. However, no research was found which utilizes the clustering process for developing a probabilistic pavement deterioration model.

The clustering process is an unsupervised learning algorithm; therefore, the cluster's validation is required. The formed cluster's validation can be done by creating different data visuals, such as 2D and 3D scatterplot. However, creating data visuals is not possible when many variables are available. Many research has been going on to determine a method to derive the optimum number of data clusters, but no reliable method is available. Generally, for determining the optimal number of clusters for a dataset, the elbow method is used. However, selecting the number of clusters is a subjective topic and can be done based on the available data, analysis's requirement, and judgment (Karimzadeh et al., 2020b).

In this research, the clustering technique was used to appropriately group pavement sections into different groups and develop each cluster's deterioration models. The sections within the cluster are homogeneous with each other and heterogeneous between other clusters. The data for developing the model was collected from the LTPP database. This data was unlabeled, and the grouping of data was needed to be done. Therefore, the *K*-means clustering algorithm was suitable for this kind of dataset and used in this research. The data was divided into three clusters using the *K*-means algorithm. The number of clusters in this research was determined based on the judgment and evenly distributing sections in each cluster.

2.4 LTPP Database

The long-term pavement performance (LTPP) InfoPave program was developed by the Transportation Research Board (TRB) in 1987 as a part of the Strategic Highway Research Program (SHRP). This program's overall objective is to analyze and document long-term pavement performance under various loading and environmental conditions throughout its lifespan. Under this program, pavement test sections were constructed on in-service roadways. Each test section has assigned a unique section number which identifies the pavement type, location, material, and thickness properties. The test section's length is 152-meter, extended by a maintenance control zone of 152-meter and 76-meter on either side of the test section (Elkins et al., 2003). The illustration of a typical LTTP test section is shown in Figure 8. The effect of traffic loading, environmental conditions, material properties, construction quality, and maintenance on test sections is documented and uploaded on the LTPP InfoPave website. The desired pavement information can be filtered and extracted from this database. In the LTPP InfoPave database, there are 2581 pavement sections across the United States and Canada. Some of these sections are routinely monitored, while others are not. The routinely monitored sections are those whose data collection and updating process is presently ongoing; therefore, these sections are given active status while others are given inactive status. The LTPP InfoPave database was used to collect data for this research, and the sections in active status are selected for the analysis.

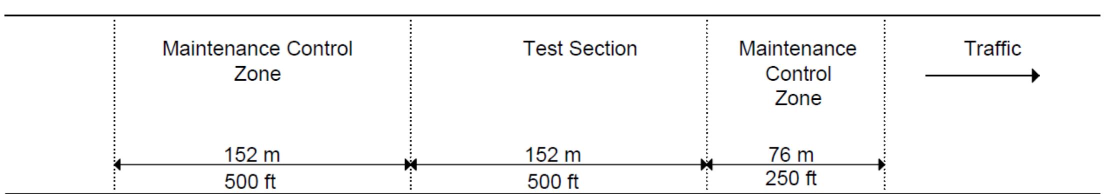


FIGURE 8: Typical layout of an LTPP test section (Adapted from Elkins et al. 2003)

The database contains different types of information on each monitored pavement section. This information is grouped into five distinct primary categories: general information, structural information, climatic information, traffic information, and performance information. The general information category contains information about pavement age, experiment type, study, group, monitoring status, section type, treatment type, section location, maintenance and rehabilitation details, and roadway functional class. The structural information category contains material details of the pavement sections' surface, base, and subgrade layers. This category is used to differentiate between asphalt concrete (AC) and Portland cement concrete (PCC) pavement sections. The climatic category contains information about the climatic region, annual freezing index, annual precipitation, and annual temperature sections. The traffic category contains traffic loading information in terms of average annual daily traffic (AADT) and average annual daily truck traffic (AADTT) of the sections. The performance category contains information about pavement deflection, fatigue cracking, faulting, longitudinal cracking, longitudinal profile (IRI), transverse cracking, and transverse profile of all the sections. These five information categories are specified on the left-hand side of the website and can be used as the filtering tool for extracting the desired pavement data.

The data collection begins by clicking on the Data tab on the LTPP InfoPave website, selecting the pavement sections' desired attributes using the available filtering tools, and then downloading the data in the desired format. In this research, the LTPP InfoPave database was used for extracting the historical data from 2015 to 2019 of the pavement sections. The extracted data contains historical information of pavement attributes such as traffic loading, temperature, precipitation, and roughness. These attributes were used in the analysis for developing the probabilistic pavement deterioration model. The analysis of these parameters is explained in chapter 3 of this document.

2.5 Flooding Hazards

A flood is termed when the streamflow level of water becomes relatively high and inundates the banks of the stream (USGS, 2019). An event or physical condition that can cause fatalities, accidents, belongings damage, infrastructure harm, etc., is termed a hazard (Drane et al., 2020). The floods can cause loss and damage to lives and property by the inundation of water. Therefore, the hazards associated with floods are termed as flood hazards. The floods comprise three elements, i.e., severity, which includes the magnitude, duration, extent of flooding; occurrence probability; and the start speed of flooding (FEMA, 2019).

2.5.1 Impact of Flood on Pavement Performance

The emission of greenhouse gases is increasing exponentially every decade, which causes climate change, and these climate changes are resulting in extreme weather events such as frequent heavy rainfall, floods, increasing temperature, and other weather extremes (Parry et al., 2007). Extreme weather events adversely affect the performance of pavement by reducing its structural strength of base and sub-base layers and increasing surface failure characteristics of the pavement network (Helali et al., 2008; Khan et al., 2014a; Mallick et al., 2017; Sultana et al., 2015; Zhang et al., 2008). The analysis for understanding the impact of flooding on pavement performance is very important for efficiently managing the road assets in the long run. The flood damage affects the pavement performance in various patterns, termed as the delayed effect, jump effect, jump and failure effect, and direct failure effect (Lu et al., 2017). These flooding effects are caused due to various flood loads such as flood depth, velocity, duration, debris, and contaminants, which cause various damages specified in Table 3 (Lu et al., 2018).

TABLE 3: Types of load and its effect on the pavement (Adapted from D. Lu et al., 2018)

Load type	Pavement damage description
Flood depth	Absorption of water and moisture
Flood duration	Absorption of water and moisture
Flood velocity	Force of water causes the removal of materials
Flood debris	Debris carried water reduces surface characteristics
Flood contaminants	Flood contaminants absorption or adhesion

The pavement sections are designed as multi-layer composite systems that transfer the traffic load uniformly from the exposed surface layer to the bottom-most subgrade layer. Pavements' serviceability is dependent on load distributing characteristics of traffic loads. However, flooding and frequent heavy rainfall adversely affect these characteristics. The research done by Lu et al. (2018) shows that pavement damage ratio increases as the increase in the cycle of flooding and frequent hefty rainfall increases.

2.5.2 Flood Risk

The flood risk assessment requires sound scientific and technical analysis to minimize the flood loss by preparing an effective flood responding strategy ahead of time. The occurrence of any magnitude of a flood is expressed in a 1-percent annual exceeding probability (AEP). The 1-percent AEP flood has the 1% probability of being equal to or exceeding the flood level or peak in any given year and has an average recurrence of 100 years and is generally termed a 100-year flood (USGS, 2019). Statistical procedures are used to derive the flooding recurrence probability. The height of the water and the quantity of streamflow is examined at the stream gauge. The stream gauge is used to derive the flooding recurrence probability by analyzing the annual peak flows measured at the water body (USGS, 2019). Generally, it is found that a catastrophic flooding event has less recurrence probability than a smaller flooding event; similarly, the recurrence interval of catastrophic foods is less than smaller floods (FEMA, 2019)..

The risk is termed as an undesirable event that will cause loss and damages to life and property. The risk of any event can be expressed in terms of the probability of that event's occurrence. The probability of an event's occurrence represents the event's likelihood, ranging from 0 to 1. The probability of flooding is derived using the following techniques (a) statistical analysis of stream-flow records, (b) regional methods, (c) transfer methods, (d) empirical equations, and (e) watershed modeling (FEMA, 2019).

2.5.3 Flooding Events in North Carolina

North Carolina, located on the Atlantic Seaboard, is regularly affected by tropical thunderstorms, causing heavy rainfall and floods. The Okeechobee hurricane (1928), Hurricane Floyd (1999), and Hurricane Florence (2018) are some examples of significant flooding events that occurred in the state (NWS, 2019). The Okeechobee hurricane in 1928 was decimated from Puerto Rico, which causes 4 to 9 inches of rainfall in eastern North Carolina. In 1999, the hurricane Floyd was decimated near the Cape Verde islands, which caused 15 to 20 inches of rainfall in the eastern region of North Carolina. Hurricane Florence in 2018 was also decimated from the Cape Verde islands, which caused devastating rainfall of 20 to 30 inches in the eastern region of North Carolina (NWS, 2019). The weather history of the state of North Carolina shows that various tropical floods occurred in the state over time, and the eastern part of North Carolina is the most affected geographical region of the state. Therefore, it is essential to develop a framework that shows the effect of flooding on the pavement surface.

2.5.4 Incorporation flooding in pavement deterioration model

This research's primary objective is to propose a framework that shows the flood's impact on the composite pavement sections and reflects it in a deterioration model. Khan et al. (2014a) had proposed the methodology for incorporating the effect of flooding in the deterioration model. However, the deterioration model developed by them used an unsophisticated clustering method. This research collected the pavement data from the LTPP InfoPave database, which does not contain the flood-affected pavement sections' information. Moreover, no reliable data was available, showing the historical relationship between flood occurrence and its impact on the pavement’s performance. Therefore, a hypothetical flooding event is assumed to occur between 2020 and 2021, and its impact on pavements’ roughness was predicted. The flood's impact on pavements’ roughness was predicted based on the accumulation of flooded water on the pavement surface. For this, the flood recurrence interval was studied, and annual flooding probability was determined based on it.

For deriving the deterioration model at various probabilities of flooding, the Monte Carlo simulation was conducted. The actual non-flood TPM and a hypothetical flood TPM are generated and used in the Monte Carlo simulation. The effect of a flood can be seen by the increment in the pavement surface's IRI value. The increment in the IRI value depends on the probability of flood occurrence, i.e., the higher the probability of flood occurrence, the higher the IRI's increment. The framework for incorporating the flooding effect in the deterioration model is explained in detail in section 3.5 of this document.

2.6 Literature Review Summary

A comprehensive review of the literature was conducted for this thesis work. The change in pavement condition with respect to time is called pavement deterioration. Pavement deterioration is dependent on various parameters such as traffic, environmental, geographical, and material. These parameters affect the pavements' performance and ride quality. The roughness (IRI) is the parameter used for calculating the ride quality of the pavement. The higher the IRI value, the lesser the ride quality. The IRI is an essential factor directly related to the pavement's performance and serviceability index; therefore, it is used in this research for developing a pavement deterioration model. The IRI is stochastic in nature; therefore, probabilistic methods are required for predicting it accurately.

Various research was conducted in the past for deriving the pavement deterioration models. Generally, two types of pavement deterioration models were found in the literature review, i.e., deterministic pavement deterioration models and probabilistic pavement deterioration models.

Deterministic pavement deterioration models are generally based on regression analysis, which is relatively simple and easy to use. Most of the pavement management system uses deterministic pavement deterioration models. However, these models have few disadvantages, such as it does not account for the IRI's uncertain nature, caused due to variable traffic loading and environmental conditions. Also, it requires a large amount of data sets for accurate deterioration prediction.

Probabilistic pavement prediction models use the Markov Chain theory for developing the deterioration model. The advantage of using the probabilistic pavement deterioration model is that it accounts for the IRI's uncertain behavior caused due to variable traffic loading and environmental factors. Moreover, the data required for developing a deterioration model is less compared to the deterministic models. In this research, the pavement deterioration model was developed using the probabilistic method for accounting for the IRI's uncertain nature.

The Markov Chain model is probabilistic, which accounts for the stochastic nature of the pavement's roughness. The Markov process consists of TPMs that predict the change in pavement conditions from one state to another based on the current state and not the state before it. The Markov chain time-independent and time-dependent TPMs can be used for deriving the pavement deterioration model. The TPMs can be derived using five methods specified earlier in section 2.2.4. The qualitative analysis was conducted on these methods by thoroughly analyzing previous literature. The salient features and evaluation of all these methods are mentioned in table 2. Based on this evaluation, the percentage transition method was the most feasible method for deriving TPMs as it addresses different explanatory variables used in developing a pavement deterioration model.

Cluster analysis or clustering is a statistical algorithm used to group a set of data objects into clusters so that the properties of data objects in a cluster are similar. In this research, *K*-means clustering was used to group the pavement sections based on their traffic loading, temperature, and precipitation data. In many research, the grouping of pavement sections was done in an unsophisticated manner. The selection of clusters' numbers was made based on the judgment for evenly distributing sections in each cluster.

In this research, the LTPP InfoPave database was used for extracting the historical data from 2015 to 2019 of the pavement sections. The extracted data contains historical information of pavement attributes such as traffic loading, temperature, precipitation, and roughness. After collecting the data, the *K*-means clustering algorithm was used to group this data into three clusters.

In this century, climate change results in extreme weather events such as excessive temperature, drought, floods, excessive snow, and various other unanticipated natural events. The pavement infrastructure system throughout the world is experiencing performance loss due to extreme weather events. This research proposes a framework for evaluating the change in pavements’ roughness due to the flooding event's occurrence and representing it through a deterioration model. Due unavailability of the flood-affected pavement data, a hypothetical flooding event is assumed to occur between 2020 and 2021, and its impact on pavements’ roughness is predicted.. The actual non-flood TPM and a hypothetical flood TPM are generated and used in the Monte Carlo simulation to predict pavements’ deterioration at the various probability of flood.

Most research does not use clustering techniques and developed family-based prediction models. However, clustering is an effective mathematical tool that can be used for efficient data analysis when less data is available. Moreover, very few research was conducted on developing a probabilistic pavement deterioration model for composite pavements. The LTPP database was not used to derive a roughness-based probabilistic pavement deterioration model by incorporating flooding effects. Therefore, a study was required that proposes a deterioration model by addressing all these issues. Also, the states' transportation department can utilize the deterioration model to improve its pavement management system. The methodology utilized in this research is thoroughly explained in chapter 3 of this document.

CHAPTER 3: METHODOLOGY

In this research, a cluster-based probabilistic pavement deterioration model was developed. This model is a roughness-based deterioration model as roughness is the most significant factor contributing to pavements' serviceability. The rate of deterioration in pavement accelerates when extreme weather events such as floods occur. Therefore, this research also proposes a framework to incorporate the effect of flooding in the deterioration model. This framework will help predict the pavement's roughness based on the probability of flood occurrence. In this section, the methodology utilized to develop a cluster-based deterioration model and the incorporation of flooding in the deterioration model is explained in detail. The overview of the methodology is shown in figure 9.

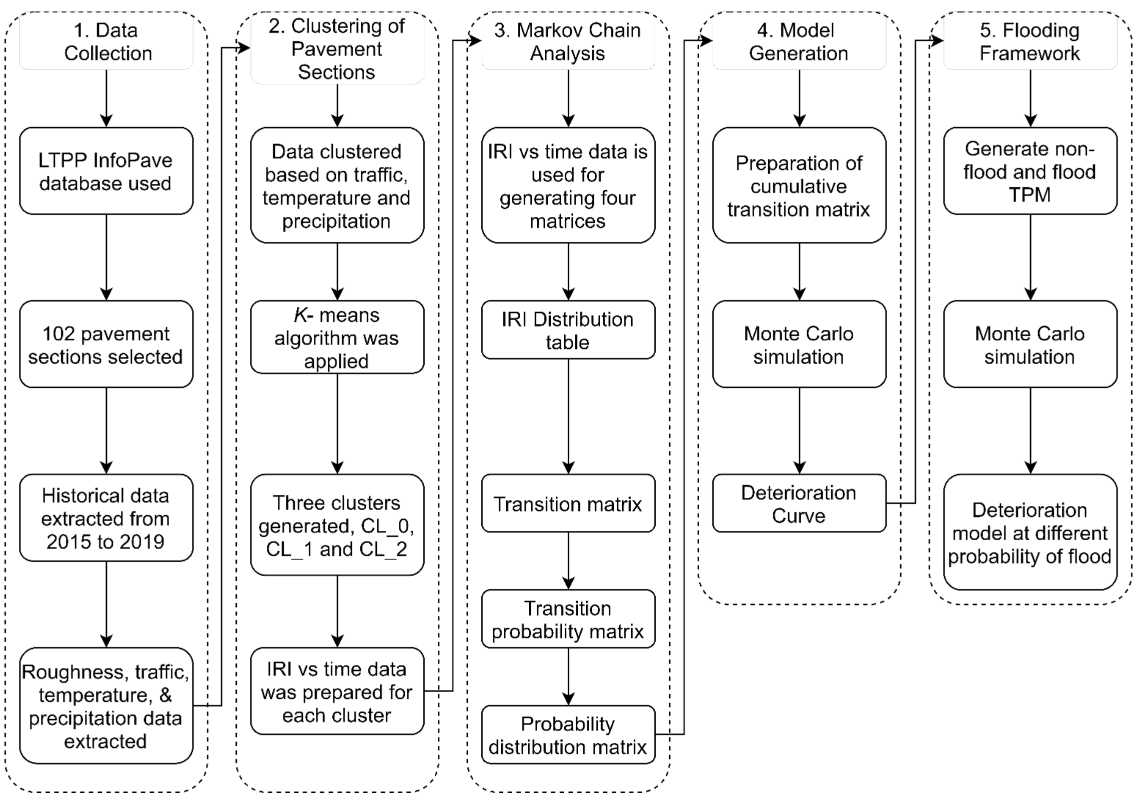


FIGURE 9: Overall methodology of the analysis

3.1 Data collection

The development of a distress-based deterioration model requires the historical distress data of the pavement. Thus, the entire pavement is divided into smaller sections; then, the historical data is extracted for these sections. These sections' distress data are analyzed and then combined for deriving a distress-based deterioration model for the entire pavement. A similar kind of methodology was used for deriving the deterioration model in this research.

The deterioration model was developed using the roughness (IRI) characteristics of the pavement in this research. The historical data such as IRI, traffic loading, temperature, and precipitation of 102 pavement sections from 2015 to 2019 was extracted from the LTPP database for the analysis purpose. The clustering technique is a robust tool used to determine patterns and structures in labeled and unlabeled datasets (Sandra & Sarkar, 2015). The data extracted from the LTPP test sections were not a part of the same roadway and belonged to different roadways. These test sections are also situated in different geographical and environmental locations, and section-wise, IRI data of one entire pavement was not available for analysis. Moreover, a single deterioration model developed for a variety of pavements underestimates or overestimates the pavement condition (Sunitha et al., 2012). Due to these issues, the clustering algorithm was applied to determine patterns within these pavement sections.

The clustering algorithm was used to group pavement sections into different clusters based on their traffic loading, temperature, and precipitation characteristics. The sections within a cluster are homogeneous while heterogeneous with the sections from different clusters. Therefore, it was assumed that the combination of the pavement sections in each cluster represents one entire pavement. The deterioration model of each cluster predicts the future condition of one entire pavement.

3.1.1 Pavement Section Selection & Data Preparation

Five years of historical data from 2015 to 2019 were obtained to develop a probabilistic pavement deterioration model that incorporates flooding. A case study was then conducted on 102 pavement sections located in the United States' eastern region. The selection of pavement sections for the analysis was made based on three criteria:

- The pavement sections should be of asphalt concrete or composite pavement,
- Sections must be in an active monitoring condition, and
- After 2015 no maintenance was conducted on these sections.

The composite pavement sections in this research are consist of an asphalt concrete layer over a Portland cement concrete layer. The typical cross-section of composite pavement is shown in figure 10.

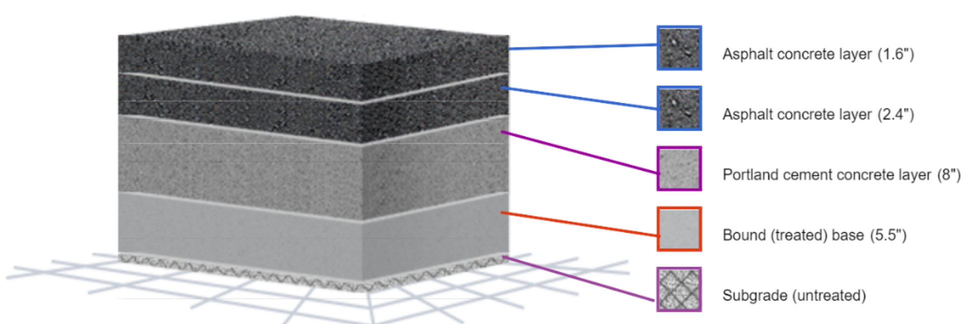


FIGURE 10 Typical cross-section of a composite test section (Adapted from LTPP InfoPave database)

After applying these filters, historical data from 2015 to 2019 was collected for 102 pavement sections from the LTPP database. The historical data of these sections consists of parameters that are: traffic (AADTT), temperature (°F), precipitation (in), and roughness (m/km). The data extraction process was conducted by using the 'data' tab on the LTPP InfoPave website. Filtering tools available on the LTPP website was used for selecting the data of interest. The data was extracted in the form of a downloadable Microsoft Excel file from the database. The extracted Excel sheet contains a basic section overview, history of the selected section's structure, climate, traffic, and profile information. The data extracted from the LTPP database requires manual formatting; therefore, every section's parameters were combined in one excel sheet for the analysis. The state-wise section list is shown in table 4. The highlighted states in figure 11 represent the geographical location of selected pavement sections.

TABLE 4: State-wise Section List

Row	State	No. of Sections
1	Alabama	9
2	Florida	15
3	Georgia	6
4	Illinois	1
5	Indiana	8
6	Kentucky	1
7	Maryland	1
8	Mississippi	7
9	Missouri	3
10	North Carolina	17
11	Ohio	8
12	Pennsylvania	5
13	South Carolina	1
14	Tennessee	3
15	Virginia	16
16	West Virginia	1
	Total	102

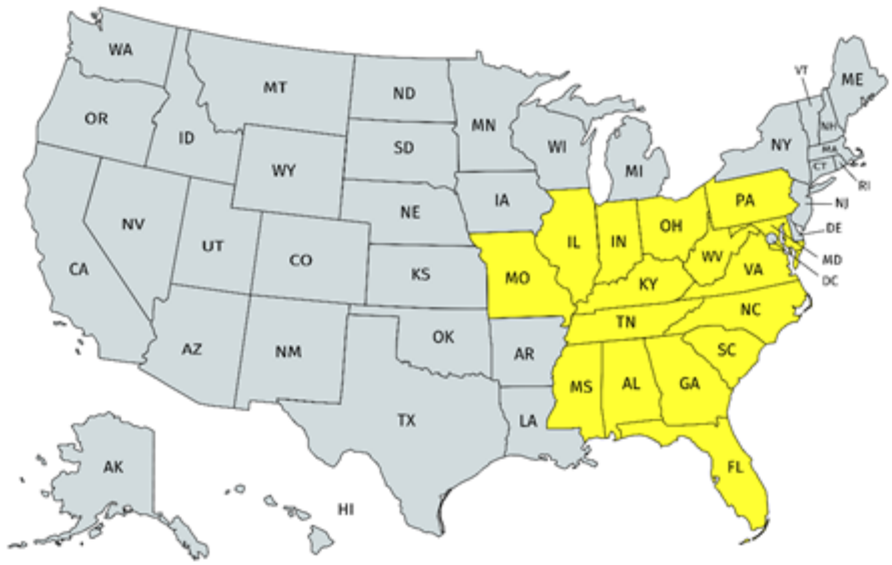


FIGURE 11: Geographical location of selected sections

3.2 Pavement Section Grouping

The selected sections are combined into different groups based on the collected historical data from the LTPP InfoPave database. These different groups are called clusters, and the process of dividing pavement sections into different clusters is known as clustering. The clustering process is an unsupervised algorithm in which there is no prior knowledge about the relationship between the observations. Clustering's aim in this research was to combine the pavement sections into various clusters based on their traffic loading, temperature, and precipitation characteristics.

Various clustering algorithms were found in the literature. Overall, the *K*-means clustering algorithm is easy to apply, accurate, and effective in handling a large amount of data (Alashwal et al., 2019; Perera et al., 1998; Sandra & Sarkar, 2015; Sunitha et al., 2012). Also, the data collected from the LTPP database was unlabeled and not in defined categories or groups. Therefore, the *K*-means clustering algorithm was found suitable for this kind of dataset and further used in this research for analysis. This algorithm's main objective was to find and assign groups to the dataset.

The variable *K* represents the number of groups. Selecting the optimal number of clusters is a subjective topic and can be done based on the available data, analysis's requirement, and judgment (Karimzadeh et al., 2020b). In this research, the optimal number of clusters was derived based on evenly distributing pavement sections into each cluster. It was assumed that each cluster represents one entire pavement. Therefore, three separate deterioration models were developed for three clusters (i.e., Cluster-based deterioration models).

Two datasets were created containing traffic, precipitation, temperature, and IRI information of all the 102 pavement sections. The first dataset contains the parameters' actual values, while the other dataset contains the scaled value between 0 to 1 of these parameters. The dataset, which contains scaled values of the parameter, was used for clustering. Python was used for applying the *K*-means clustering algorithm for the grouping of sections into different clusters. The dataset containing the scaled value of the parameters was imported into Python. The clusters were then generated using the *K*-means clustering algorithm based on traffic, temperature, and precipitation data from 2015 to 2019. In total, the clustering algorithm was applied on fifteen pavement attributes (3 characteristics x 5 years). The results derived from the clustering analysis were that the pavement sections are grouped into three different clusters named CL_0, CL_1 & CL_2. These clusters were assumed to represent three different pavements as a whole, and the deterioration model of these clusters was developed separately. The scatterplot of sections representing its cluster identity was plotted using their geographical location and is shown in figure 12.

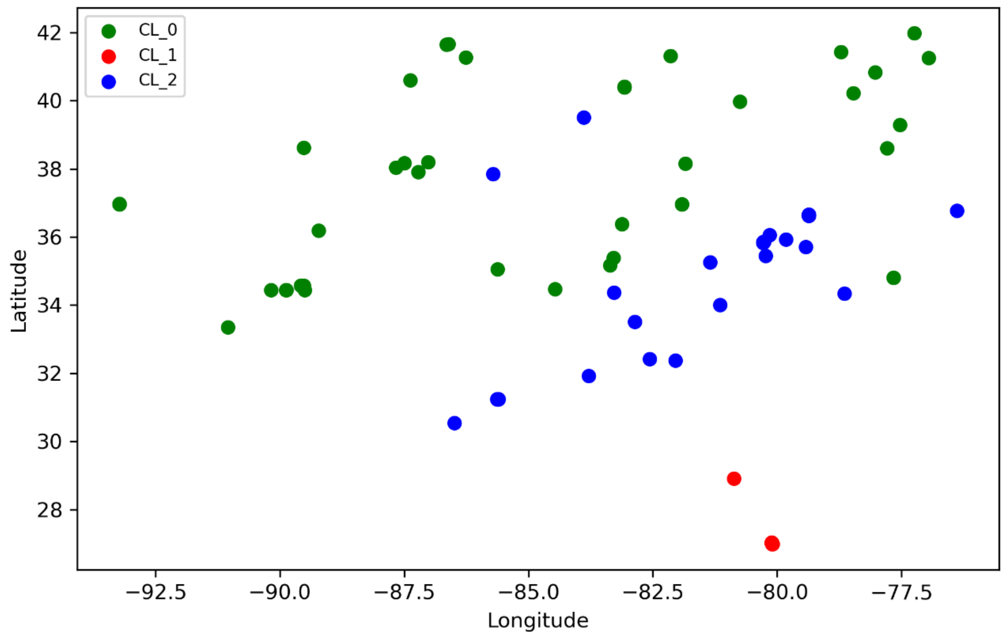


FIGURE 12: Cluster identity and geographical location of sections

3.2.1 Clustered data visualization

In the cluster CL_0, CL_1, and CL_2, there are 44, 14, and 44 pavement sections, respectively. Each cluster comprises of sections from various states of the United States. Based on each cluster's traffic, temperature, and precipitation characteristics, a cluster summary table was prepared and shown in table 5. This summary table was then further used to define each cluster's properties, and a name code was assigned to each cluster. The properties and name code of these clusters are:

- The cluster CL_0 characterized moderate traffic loading, low temperature, and high precipitation conditions and represented by the code MTr_LTe_HP
- The cluster CL_1 characterized low traffic loading, high temperature, and low precipitation conditions and represented by the code LTr_HTe_LP.

- The cluster CL_2 characterized high traffic, moderate temperature, and moderate precipitation conditions and represented by the code HTr_MTe_MP

TABLE 5: Section summary of each cluster

Cluster Name	Number of Sections	Traffic (AADTT)			Temperature (°F)			Precipitation (In)		
		Max	Min	Range	Max	Min	Range	Max	Min	Range
CL_0	44	4748.0	4.0	4744.0	65.8	45.1	20.7	91.3	40.7	50.6
CL_1	14	163.0	62.0	101.0	77.2	71.6	5.6	63.5	31.3	32.2
CL_2	44	5731.0	17.0	5714.0	69.8	52.5	17.3	77.2	34.0	43.3

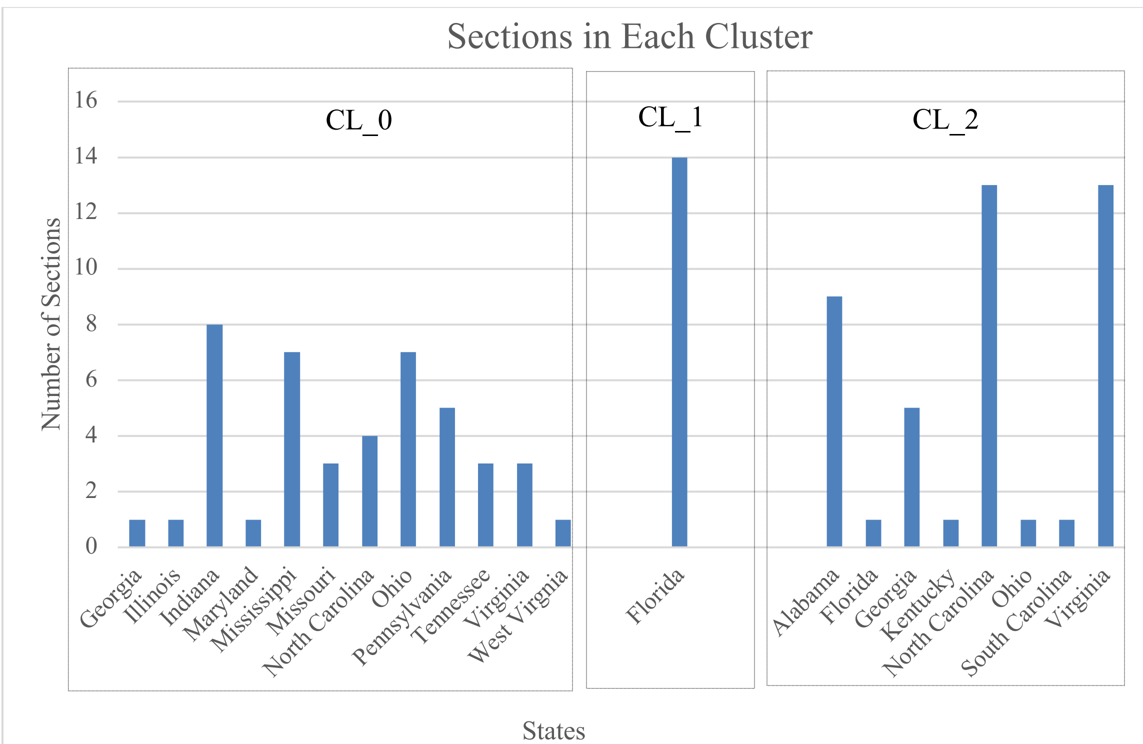
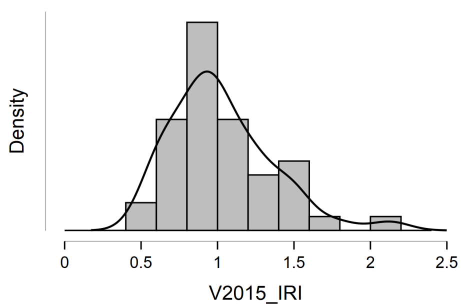


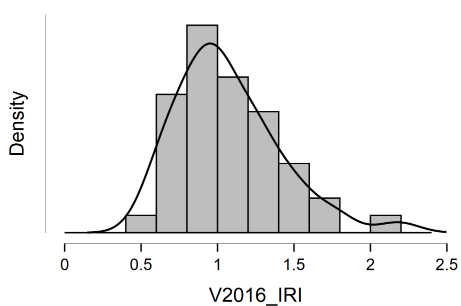
FIGURE 13: State-wise section list of each cluster

The state-wise section list of each cluster is shown in figure 13. The IRI data of each cluster was analyzed to understand its data quality and validate it. For this, the IRI distribution chart and its descriptive statistic were prepared and shown in table 6. These visuals show the interquartile IRI value of pavement sections each year. The IRI value of the 25th, 50th, and 75th percentile of the pavement sections in 2015 are 0.802, 0.963, and 1.206 m/km, while for 2016, these values are 0.837, 1.012, and 1.262 m/km, respectively. As per the descriptive statistics, sections' IRI value in the specified interquartile range tends to increase every year. It suggests that the pavement sections tend to shift towards the right side of the curve, representing higher IRI values. Higher IRI values represent deterioration in pavement sections. Hence, the IRI data collected showed pavement deterioration and was suitable for developing a deterioration model. The IRI distribution for cluster CL_2 is shown in table 6. After completing the clustering analysis and grouping pavement sections, the Markov Chain analysis was performed on these clusters to derive a deterioration model.

TABLE 6: IRI distribution chart and descriptive statistics of cluster CL_2

Distribution Charts





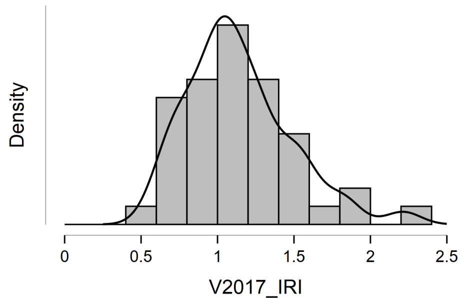
Descriptive Statistics

	IRI
Mean	1.027
Std. Deviation	0.336
Minimum	0.554
Maximum	2.126
25th percentile	0.802
50th percentile	0.963
75th percentile	1.206

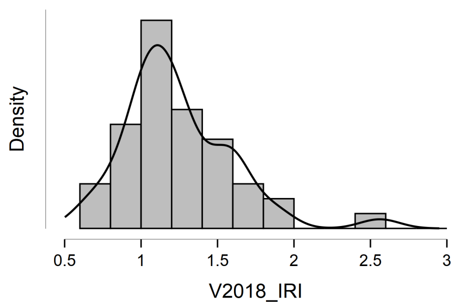
	IRI
Mean	1.070
Std. Deviation	0.342
Minimum	0.550
Maximum	2.181
25th percentile	0.837
50th percentile	1.012
75th percentile	1.262

TABLE 6 (Contd.): IRI distribution chart and descriptive statistics of cluster CL_2

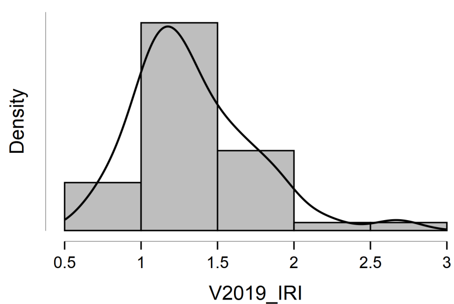
Distribution Charts | **Descriptive Statistics**



	IRI
Mean	1.141
Std. Deviation	0.344
Minimum	0.595
Maximum	2.213
25th percentile	0.923
50th percentile	1.063
75th percentile	1.288



	IRI
Mean	1.240
Std. Deviation	0.357
Minimum	0.637
Maximum	2.561
25th percentile	1.026
50th percentile	1.159
75th percentile	1.438



	IRI
Mean	1.343
Std. Deviation	0.396
Minimum	0.665
Maximum	2.673
25th percentile	1.085
50th percentile	1.276
75th percentile	1.573

3.3 Markov Chain Analysis

This research intends to develop a probabilistic pavement deterioration model and integrate flooding in the model. The Markov Chain analysis is the basis on which the intended model was developed. The Markov Chain analysis comprises developing four matrices: IRI distribution table, transition matrix, transition probability matrix, and probability distribution matrix. First, the IRI distribution table was created, and then the other three matrices were developed based on it. The process acquired for developing these matrices is explained below.

3.3.1 Development of Matrices

1. IRI Distribution Table

The IRI distribution table was prepared by analyzing each section's historical IRI data and then grouping it into their respective IRI bucket each year. A bucket is the IRI range value in m/km. For example, the IRI range of 0.5 to 1.0 m/km can be termed as the IRI bucket range. The derivation of the IRI bucket range was based on the following reasons:

- Pavement sections generally deteriorate with the increment in the IRI value within the range of 0.10 to 0.25 m/km every year (Sayers, 1998),
- The maximum number of pavement sections in each cluster falls within the IRI range of 0.5 to 1.75 m/km; therefore, for uniformly distributing sections into each range bucket, the selection of range buckets was made, and
- Due to the low range of the IRI bucket, small changes in the sections were monitored.

The IRI bucket range was selected 0.25 m/km for cluster CL_0 & CL_2, while 0.2 m/km for cluster CL_1. In cluster CL_1, more than 90% of the sections have the IRI value less than 1.1 m/km, and the number of sections in this cluster is less compared to the other two clusters. Therefore, for uniformly distributing sections in each bucket, the IRI range bucket of the cluster CL_1 was selected 0.2 m/km. Sections in the IRI distribution table formed the basis of developing all three matrices specified earlier.

The IRI distribution table consists of six columns. The first column represents the IRI bucket range in m/km units. The second column represents the number of sections, which is divided into five sub-sections representing the number of sections in five different years from 2015 to 2019. The third column represents the total sections, calculated by adding the number of sections each year in a particular range bucket. The fourth column represents the percentage of sections, calculated by taking the summation of total sections in each range bucket and then dividing it by the sum of total sections row. The fifth row represents the cumulative percentage of the sections, calculated by adding the percentage value of the section in each range bucket to the sum of all of the previous percentage values. The sixth row represents the lower end range of the bucket. The IRI distribution table for all the clusters is shown in tables 7, 8, and 9.

For example, in table 7, there are eight sections in the range bucket of 0.5 to 0.75 m/km in 2015, while six sections in 2017 in the same bucket. The IRI distribution table shows that the number of pavement sections tends to shift in the higher IRI range bucket as the time increases. This trend indicates that the IRI of pavement sections is deteriorating with time.

TABLE 7: IRI Distribution table for cluster CL_0

IRI Range (m/km)	Number of Sections					Total Sections	Percent Section	Cumulative Percent	Lower Limit
	2015	2016	2017	2018	2019				
0.500 - 0.750	8	8	6	2	2	26	0.118	0.000	0.500
0.751 - 1.000	15	10	6	2	2	35	0.159	0.118	0.751
1.001 - 1.250	8	11	9	8	7	43	0.195	0.277	1.001
1.251 - 1.500	4	3	9	11	3	30	0.136	0.473	1.251
1.501 - 1.750	3	5	4	8	12	32	0.145	0.609	1.501

TABLE 7 (Contd.): IRI Distribution table for cluster CL_0

IRI Range (m/km)	Number of Sections					Total Sections	Percent Section	Cumulati ve Percent	Lower Limit
	2015	2016	2017	2018	2019				
1.751 - 2.000	5	4	6	5	7	27	0.123	0.755	1.751
2.001 - 2.250	1	2	3	5	4	15	0.068	0.877	2.001
2.251 - 2.500	0	1	0	1	2	4	0.018	0.945	2.251
2.501 - 2.750	0	0	1	2	3	6	0.027	0.964	2.501
2.751 - 3.000	0	0	0	0	2	2	0.009	0.991	2.751
Total	44	44	44	44	44	220	1.000	1.000	

TABLE 8: IRI Distribution table for cluster CL_1

IRI Range (m/km)	Number of Sections					Total Sections	Percent Section	Cumulati ve Percent	Lower Limit
	2015	2016	2017	2018	2019				
0.500 - 0.700	13	11	8	3	1	36	0.514	0.000	0.500
0.701 - 0.900	0	2	4	8	10	24	0.343	0.514	0.701
0.901 - 1.100	1	1	2	0	0	4	0.057	0.857	0.901
1.101 - 1.300	0	0	0	2	2	4	0.057	0.914	1.101
1.301 - 1.500	0	0	0	1	0	1	0.014	0.971	1.301
1.501 - 1.700	0	0	0	0	1	1	0.014	0.986	1.501
Total	14	14	14	14	14	70	1	1.000	

TABLE 9: IRI Distribution table for cluster CL_2

IRI Range (m/km)	Number of Sections					Total Sections	Percent Section	Cumulati ve Percent	Lower Limit
	2015	2016	2017	2018	2019				
0.500 - 0.750	8	7	5	2	1	23	0.105	0.000	0.500
0.751 – 1.000	16	12	11	8	5	52	0.236	0.105	0.751
1.001 - 1.250	9	10	12	11	9	51	0.232	0.341	1.001
1.251 - 1.500	7	8	5	7	8	35	0.159	0.573	1.251

TABLE 9 (Contd.): IRI Distribution table for cluster CL_2

IRI Range (m/km)	Number of Sections					Total Sections	Percent Section	Cumulati ve Percent	Lower Limit
	2015	2016	2017	2018	2019				
1.501 - 1.750	1	2	5		6	22	0.100	0.732	1.501
1.751 – 2.000	1	2	2	3	6	14	0.064	0.832	1.751
2.001 - 2.250	1	2	1	2	4	10	0.045	0.895	2.001
2.251 - 2.500	1	1	2	2	2	8	0.036	0.941	2.251
2.501 - 2.750	0	0	1	1	1	3	0.014	0.977	2.501
2.751 – 3.000	0	0	0	0	2	2	0.009	0.991	2.751
Total	44	44	44	44	44	220	1.000	1.000	

2. Transition Matrix

The transition matrix is an *m x m* matrix, where *m* represents the number of IRI range buckets. This matrix is prepared by noting the pavement sections' transition from one IRI bucket range to another in the next year. Overall, the transition matrix represents the number of sections that will change its IRI value from one range bucket to another in the next year. The IRI data in each cluster were analyzed and grouped into their respective IRI range bucket for developing the transition matrix. The IRI range bucket was specifically selected to show the section's possible transition point and reflect it in the deterioration model. The number of IRI range buckets for cluster CL_0 & CL_2 is 10, while for cluster CL_1 is 6. Therefore, a ten by ten transition matrix was developed for CL_0 & CL_2, and a six by six transition matrix was developed for CL_1. In total, three transition matrices were developed.

The transition matrix was developed using the five years of IRI data from 2015 to 2019 of each cluster's sections. The IRI data were combined into four groups which were representing the change in IRI between each consecutive year. The four-year groups were:

2015-2016, 2016-2017, 2017-2018, and 2018-2019. The pavement sections that show improvement in the condition were not considered in the analysis because data showing such a decrease in IRI value without any maintenance is not realistic for pavement deterioration. The transition matrix for each cluster is shown in tables 10, 11 & 12. These tables represent the deterioration of pavement sections. Each cell's values represent the number of sections that transitioned its state from one IRI bucket range to another in the next year. This matrix satisfies the requirements of the Markov property. Therefore, it was further used in developing the Markov Chain prediction model.

In tables 10 to 12, the cells showing a zero represent no transition of pavement sections in this IRI bucket range for the next year. For example, in table 10, 3 pavement sections transitioned from the IRI bucket range of 0.5-0.75 m/km to 0.751-1.0 m/km range, while one sections transitioned from 0.5-0.75 m/km to 1.0-1.250 m/km range. The computer program was written on Microsoft Visual Basic for developing a transition matrix for each cluster. This computer program was used because the data collected from the LTPP database was in Excel format, and MS Visual Basic can be used efficiently on Excel sheets. This transition matrix was further used in developing the Markovian Chain transition probability matrix.

TABLE 10: Transition Matrix for the Cluster CL_0

IRI Bucket Range (m/km)	Pavement section transition details										
	Number of sections in next year										
Current Year IRI Bucket	0.500 - 0.750	0.751 - 1.000	1.001 - 1.250	1.251 - 1.500	1.501 - 1.750	1.751 - 2.000	2.001 - 2.250	2.251 - 2.500	2.501 - 2.750	2.751 - 3.000	Total
0.500 - 0.750	16	3	1	0	2	0	0	0	0	0	22
0.751 - 1.000	0	17	9	4	1	2	0	0	0	0	33

TABLE 10 (Contd.): Transition Matrix for the Cluster CL_0

IRI Bucket Range (m/km)	Pavement section transition details Number of sections in next year										
Current Year IRI Bucket	0.500 - 0.750	0.751 - 1.000	1.001 - 1.250	1.251 - 1.500	1.501 - 1.750	1.751 - 2.000	2.001 - 2.250	2.251 - 2.500	2.501 - 2.750	2.751 - 3.000	Total
1.001 - 1.250	0	0	24	9	2	0	1	0	0	0	36
1.251 - 1.500	0	0	1	12	11	1	2	0	0	0	27
1.501 - 1.750	0	0	0	1	13	6	0	0	0	0	20
1.751 - 2.000	0	0	0	0	0	13	5	1	1	0	20
2.001 - 2.250	0	0	0	0	0	0	6	3	1	1	11
2.251 - 2.500	0	0	0	0	0	0	0	0	3	2	5
2.501 - 2.750	0	0	0	0	0	0	0	0	2	1	3
2.751 - 3.000	0	0	0	0	0	0	0	0	0	2	2

TABLE 11: Transition Matrix for the Cluster CL_1

IRI Bucket Range (m/km)	Pavement section transition details Number of sections in next year						
Current Year IRI Bucket	0.500 - 0.700	0.701 - 0.900	0.901 - 1.100	1.101 - 1.300	1.301 - 1.500	1.501 - 1.700	Total
0.500 - 0.700	23	11	0	0	0	0	34
0.701 - 0.900	0	13	1	0	0	0	14
0.901 - 1.100	0	0	2	2	0	0	4
1.101 - 1.300	0	0	0	1	1	0	2
1.301 - 1.500	0	0	0	0	0	1	1
1.501 - 1.700	0	0	0	0	0	1	1

TABLE 12: Transition Matrix for the Cluster CL_2

IRI Bucket Range (m/km)	Pavement section transition details										
	Number of sections in next year										
Current Year IRI Bucket	0.500 - 0.750	0.751 - 1.000	1.001 - 1.250	1.251 - 1.500	1.501 - 1.750	1.751 - 2.000	2.001 - 2.250	2.251 - 2.500	2.501 - 2.750	2.751 - 3.000	Total
0.500 - 0.750	16	5	0	1	2	0	0	0	0	0	24
0.751 - 1.000	0	34	14	2	0	0	0	0	0	0	50
1.001 - 1.250	0	0	41	8	0	1	0	0	0	0	50
1.251 - 1.500	0	0	1	21	6	0	0	0	0	0	28
1.501 - 1.750	0	0	0	0	13	3	0	0	0	0	16
1.751 - 2.000	0	0	0	0	0	5	1	0	0	0	6
2.001 - 2.250	0	0	0	0	0	0	2	0	1	0	3
2.251 - 2.500	0	0	0	0	0	0	0	4	2	0	6
2.501 - 2.750	0	0	0	0	0	0	0	0	2	1	3
2.751 - 3.000	0	0	0	0	0	0	0	0	0	1	1

3. Transition Probability Matrix

The transition probability matrix was developed using the transition matrix shown in tables 10 to 12. The percentage transition method was found feasible for generating the Markov Chain transition probability matrix. The reason behind using the percentage transition method for deriving TPMs is specified in the literature review section. Equation 3, specified in the literature review section, was used for deriving the TPMs. A detailed explanation of the generation of TPMs using the percentage transition method is explained in the literature review section.

The transition probability matrix for each cluster is shown in tables 13, 14, and 15. For example, in table 13, 72.7% of pavement sections remain in the same IRI bucket range of 0.5 – 0.75 m/km for the next year, while 13.6% of pavement sections change its state to

0.751-1.000 m/km IRI range bucket. This matrix was further used in developing a probability distribution matrix. The TPM generated in this section represents a non-flood condition. These TPMs are further used in developing a probabilistic pavement deterioration model for non-flood conditions.

TABLE 13: Transition Probability Matrix of Cluster CL_0

IRI Bucket Range (m/km)	Transition Probability of Sections										
	Percentage of Sections in Next Year										
Current Year IRI Bucket	0.500 - 0.750	0.751 - 1.000	1.001 - 1.250	1.251 - 1.500	1.501 - 1.750	1.751 - 2.000	2.001 - 2.250	2.251 - 2.500	2.501 - 2.750	2.751 - 3.000	Total
0.500 - 0.750	0.727	0.136	0.045	0	0.091	0	0	0	0	0	1
0.751 - 1.000	0	0.515	0.273	0.121	0.030	0.061	0	0	0	0	1
1.001 - 1.250	0	0	0.667	0.250	0.056	0	0.028	0	0	0	1
1.251 - 1.500	0	0	0.037	0.444	0.407	0.037	0.074	0	0	0	1
1.501 - 1.750	0	0	0	0.050	0.650	0.300	0	0	0	0	1
1.751 - 2.000	0	0	0	0	0	0.650	0.250	0.050	0.050	0	1
2.001 - 2.250	0	0	0	0	0	0	0.545	0.273	0.091	0.091	1
2.251 - 2.500	0	0	0	0	0	0	0	0	0.600	0.400	1
2.501 - 2.750	0	0	0	0	0	0	0	0	0.667	0.333	1
2.751 - 3.000	0	0	0	0	0	0	0	0	0	1.000	1

TABLE 14: Transition Probability Matrix of Cluster CL_1

IRI Bucket Range (m/km)	Transition Probability of Sections						
	Percentage of Sections in Next Year						
Current Year IRI Bucket	0.500 - 0.700	0.701 - 0.900	0.901 - 1.100	1.101 - 1.300	1.301 - 1.500	1.501 - 1.700	Total
0.500 - 0.700	0.676	0.324	0	0	0	0	1
0.701 - 0.900	0	0.929	0.071	0	0	0	1

TABLE 14 (Contd.): Transition Probability Matrix of Cluster CL_1

IRI Bucket Range (m/km)	Transition Probability of Sections — Percentage of Sections in Next Year						
Current Year IRI Bucket	0.500 - 0.700	0.701 - 0.900	0.901 - 1.100	1.101 - 1.300	1.301 - 1.500	1.501 - 1.700	Total
0.901 - 1.100	0	0	0.500	0.500	0	0	1
1.101 - 1.300	0	0	0	0.500	0.500	0	1
1.301 - 1.500	0	0	0	0	0	1.000	1
1.501 - 1.700	0	0	0	0	0	1.000	1

TABLE 15: Transition Probability Matrix of Cluster CL_2

IRI Bucket Range (m/km)	Pavement section transition details for the next year — Sections in Next Year IRI Bucket										
Current Year IRI Bucket	0.500 - 0.750	0.751 - 1.000	1.001 - 1.250	1.251 - 1.500	1.501 - 1.750	1.751 - 2.000	2.001 - 2.250	2.251 - 2.500	2.501 - 2.750	2.751 - 3.000	Total
0.500 - 0.750	0.667	0.208	0.000	0.042	0.083	0	0	0	0	0	1
0.751 - 1.000	0	0.680	0.280	0.040	0	0	0	0	0	0	1
1.001 - 1.250	0	0	0.820	0.160	0	0.020	0	0	0	0	1
1.251 - 1.500	0	0	0.036	0.750	0.214	0.000	0	0	0	0	1
1.501 - 1.750	0	0	0	0	0.813	0.188	0	0	0	0	1
1.751 - 2.000	0	0	0	0	0	0.833	0.167	0	0	0	1
2.001 - 2.250	0	0	0	0	0	0	0.667	0	0.333	0	1
2.251 - 2.500	0	0	0	0	0	0	0	0.667	0.333	0	1
2.501 - 2.750	0	0	0	0	0	0	0	0	0.667	0.333	1
2.751 - 3.000	0	0	0	0	0	0	0	0	0	1.000	1

4. Probability Distribution Matrix

The probability distribution matrix is used to predict the future condition of the pavement at any given year. Equation 2 was used for developing the probability distribution matrix of each cluster. The vector *a(0)* in equation (2) represents the pavement's initial condition. It was assumed that the initial condition of the pavement was perfect. The initial state vector *a(0)* is shown by the matrix below. The variable P^t is the transition probability matrix, shown in tables 13 to 15 for each cluster. The probability distribution matrix was generated by substituting the variables *a(0)* and P^t in equation 2. The probability distribution matrix for each cluster is shown in tables 16, 17, & 18.

$$a(0) = [1 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0]$$

TABLE 16: Probability Distribution Matrix for cluster CL_0

IRI Range (m/km)	Year1	Year2	Year3	Year4	Year5
0.500 - 0.750	0.727	0.529	0.385	0.280	0.203
0.751 - 1.000	0.136	0.169	0.159	0.135	0.107
1.001 - 1.250	0.045	0.101	0.138	0.156	0.157
1.251 - 1.500	0.000	0.032	0.067	0.091	0.105
1.501 - 1.750	0.091	0.132	0.158	0.177	0.191
1.751 - 2.000	0.000	0.036	0.074	0.108	0.135
2.001 - 2.250	0.000	0.001	0.015	0.035	0.057
2.251 - 2.500	0.000	0.000	0.002	0.008	0.015
2.501 - 2.750	0.000	0.000	0.002	0.008	0.018
2.751 - 3.000	0.000	0.000	0.000	0.003	0.012
Total	1.000	1.000	1.000	1.000	1.000

TABLE 17: Probability Distribution Matrix for cluster CL_1

IRI Range (m/km)	Year1	Year2	Year3	Year4	Year5
0.500 - 0.700	0.676	0.457	0.309	0.209	0.141
0.701 - 0.900	0.324	0.520	0.631	0.686	0.705
0.901 - 1.100	0.000	0.023	0.048	0.069	0.083

TABLE 17 (Contd.): Probability Distribution Matrix for cluster CL_1

IRI Range (m/km)	Year1	Year2	Year3	Year4	Year5
1.101 - 1.300	0.000	0.000	0.012	0.030	0.049
1.301 - 1.500	0.000	0.000	0.000	0.006	0.015
1.501 - 1.700	0.000	0.000	0.000	0.000	0.006
Total	1.000	1.000	1.000	1.000	1.000

TABLE 18: Probability Distribution Matrix for cluster CL_2

IRI Range (m/km)	Year1	Year2	Year3	Year4	Year5
0.500 - 0.750	0.667	0.444	0.296	0.198	0.132
0.751 - 1.000	0.208	0.281	0.283	0.254	0.214
1.001 - 1.250	0.000	0.060	0.130	0.189	0.230
1.251 - 1.500	0.042	0.067	0.090	0.112	0.133
1.501 - 1.750	0.083	0.132	0.159	0.173	0.181
1.751 - 2.000	0.000	0.016	0.039	0.065	0.090
2.001 - 2.250	0.000	0.000	0.003	0.008	0.016
2.251 - 2.500	0.000	0.000	0.000	0.000	0.000
2.501 - 2.750	0.000	0.000	0.000	0.001	0.003
2.751 - 3.000	0.000	0.000	0.000	0.000	0.000
Total	1.000	1.000	1.000	1.000	1.000

Tables 16 to 18 represent the prediction of pavement sections in a particular IRI range bucket. For example, in table 16, at year 1, 72.7% of sections will remain in the IRI range of 0.5-0.75 m/km, while in year 5, 20.3% of the section will remain this IRI range, and so on forth. After the generation of these three matrices, the Markovian Chain analysis is completed. These generated matrices were used in developing the deterioration model for all the clusters. In section 3.5, these matrices were used for proposing a framework for incorporating the effect of flooding in the deterioration model.

3.4 Model Generation

The main objective of this research is to generate a probabilistic pavement deterioration model. The Monte Carlo simulation was used for generating a probabilistic deterioration model and achieving this objective. In the Monte Carlo simulation, an uncertain variable is assigned multiple values by random variables' intervention to achieve multiple results. The multiple results are achieved through numerous trials, and then the average of these trials is taken to estimate the most favorable outcome. The random variables can be any numerical values that are used in numerous trials to predict the outcome. The pavements' roughness (IRI) is uncertain, as shown in tables 16 to 18, also represents different distribution trends shown in table 6. Therefore, the Monte Carlo simulation was used to incorporate IRI's probabilistic nature with numerous trials and then generate the most favorable deterioration model for each cluster.

The implementation of the Monte Carlo simulation was done by transforming the TPM into cumulative TPM. The cumulative TPM for cluster CL_0 is shown in table 19 and appendix A1 & A2 for cluster CL_1 and CL_2, respectively. The Monte Carlo simulation algorithm was developed in Microsoft Visual Basic. In this simulation, a comparison was made between the cumulative TPM values and the random numbers for predicting the future IRI values of the pavement sections. In the simulation, 20 uniformly distributed random numbers between 0 and 1 were generated for each iteration. This process was repeated for 1500 iterations, and in total, 30,000 random numbers were generated in the simulation. These random numbers represent the IRI probability and are used to predict future IRI values. In each iteration, the pavement deterioration model for the next 20 years was generated, and in total, 1500 deterioration models were generated. The model generated in this section is a probabilistic pavement deterioration model, and it accounts for non-flooding conditions.

TABLE 19: Cumulative Transition Probability Matrix for CL_0

IRI Bucket Range (m/km)	Cumulative TPM next year									
Current Year IRI Bucket	0.500 - 0.750	0.751 - 1.000	1.001 - 1.250	1.251 - 1.500	1.501 - 1.750	1.751 - 2.000	2.001 - 2.250	2.251 - 2.500	2.501 - 2.750	2.751 - 3.000
0.500 - 0.750	0.7273	0.8636	0.9091	0.9091	1	1	1	1	1	1
0.751 - 1.000	0	0.5152	0.7879	0.9091	0.9394	1	1	1	1	1
1.001 - 1.250	0	0	0.6667	0.9167	0.9722	0.9722	1	1	1	1
1.251 - 1.500	0	0	0.037	0.4815	0.8889	0.9259	1	1	1	1
1.501 - 1.750	0	0	0	0.05	0.7	1	1	1	1	1
1.751 - 2.000	0	0	0	0	0	0.65	0.9	0.95	1	1
2.001 - 2.250	0	0	0	0	0	0	0.5455	0.8182	0.9091	1
2.251 - 2.500	0	0	0	0	0	0	0	0	0.6	1
2.501 - 2.750	0	0	0	0	0	0	0	0	0.6667	1
2.751 - 3.000	0	0	0	0	0	0	0	0	0	1

3.4.1 Monte Carlo simulation logic

The random number in each iteration was compared with the cumulative TPM values. The IRI range bucket of 0.5 to 0.75 m/km represents the perfectly smooth pavement surface; therefore, the comparison starts from this IRI range. The comparison was started from 0.5 to 0.75 m/km range, continued towards the right side of the matrix, and stopped when the cumulative TPM value was greater than the random number. The IRI range in which the comparison stops represents the IRI value of the following year's pavement section. The next iteration will start from the same IRI range in which the last iteration stopped. This procedure was continued for each of the 20 random numbers in a trial and repeated for 1500 trials.

For example, in reference to table 19, the first random number generated was 0.52. This random number was compared with the cumulative TPM value 0.727 that is located in the leftmost corner, in the 0.5 to 0.75 m/km IRI range. The random number 0.52 is less than 0.727; therefore, the IRI transition in the first year did not happen, and the IRI value of 0.5 was allocated in this trial. If the second next random number generated was 0.75, it is compared with 0.727, i.e., 0.5 to 0.75 m/km IRI range. The random number 0.75 is greater than 0.727, so the comparison moves to the next IRI range, which is 0.75 to 1.0 m/km. The cumulative TPM value in this IRI range is 0.864, which is greater than 0.75; therefore, the comparison stops here, and the IRI value of 0.751 was allocated for the second year. If the third random number again stopped in the same IRI range of 0.75 to 1.0 m/km, the IRI value allocated for the third year would be 0.752. This kind of pattern will continue until a random number stopped in a different IRI range.

This process will continue for each of the 20 random numbers and 1500 iterations. Each iteration generates a pavement deterioration model for the next 20 years; therefore, a total of 1500 deterioration models were generated by the simulation. The final deterioration model was generated by taking the average of all the IRI values in their respective year in each iteration. Each year's average IRI value is plotted against time, in years, to obtain the deterioration model. The deterioration model of each cluster is shown in figures 15 to 17. These models represent pavements' deterioration for the non-flooding situation. The results derived from the deterioration models are explained in section 4.1.

3.5 Flooding Framework

A framework was designed for incorporating the effect of flooding in the deterioration model, and it is explained thoroughly in this section. This framework was developed based on these four assumptions:

- The initial pavement condition is excellent,
- Roughness is majorly affected by the accumulation of flooded water on the pavement surface,
- Hypothetical flooding event will occur between the year 2020 and 2021, and
- No rehabilitation work will be done for the next 3-4 years.

Hence, a pavement deterioration model was developed that shows a change in the IRI of the pavement surface at a different flooding probability. A Monte Carlo simulation code was developed for incorporating the effect of flooding in the deterioration model. In the simulation code, two types of TPMs were used: non-flooding TPM and flooding TPM. The non-flooding TPM represents no flood event that occurred in the past, while the flood TPM represents flood occurrence in the past. Ideally, both TPMs needed to be developed based on the flood-affected pavement sections' historical IRI data. In this research, the LTPP InfoPave database was used to collect the pavement sections' historical IRI data from 2015 to 2019. The LTPP database does not contain the IRI data of flood-affected pavement sections. Therefore, it was not possible to develop an actual flood TPM; instead, a hypothetical flood TPM was developed for proposing the framework for incorporating flooding in the deterioration model. The hypothetical flooding matrix was developed for each cluster. The TPMs developed in the previous section are the non-flood TPMs and was used in the place of non-flood TPMs in the simulation. The Non-flooding matrix used in the simulation for cluster CL_0 is shown in table 13, while the hypothetical flooding matrix for the cluster CL_0 is shown in table 20. Both of these matrices satisfy the Markov Chain property.

TABLE 20: Hypothetical Flooding TPM for cluster CL_0

IRI Bucket Range (m/km)	Percentage of Sections in Next Year										
Current Year IRI Bucket	0.500 - 0.750	0.751 - 1.000	1.001 - 1.250	1.251 - 1.500	1.501 - 1.750	1.751 - 2.000	2.001 - 2.250	2.251 - 2.500	2.501 - 2.750	2.751 - 3.000	Total
0.500 - 0.750	0.200	0.8	0	0	0	0	0	0	0	0	1
0.751 - 1.000	0	0.297	0.703	0	0	0	0	0	0	0	1
1.001 - 1.250	0	0	0.48	0.52	0	0	0	0	0	0	1
1.251 - 1.500	0	0	0	0.567	0.433	0	0	0	0	0	1
1.501 - 1.750	0	0	0	0	0.727	0.273	0	0	0	0	1
1.751 - 2.000	0	0	0	0	0	0.611	0.389	0	0	0	1
2.001 - 2.250	0	0	0	0	0	0	0.615	0.385	0	0	1
2.251 - 2.500	0	0	0	0	0	0	0	0.556	0.444	0	1
2.501 - 2.750	0	0	0	0	0	0	0	0	0.444	0.556	1
2.751 - 3.000	0	0	0	0	0	0	0	0	0	1	1

The framework assumes that the pavements' roughness was majorly affected by the accumulation of flooded water on the pavement surface. The effect of the magnitude of flood on roughness was not considered in this study. For this, the flood recurrence interval was studied, and annual flooding probability was determined based on it. The flooding event of different recurrence intervals is assumed to occur in the future, which affects the pavements' performance. It was assumed that as the annual flooding probability increases, the inundation of pavement surface increases, which ultimately results in a higher rate of pavement deterioration. The flooding event with the recurrence interval of annual, 2-years, 5-years, 10-years, and 20-years was selected for developing a deterioration model. The probability associated with the above-specified flooding events is 1, 0.5, 0.2, 0.1, and 0.05, respectively. The deterioration model of pavement sections was generated at these specified flooding probabilities. The Monte Carlo simulation was used for generating the deterioration models. This code was written on MATLAB.

In the simulation, non-flood TPM and flood TPM used. For cluster CL_0, the two TPMs are shown in tables 13 and 20, respectively. An initial pavement condition matrix was developed by assuming that the pavement is in excellent condition in the year 2020. State vectors representing the transition of the pavement sections into various states were generated. Then a set of random numbers are compared with the flooding probability to determine if a non-flooding TPM or a flooding-TPM should be utilized for further process. The chance of selecting flooding TPM depends upon the chance of flood occurrence. After selecting the TPM, a second random variable is generated to estimate the pavement's future state. The final pavement state is generated by taking the average of all the simulated states. This process is repeated for 10,000 trials and 20 years. At the end of the simulation, the deterioration model was generated for different flooding probabilities. The logic of the simulation code is shown in figure 14. The deterioration model at different flooding probabilities is shown in figures 18, 19, and 20 for each cluster. The results derived from this proposed model is explained in section 4.2 of this document.

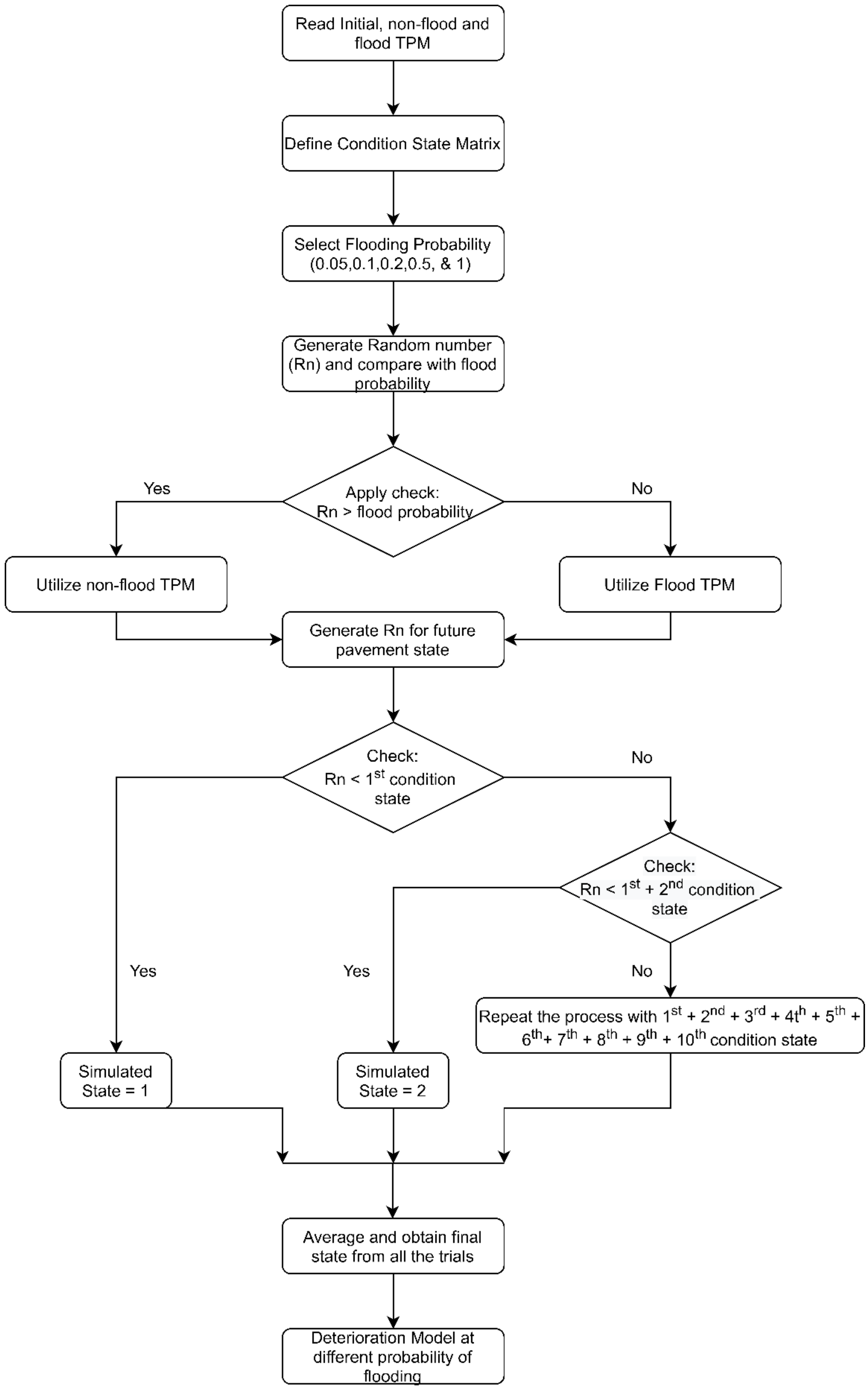


FIGURE 14: Simulation Code logic

CHAPTER 4: RESULTS

The results derived from the methodology suggested for generating the pavement deterioration model in the previous section are illustrated in this section. The deterioration model generated was based on the roughness characteristics of the pavement sections. Therefore, the results derived from the models are explained in terms of the roughness of the pavement. These results are presented in sections 4.1 and 4.2.

4.1 Results of deterioration model with no-flood

The results derived from the probabilistic pavement deterioration model, without incorporating the flooding effect, are explained in this section. The generation process of these models is explained in section 3.4. The deterioration model for cluster CL_0, CL_1, and CL_2 was generated separately.

The cluster CL_0 was subjected to moderate traffic, low temperature, and high precipitation and represented by the code MTr_LTe_HP. The deterioration curve for this cluster is shown in figure 15. The trend illustrates that the IRI will increase throughout 20 years, representing continuous pavement deterioration. From 2020 to 2025, the IRI of these sections will increase by the average rate of 0.150 m/km each year; after 2026, it will increase by the average rate of 0.135 m/km till 2029; then from 2030, it will increase by the average rate of 0.093 every year till 2034; and then from 2034, it will increase by the average rate of 0.052 m/km every year till 2039. The increment in the deterioration rate will be highest for this cluster compared to the other two clusters. This pattern will be shown because the sections are subjected to heavy precipitation and moderate traffic loading conditions. Therefore, for initial years, the rate of deterioration is high, and it tends to decrease as the age of pavement increases.

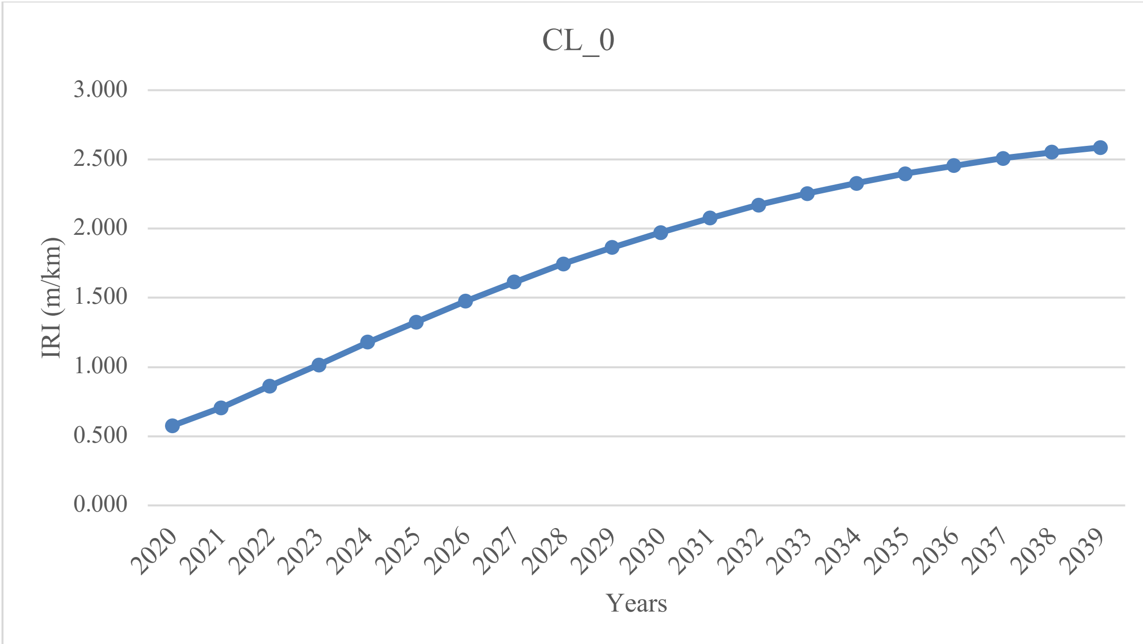


FIGURE 15: Deterioration Model for cluster CL_0

The cluster CL_1 was subjected to low traffic loading, high temperature, and low precipitation condition and represented by the code LTr_HTe_LP. The deterioration curve for this cluster is shown in figure 16. This cluster's roughness value will start at 0.50 m/km in 2020 and reach 1.15 m/km in 2039. The trend illustrates that the IRI will increase throughout 20 years, representing continuous pavement deterioration. From 2020 to 2025, the IRI of these sections will increase by the average rate of 0.02 m/km each year; after 2026, it will increase by the average rate of 0.043 m/km till 2034; and then from 2035, it will increase by 0.030 m/km every year until 2039. Initially, from 2020 to 2025, this cluster's deterioration rate is low compared to the other two clusters because the sections of this cluster are subjected to lower traffic and precipitation. However, the deterioration rate will increase as the pavement age increases because these sections are subjected to high temperatures.

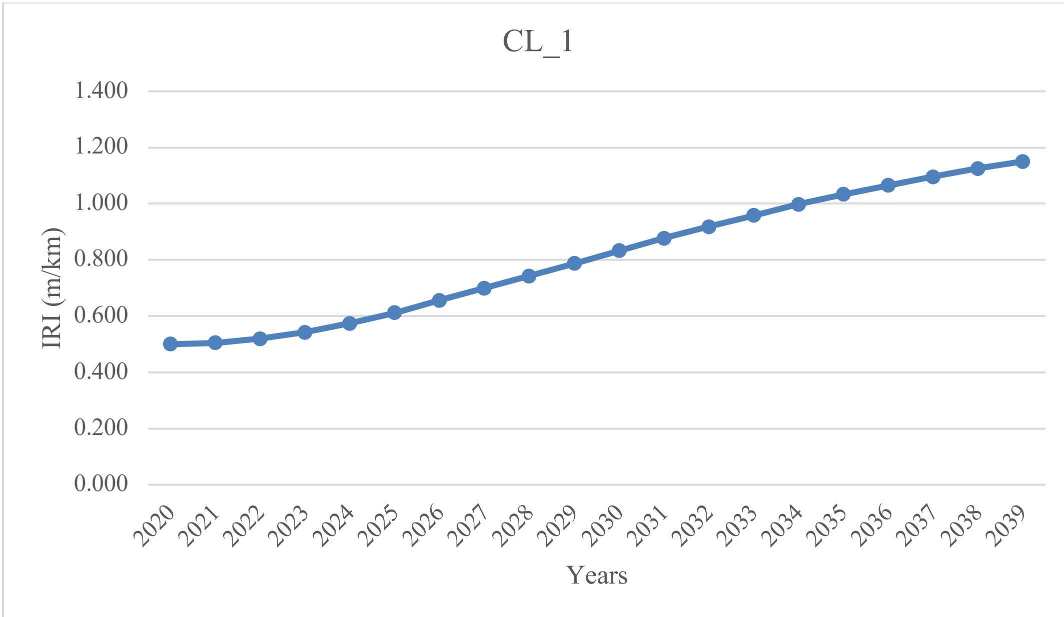


FIGURE 16: Deterioration Model for cluster CL_1

The cluster CL_2 was subjected to high traffic loading, moderate temperature, and moderate precipitation condition and represented by the code HTr_MTe_MP. The deterioration curve for this cluster is shown in figure 17. This cluster's roughness value will start at 0.585 m/km in 2020 and reach 2.005 m/km in 2039. The trend illustrates that the IRI will increase throughout 20 years, representing continuous pavement deterioration. From 2020 to 2026, the IRI will increase by the average rate of 0.103 m/km each year; after 2026, it will increase by the average rate of 0.067 m/km till 2032; and then from 2033, it will increase by the average rate of 0.057 m/km each year until 2039. Usually, during the initial years, the rate of pavement deterioration is maximum. These sections are also subjected to heavy loading conditions; therefore, this trend aligns with expectations. The rate of deterioration of this cluster will be high from 2020 to 2026, and then it will start decreasing as the year progresses. This cluster represents the classic example of a newly constructed pavement subjected to heavy traffic loading and moderate climatic conditions.

These kind of pavements tend to deteriorate faster just after its construction, and the deterioration rate tends to decrease as the age of pavement increases.

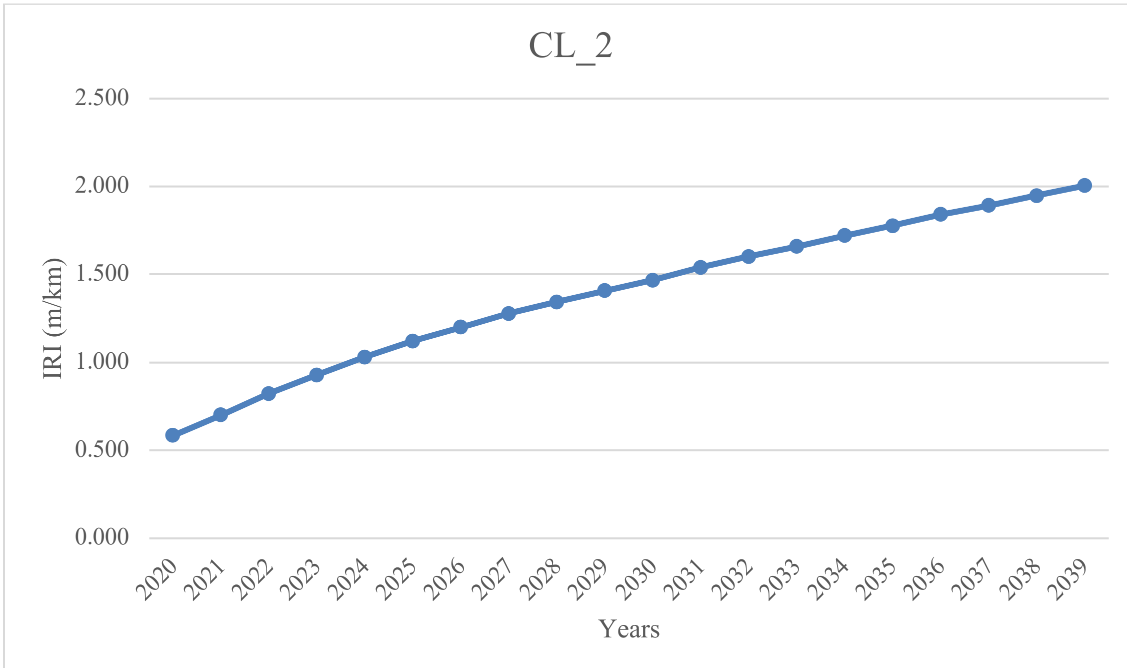


FIGURE 17: Deterioration model for cluster CL_2

Generally, the pavement tends to deteriorate faster in initial years if subjected to heavy or moderate traffic loading and precipitation condition. This rate will tend to decrease as the age of pavement increases. Similar kinds of trends were shown by the deterioration model of cluster CL_0 and CL_2. If the pavement sections are subjected to low traffic loading and low precipitation, the rate of deterioration for initial years is less, but it will tend to increase as the age of pavement increases. The deterioration model of cluster CL_1 showed a similar kind of trend.

4.2 Results of deterioration model with flood

The results derived from the probabilistic pavement deterioration model by incorporating flooding are explained in this section. The generation process of these models is explained in section 3.5. The graphs in figures 18, 19, and 20 illustrate the expected change in pavement's roughness by predicting a jump in the IRI value due to hypothetical flooding events. These graphs represent pavement deterioration envelop at different flooding probabilities and help to understand the framework proposed in section 3.5 of this document.

Figures 18, 19, and 20 show the predicted roughness of sections in cluster CL_0, CL_1, and CL_2, respectively, at different flooding probabilities. The graphs were prepared by assuming that flooding events will occur between 2020 to 2021, and the impact will be represented by the increment in the sections' roughness. The graph illustrates that the impact of flooding on the pavement's roughness is maximum when the probability of flooding is maximum. In other words, the higher the probability of flooding, the higher the inundation of the pavement surface, which ultimately results in a higher rate of pavement deterioration. For example, in figure 18, the roughness of pavement in 2021 will be 1.603 m/km at a 5% probability of flood, while it will be 1.747 m/km at a 50 % probability of flood. The Monte Carlo simulation was used to predict the results. The impact of flooding on pavement sections is maximum initially; therefore, the deterioration rate is maximum for the initial period. The roughness-based deterioration model tends to decrease when post-flood maintenance is applied of the sections; therefore, roughness prediction for the first few years was shown at different flooding probabilities. The flood's maximum impact is shown in 2021 because it occurred between 2020 and 2021, in all the figures. This impact tends to reduce as time increases. This framework can be utilized when the actual flooding data is available to predict the pavement performance for post-flood times and prepare the maintenance budget.

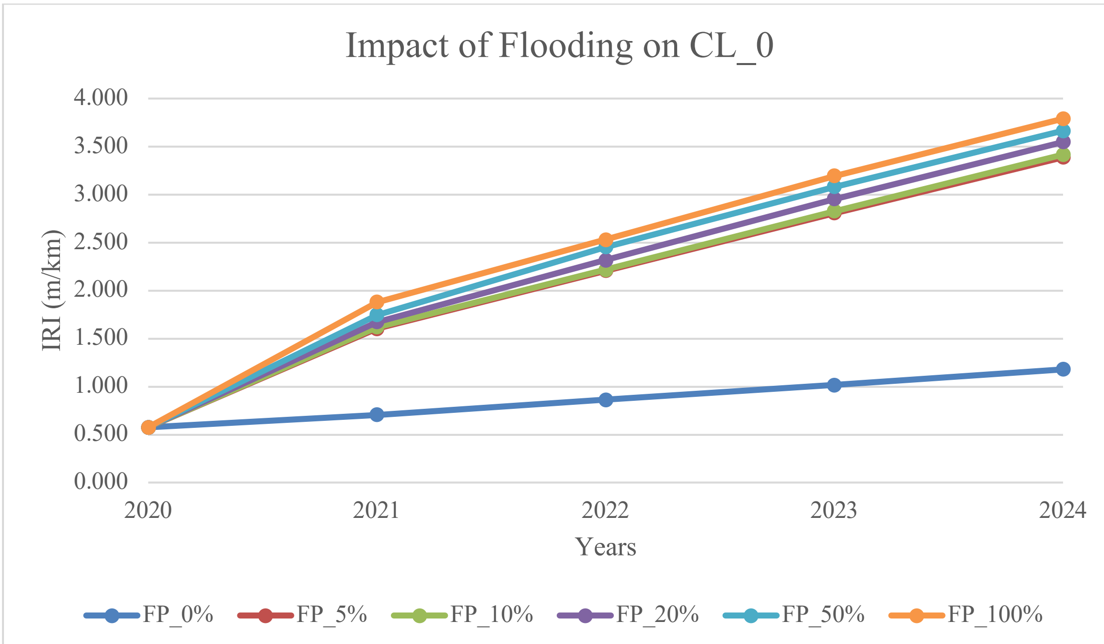


FIGURE 18: Pavement deterioration model at different probability of flooding for CL_0

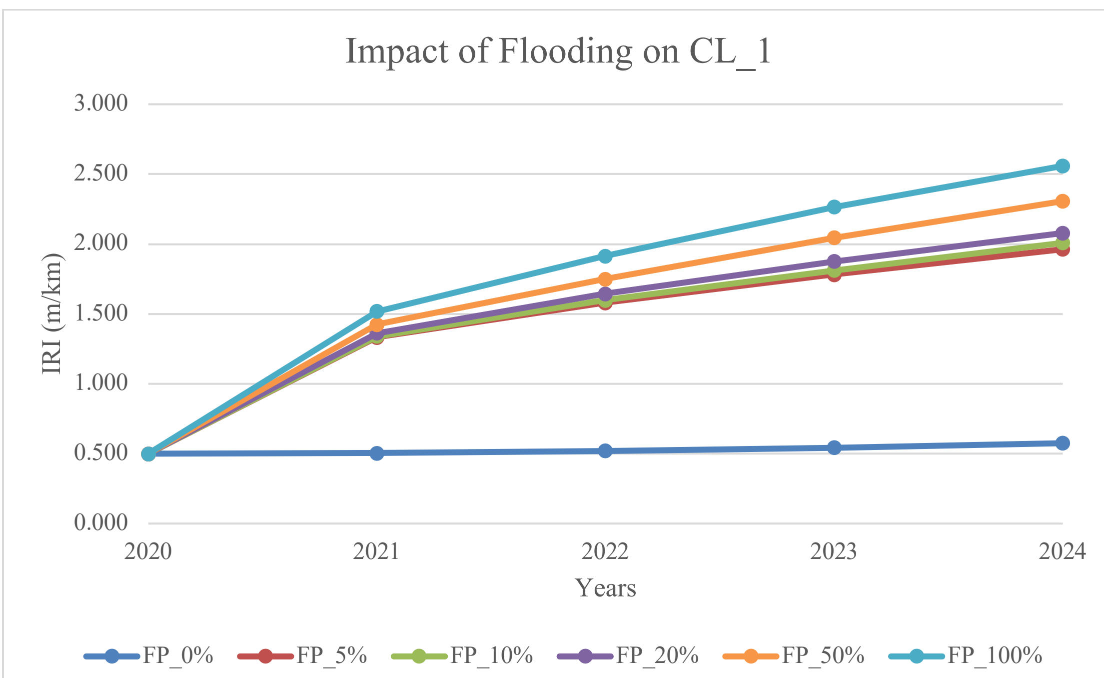


FIGURE 19: Pavement deterioration model at different probability of flooding for CL_1

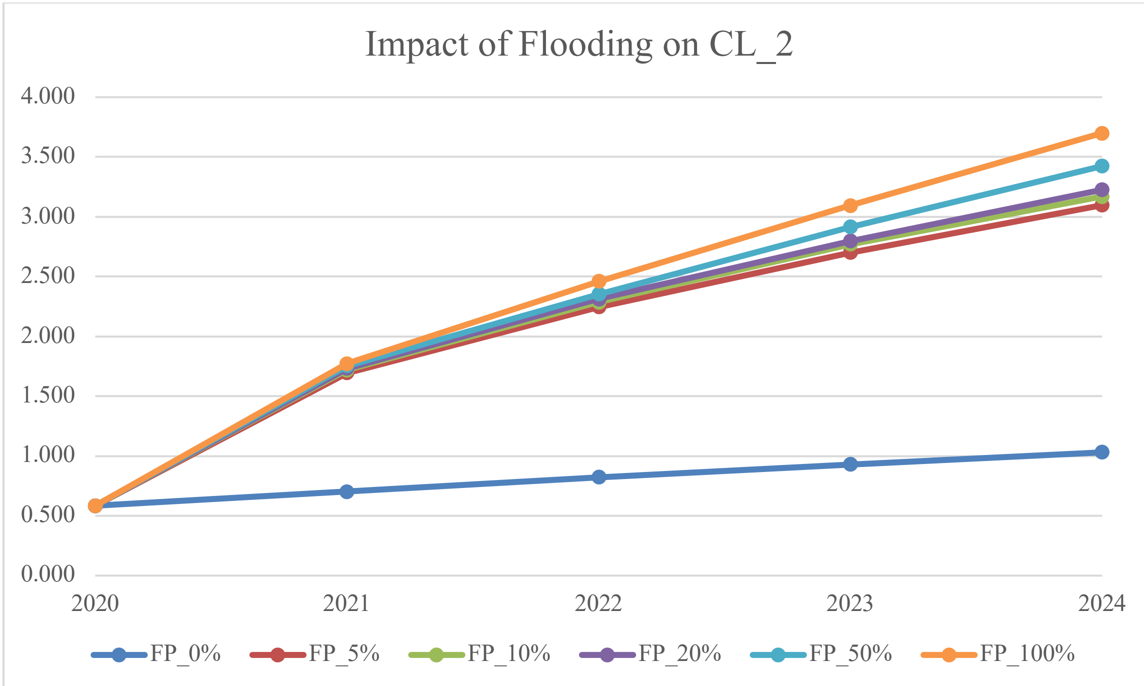


FIGURE 20: Pavement deterioration model at different probability of flooding for CL_2

CHAPTER 5: CONCLUSION

5.1 Necessity of this research

The primary objective of this research was to develop a cluster-based probabilistic pavement deterioration model for composite pavements. Also, to propose a framework for incorporating the effect of flooding in the pavement deterioration model. The research objective was developed based on the gaps identified in the literature review. A thorough literature review was conducted and documented previously in chapter 2. In the literature review, limited studies have focused on using the LTPP InfoPave database for developing a probabilistic pavement deterioration model and integrate the effect of flooding for composite pavements.

Extreme weather events such as flooding and frequent heavy rainfall cause increment in the deterioration rate and a significant reduction in the strength of pavement sections; however, few studies were conducted to address this issue. Regression-based deterministic deterioration models were developed in previous research that does not address flooding in the model. Therefore, a study was required that proposes a deterioration model by addressing all these issues so that the states' transportation department can utilize this model to improve their pavement management system. It is believed that the current study addresses these issues appropriately, and the framework can be utilized when the actual flooding data will be available.

5.2 Findings of this research

The methodology utilized for achieving the objective is thoroughly explained in chapter 3 of this document. In the current study, the cluster-based pavement deterioration model was developed using the Markov Chain analysis, and a framework is proposed to incorporate the effect of flooding in a deterioration model. Historical roughness (IRI), traffic loading, temperature, and precipitation data from 2015 to 2019 were used to develop the model. The LTPP InfoPave database was used for extracting the data of all these parameters. A total of 102 pavement sections from 16 different states of the USA was selected for the analysis. These sections were divided into three groups known as clusters with the help of the *K*-means clustering technique. These pavement sections are clustered based on each section's traffic loading, temperature, and precipitation characteristics, and the clusters are named CL_0, CL_1, and CL_2. The deterioration model for all three clusters was developed separately based on each section's roughness. The pavement's roughness is stochastic; therefore, the Markov Chain analysis was used to account for this behavior. The IRI vs. time data was prepared for each cluster, and then Markov Chain's analysis was applied for preparing four matrices, namely, IRI distribution table, transition matrix, transition probability matrix, and probability distribution matrix. These matrices were further used in the Monte Carlo simulation to generate each cluster's deterioration model for the next 20 years. A framework was proposed for incorporating the effect of flooding in the deterioration model. In this framework, two types of TPMs were generated: non-flood TPM and flood TPM. These TPMs are further used in the Monte Carlo simulation for generating a deterioration model of each cluster, based on the probability of flood reoccurrence. The results achieved in this study are explained thoroughly in chapter 4 of this document.

5.3 Limitations of this research

The proposed model of this study was conducted using limited data. It is believed that the increase in the quantity of data will improve the accuracy of this data-driven model.

In addition, the actual roughness data of flood-affected pavement sections were not available; therefore, a hypothetical flooding transition probability matrix was developed to understand the framework's gist. The framework is proposed by assuming that no maintenance work on the pavements will be conducted after the flooding event, and the magnitude of the flood was not considered in the framework.

5.4 Future work suggestion

In the future, when the predicted sections are observed, the pavement's actual roughness data can be compared statistically with the proposed model's data by applying a t-test. Also, the actual roughness data of flood-affected pavement sections must be collected for generating actual flooding TPM and then validate it with the field data.

5.5 Expected application of the study

The transportation departments can use this framework to predict pavement's future condition based on the probability of the flood-occurrence. This framework can be utilized when the actual flooding data is available to predict the pavement performance for post-flood times and prepare the maintenance budget.

REFERENCES

AASHTO, G. (1993). Guide for design of pavement structures. *American Association of State Highway and Transportation Officials, Washington, DC*.

Alashwal, H., El Halaby, M., Crouse, J. J., Abdalla, A., & Moustafa, A. A. (2019). The application of unsupervised clustering methods to Alzheimer's Disease. *Frontiers in computational neuroscience, 13*, 31.

Barnard, G. M. D. a. J. M. (1995). Hierarchical and non-Hierarchical Clustering. *Barnard Chemical Information Ltd.* Retrieved from https://www.daylight.com/meetings/mug96/barnard/E-MUG95.html

Battiato, G., Ame, E., & Wagner, T. (1994). *Description and implementation of RO. MA. for urban road and highway network maintenance.* Paper presented at the CONFERENCE PROCEEDINGS 1.

Butt, A. A., Shahin, M., Carpenter, S. H., & Carnahan, J. V. (1994). *Application of Markov process to pavement management systems at network level.* Paper presented at the 3rd international conference on managing pavements.

Chen, D., Hildreth, J., & Nicholas, T. (2014). Development of IRI limits and targets for network management and construction approval purposes. In: Research project funded through the North Carolina Department of ….

Chen, X., Dong, Q., Gu, X., & Mao, Q. (2019). Bayesian Analysis of Pavement Maintenance Failure Probability with Markov Chain Monte Carlo Simulation. *Journal of Transportation Engineering, Part B: Pavements, 145*(2). doi:10.1061/jpeodx.0000107

Chowdhury, S., Sultan, M., Chai, G., & Martin, T. (2016). *A Study on the Flood Affected Flexible Pavements in Australia*.

Drane, K. W., Gattis, D. R., Homewood, G. M., Parkinson, P. W., & Schwab, J. C. (2020). Hazard Mitigation Policy Guide. *American Planning Association*. Retrieved from https://planning-org-uploaded-media.s3.amazonaws.com/publication/download_pdf/hazard-mitigation-policy-guide.pdf

Elkins, G. E., Schmalzer, P. N., Thompson, T., & Simpson, A. (2003). *Long-term pavement performance information management system: Pavement performance database user reference guide*. Retrieved from https://www.fhwa.dot.gov/publications/research/infrastructure/pavements/ltpp/reports/03088/03088.pdf

FEMA. (2019). *Flood Risk Assessment*. Retrieved from https://training.fema.gov/hiedu/docs/fmc/chapter%204%20-%20flood%20risk%20assessment.pdf

Fu, G., & Devaraj, D. (2008). Methodology of homogeneous and non-homogeneous Markov Chains for modeling bridge element deterioration.

Gagniuc, P. A. (2017). *Markov chains: from theory to implementation and experimentation*: John Wiley & Sons.

George, K., Rajagopal, A., & Lim, L. (1989). Models for predicting pavement deterioration. *Transportation Research Record*(1215).

Gong, X., & Richman, M. B. (1995). On the application of cluster analysis to growing season precipitation data in North America east of the Rockies. *Journal of climate, 8*(4), 897-931.

Helali, K., Robson, M., Nicholson, R., & Bekheet, W. (2008). *Importance of a pavement management system in assessing pavement damage from natural disasters: a case study to assess the damage from Hurricanes Katrina and Rita in Jefferson Parish, Louisiana.* Paper presented at the 7th International Conference on Managing Pavement Assets.

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

Hudson, S., Chen, X., & Perrone, E. (1998). *Development and implementation of a network level pavement management system for Oklahoma.* Paper presented at the Proceedings of the 4th International Conference on Managing Pavements, Durban, South Africa. University of Pretoria, Pretoria, South Africa.

Jiang, Y., Saito, M., & Sinha, K. C. (1988). *Bridge performance prediction model using the Markov chain*.

Karimzadeh, A., Sabeti, S., & Shoghli, O. (2020a). Clustering-based Similarity Detection of Pavement Segments Considering Multiple Contributors to Deterioration. *ASCE Construction Research Congress (CRC), Tempe, Arizona*.

Karimzadeh, A., Sabeti, S., & Shoghli, O. (2020b). Optimal Clustering of Pavement Segments Using K-prototype Algorithm in a High-dimensional Mixed Feature Space. *Journal of Management in Engineering*.

Karimzadeh, A., & Shoghli, O. (2020). Predictive Analytics for Roadway Maintenance: A Review of Current Models, Challenges, and Opportunities. *Civil Engineering Journal, 6*(3), 602-625.

Karypis, M. S. G., Kumar, V., & Steinbach, M. (2000). *A comparison of document clustering techniques.* Paper presented at the TextMining Workshop at KDD2000 (May 2000).

Kenley, R., & Harfield, T. (2014). Managing road assets in times of multiple extreme flooding events. *Conference Proceedings*.

Kerali, H., & Snaith, M. (1992). NETCOM: the TRL visual condition model for road networks.

Khan, M. U., Mesbah, M., Ferreira, L., & Williams, D. J. (2014a). *Developing a new road deterioration model incorporating flooding.* Paper presented at the Proceedings of the Institution of Civil Engineers-Transport.

Khan, M. U., Mesbah, M., Ferreira, L., & Williams, D. J. (2014b). Development of road deterioration models incorporating flooding for optimum maintenance and rehabilitation strategies. *Road & Transport Research: A Journal of Australian and New Zealand Research and Practice, 23*(1), 3.

Kim, K.-D., Heo, J.-H., Nam, W.-S., & Shin, E.-W. (2004). Applicability of various clustering techniques for regional frequency analysis. In *Critical Transitions in Water and Environmental Resources Management* (pp. 1-9).

Kostuk, K. J. (2003). *Quantifying the Uncertainty Associated with Long Term Maintenance Contracts.* University of Saskatchewan,

Kulkarni, R. B. (1984). Dynamic decision model for a pavement management system. *Transportation Research Record, 997*, 11-18.

Kumar, A., & Swamy, A. (2015). Comparison of clustering approaches on temperature zones for pavement design. *Bituminous mixtures and pavements VI*.

Li, N. (1997). Development of a probabilistic based, integrated pavement management system.

Li, N., Huot, M., Xie, W., & Haas, R. (1995). *Applied Reliability And Markov Process Approch To Modelling Pavement Deterioration.* Paper presented at the XXTH World Road Congress, Montreal, 3-9 September 1995. individual papers presented under the auspices of the committees and working groups.

Li, Z. (2005). *A probabilistic and adaptive approach to modeling performance of pavement infrastructure.*

Lu, D., Tighe, S. L., & Xie, W.-C. (2017). *Pavement fragility modeling framework and build-in resilience strategies for flood hazard*. Retrieved from

Lu, D., Tighe, S. L., & Xie, W.-C. (2018). Impact of flood hazards on pavement performance. *International Journal of Pavement Engineering*, 1-7.

Lu, Q., & Harvey, J. T. (2006). Characterization of truck traffic in California for mechanistic–empirical design. *Transportation Research Record, 1945*(1), 61-72.

Madanat, S., Bulusu, S., & Mahmoud, A. (1995). Estimation of infrastructure distress initiation and progression models. *Journal of Infrastructure Systems, 1*(3), 146-150.

Mallick, R. B., Tao, M., Daniel, J. S., Jacobs, J., & Veeraragavan, A. (2017). Development of a methodology and a tool for the assessment of vulnerability of roadways to flood-induced damage. *Journal of Flood Risk Management, 10*(3), 301-313. doi:10.1111/jfr3.12135

Mills, L. (2010). *Hierarchical Markov chain Monte Carlo and pavement roughness model.* University of Delaware,

NCDOT. (2019). *North Carolina Department of Transportation Asset Management Plan*. Retrieved from https://connect.ncdot.gov/resources/Asset-Management/TAMP/Final%20TAMP%20-%20June%202019.pdf

NWS. (2019). *National Weather Service*. Retrieved from https://www.weather.gov/safety/flood-states-nc

Odoki, J., & Kerali, H. G. (2000). Analytical framework and model descriptions. *The Highway Development and Management Series, 4*.

Ortiz-García, 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.

Paerl, H. W., Hall, N. S., Hounshell, A. G., Luettich, R. A., Rossignol, K. L., Osburn, C. L., & Bales, J. (2019). Recent increase in catastrophic tropical cyclone flooding in coastal North Carolina, USA: Long-term observations suggest a regime shift. *Scientific reports, 9*(1), 1-9.

Panthi, K. (2009). A methodological framework for modeling pavement maintenance costs for projects with performance-based contracts.

Papagiannakis, A., Bracher, M., & Jackson, N. (2006). Utilizing clustering techniques in estimating traffic data input for pavement design. *Journal of Transportation Engineering, 132*(11), 872-879.

Parry, M., Parry, M. L., Canziani, O., Palutikof, J., Van der Linden, P., & Hanson, C. (2007). *Climate change 2007-impacts, adaptation and vulnerability: Working group II contribution to the fourth assessment report of the IPCC* (Vol. 4): Cambridge University Press.

Perera, R. W., Byrum, C., & Kohn, S. D. (1998). *Investigation of development of pavement roughness*. Retrieved from

Pierce, L. (2003). A probabilistic approach to creating pavement deterioration models. *Project Unpublished (M. Sc). Brigham Young University*.

Ranjith, S., Setunge, S., Gravina, R., & Venkatesan, S. (2011). Deterioration prediction of timber bridge elements using the Markov chain. *Journal of Performance of Constructed Facilities, 27*(3), 319-325.

Saha, P., Ksaibati, K., & Atadero, R. (2017). Developing pavement distress deterioration models for pavement management system using Markovian probabilistic process. *Advances in Civil Engineering, 2017*.

Sandra, A. K., & Sarkar, A. K. (2015). Application of fuzzy logic and clustering techniques for pavement maintenance. *Transportation Infrastructure Geotechnology, 2*(3), 103-119.

Sayers, M. W. (1998). *The little book of profiling: basic information about measuring and interpreting road profiles*. Retrieved from

Shamsabadi, S. S., Tari, Y. S. H., Birken, R., & Wang, M. (2014). *Deterioration forecasting in flexible pavements due to floods and snow storms*.

Sultana, M., Chai, G., Martin, T., & Chowdhury, S. (2015). *A study on the flood affected flexible pavements in Australia.* Paper presented at the 9th international conference on road and airfield pavement technology.

Sunitha, V., Veeraragavan, A., Srinivasan, K. K., & Mathew, S. (2012). Cluster-based pavement deterioration models for low-volume rural roads. *International Scholarly Research Notices, 2012*.

Thompson, P., Neumann, L., Miettinen, M., & Talvitie, A. (1987). *A micro-computer Markov dynamic programming system for pavement management in Finland.* Paper presented at the Proceedings of the Second North American Conference on Managing Pavements Toronto, Canada.

Tokuyama, H. (2012). Learning from Japan's ordeal. *Public roads, 75*(6).

USGS. (2019). *Flood and Recurrence Interval*. U.S. Geological Survey Circular Retrieved from https://www.usgs.gov/special-topic/water-science-school/science/floods-and-recurrence-intervals?qt-science_center_objects=0#qt-science_center_objects

Wang, K. C., Li, Q., Hall, K. D., Nguyen, V., & Xiao, D. X. (2011). Development of truck loading groups for the mechanistic-empirical pavement design guide. *Journal of Transportation Engineering, 137*(12), 855-862.

Wang, K. C., Zaniewski, J., & Way, G. (1994). Probabilistic behavior of pavements. *Journal of Transportation Engineering, 120*(3), 358-375.

Winfrey, R. (1935). *Statistical analysis of industrial property retirements* (Vol. 125): Iowa state college of agriculture and mechanic arts.

Xu, T., & Huang, X. (2012). Investigation into causes of in-place rutting in asphalt pavement. *Construction and Building Materials, 28*(1), 525-530.

Yan, Z., Yan, X., Huang, M., Xie, L., & Wang, Z. (2011). Based on the Improved K-Means Algorithm of Tianjin Port Traffic Flow Characteristic Analysis. In *ICTIS 2011: Multimodal Approach to Sustained Transportation System Development: Information, Technology, Implementation* (pp. 1863-1870).

Zhang, W., & Durango-Cohen, P. L. (2014). Explaining Heterogeneity in Pavement Deterioration: Clusterwise Linear Regression Model. *Journal of Infrastructure Systems, 20*(2). doi:10.1061/(asce)is.1943-555x.0000182

Zhang, Z., Wu, Z., Martinez, M., & Gaspard, K. (2008). Pavement structures damage caused by Hurricane Katrina flooding. *Journal of geotechnical and geoenvironmental engineering, 134*(5), 633-643.

APPENDIX

Appendix A1

Cumulative TPM cluster CL_1

IRI Range (m/km)	Cumulative IRI TPM Table CL_1					
	Cumulative TPM next year					
Current Year IRI Bucket	0.500 - 0.700	0.701 - 0.900	0.901 - 1.100	1.101 - 1.300	1.301 - 1.500	1.501 - 1.700
0.500 - 0.700	0.676	1.000	1	1	1	1
0.701 - 0.900	0	0.929	1.000	1	1	1
0.901 - 1.100	0	0	0.500	1.000	1	1
1.101 - 1.300	0	0	0	0.500	1.000	1
1.301 - 1.500	0	0	0	0	0	1.000
1.501 - 1.700	0	0	0	0	0	1.000

Appendix A2

Cumulative TPM cluster CL_2

IRI Range (m/km)	Cumulative Transition Probability Matrix CL_2									
	Cumulative TPM next year									
Current	0.500 - 0.750	0.751 - 1.000	1.001 - 1.250	1.251 - 1.500	1.501 - 1.750	1.751 - 2.000	2.001 - 2.250	2.251 - 2.500	2.501 - 2.750	2.751 - 3.000
0.500 - 0.750	0.6667	0.875	0.875	0.9167	1	1	1	1	1	1
0.751 - 1.000	0	0.68	0.96	1	1	1	1	1	1	1
1.001 - 1.250	0	0	0.82	0.98	0.98	1	1	1	1	1
1.251 - 1.500	0	0	0.0357	0.7857	1	1	1	1	1	1
1.501 - 1.750	0	0	0	0	0.8125	1	1	1	1	1
1.751 - 2.000	0	0	0	0	0	0.8333	1	1	1	1
2.001 - 2.250	0	0	0	0	0	0	0.6667	0.6667	1	1
2.251 - 2.500	0	0	0	0	0	0	0	0.6667	1	1
2.501 - 2.750	0	0	0	0	0	0	0	0	0.6667	1
2.751 - 3.000	0	0	0	0	0	0	0	0	0	1