Multiple random slope and fixed intercept linear regression models for pavement condition forecasting

Xiaojun Lin
University of Toledo

Follow this and additional works at: http://utdr.utoledo.edu/theses-dissertations

Recommended Citation
Lin, Xiaojun, "Multiple random slope and fixed intercept linear regression models for pavement condition forecasting" (2015). Theses and Dissertations. 2063.
http://utdr.utoledo.edu/theses-dissertations/2063

This Thesis is brought to you for free and open access by The University of Toledo Digital Repository. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of The University of Toledo Digital Repository. For more information, please see the repository's About page.
A Thesis

entitled

Multiple Random Slope and Fixed Intercept Linear Regression Models for Pavement Condition Forecasting

by

Xiaojun Lin

Submitted to the Graduate Faculty as partial fulfillment of the requirements for the Master of Science Degree in Engineering

Dr. Eddie Yein-Juin Chou, Committee Chair

Dr. Liangbo Hu, Committee Member

Dr. Bhuiyan Alam, Committee Member

Dr. Patricia R. Komuniecki, Dean
College of Graduate Studies

The University of Toledo

May 2015
An Abstract of

Multiple Random Slope and Fixed Intercept Linear Regression Models for Pavement Condition Forecasting

by

Xiaojun Lin

Submitted to the Graduate Faculty as partial fulfillment of the requirements for the Master of Science Degree in Engineering

The University of Toledo
May 2015

Pavement condition forecasting plays an important role in pavement management. Accurate predictions can help pavement managers in making optimal management plans which keeps the pavements in serviceable conditions over a specific period, as well as saves costs that are spent in the pavement maintenance and rehabilitation. In general, there are two types of prediction methods for pavement condition, which are probabilistic approach and deterministic approach. Probabilistic approach has its advantage in large scale pavement network’s overall condition predictions, it focuses on the whole pavement network condition forecasting and provides prediction results in distributions manner. Deterministic approach has its advantage in small pavement section scale predictions; it focuses on specific pavement sections condition forecasting and is able to provide the result of specific pavement section’s condition.

This paper attempts to develop a deterministic forecasting approach which not only utilizes the advantage of deterministic approach that is able to provide specific pavement section condition prediction, but also attempts to adopt the advantage of probabilistic approach, which considers the effects that might have universal influence on a pavement network’s overall condition. After model simulations, the approach finally obtained on
making pavement condition forecasting in this study is a random slope and fixed intercept linear regression approach. Each pavement section is assigned with a specific slope and specific intercept based on its categorical variable values and numerical variable values. The effectiveness of the obtained models is checked by comparing their predictions with the predictions through the existing prediction program ODOTPMIS.
This thesis is dedicated to my parents and my grandparents
Acknowledgements

This thesis would not have been finished without the support and encouragement from a lot of people. First of all, I would like to give thanks to my grandparents, although one of them is not along with the family anymore. I couldn’t have made the decision to pursue studies in a foreign country which is far away from my hometown without their unexpressed support. I am sincerely grateful to my grandparents’ understanding of me couldn’t fulfill filial piety as a grandson during these years, especially when they were not doing well. I also want to express thanks to my parents, who have been giving me countless suggestions and support on living in a foreign place.

Second, I would like to thank my advisor, Dr. Eddie Y. J. Chou, for his encouragement and patience, as well as lots of professional advice and instructions. I would also like to thank Dr. Qin Shao and Qing Qin, who gave me very helpful advice on statistics analysis. Besides, I also felt lucky to receive valuable help from Shuo Wang, who instructed me a lot in understanding the database and the using of PMIS and mapping software before he left the university.

Finally, I would like to thank my friends, roommates as well as my landlord. They gave me lots of advice on everyday life.
Table of Contents

Abstract .......................................................................................................................... iii

Acknowledgements ....................................................................................................... vi

Table of Contents .......................................................................................................... vii

List of Tables ................................................................................................................... x

List of Figures ................................................................................................................. xii

1 Introduction .................................................................................................................. 1

1.1 Objective of the Study ............................................................................................... 2

2 Literature Review ........................................................................................................ 4

2.1 Pavement Performance ............................................................................................. 4

2.1.1 Factors That Could Affect Pavement Performance ............................................. 6

2.2 Performance Prediction Models ................................................................................ 8

2.2.1 Deterministic Models ............................................................................................ 9

2.2.2 Probabilistic Models .............................................................................................. 14

2.2.3 Comparing Deterministic Models with Probabilistic Models ............................. 19
3 Methodology..................................................................................................................21

3.1 Data and Data Arrangement.......................................................................................21

3.1.1 Categorical Variable .............................................................................................23

3.1.2 Quantitative Variable ............................................................................................26

3.1.3 Data Arrangement ..................................................................................................31

3.2 Forecasting Model.......................................................................................................36

3.2.1 The Concept and Formulation Development .........................................................36

3.2.2 The Concept of Baseline Group ............................................................................42

3.2.3 The Use of Variables and Parameters ....................................................................43

3.3 Softwares Implementation ..........................................................................................45

3.3.1 Models Generating ................................................................................................45

3.3.2 Results Comparison ...............................................................................................46

4 Results and Findings.......................................................................................................47

4.1 Obtained Models ........................................................................................................47

4.1.1 Models and Interpretations ....................................................................................48

4.1.2 Models Selection, Further Modifications and Considerations ..............................54

4.1.3 Models through Statistics Consulting .....................................................................61

4.2 Forecasting Results Comparison ................................................................................63

4.2.1 An Introduction of ODOTPMIS Prediction Approaches .......................................63
List of Tables

2.1 Condition States Definitions

3.1 A Brief Summary of the Variables in Database

3.2 Comparison on the Data before and after Filtering

3.3 Coding of Two Dummy Regressors

3.4 Modification of Coding of Three Dummy Regressors

4.1 Coefficients of RSFI Models with Average Weather and Traffic Data

4.2 Coefficients of RSFI Models with Cumulative Weather and Traffic Data

4.3 VIF and Tolerance Values of Variables of Table 4.1

4.4 VIF and Tolerance Values of Variables of Table 4.2

4.5 Coefficients of Updated RSFI Models of Equation No. 10

4.6 Coefficients of Models of District No. 1 after Statistics Consulting

4.7 Weighted Correlations of Prediction Comparisons from 2013~2017

4.8 Strength of Correlation Categorization

A.1 Frequency Distribution of Average Annual Rainfall Amount

A.2 Frequency Distribution of Average Annual Temperature

A.3 Frequency Distribution of Average Annual Snowfall Amount
A.4 Frequency Distribution of Average ESAL ...............................................................84
A.5 Frequency Distribution of Thickness Added ..........................................................85
C.1 Coefficients of Models of District No. 2 after Statistics Consulting .......................89
C.2 Coefficients of Models of District No. 3 after Statistics Consulting .........................90
C.3 Coefficients of Models of District No. 4 after Statistics Consulting .........................90
C.4 Coefficients of Models of District No. 5 after Statistics Consulting .........................91
C.5 Coefficients of Models of District No. 6 after Statistics Consulting .........................91
C.6 Coefficients of Models of District No. 7 after Statistics Consulting .........................92
C.7 Coefficients of Models of District No. 8 after Statistics Consulting .........................92
C.8 Coefficients of Models of District No. 9 after Statistics Consulting .........................93
C.9 Coefficients of Models of District No. 10 after Statistics Consulting .......................93
C.10 Coefficients of Models of District No. 11 after Statistics Consulting ....................94
C.11 Coefficients of Models of District No. 12 after Statistics Consulting ....................94
List of Figures

2-1 Different Effect on Pavement Performance after Different Treatments ..................8

2-2 Three Types of Regression Approach .....................................................................13

2-3 An Example of Survivor Curve .............................................................................16

2-4 An Example of Markov Model Transition Matrix ..................................................17

3-1 A Map of ODOT District Divisions of 2014 ..............................................................26

A-1 Average Annual Rainfall Amount Distribution (in percentage) .............................81

A-2 Average Annual Temperature Distribution (in percentage) ..................................82

A-3 Average Annual Snowfall Amount Distribution (in percentage) ..........................83

A-4 Average ESAL Distribution (in percentage) ..........................................................84

A-5 Thickness Added Distribution (in percentage) .......................................................85
Chapter 1

Introduction

Pavement Management Systems (PMS) are being widely used throughout the North America to assist pavement managers in finding optimal management strategies which help maintain pavements in serviceable conditions over a specific period while keeping the maintenance and rehabilitation costs as low as possible. On one hand, a PMS is used on storing information, such as construction and maintenance histories or vehicle loadings on pavement sections. On the other hand, more importantly, the information that stored in PMS can also be used to evaluate the future conditions of pavement sections or a pavement network. If the predictions of the pavements’ future conditions are accurate, then the funds for constructions and maintenances can be more properly allocated on the projects that are in need the most.

Nowadays, strategies on pavement performance prediction and management can be categorized into mainly two types, network-level and project-level. Network-level usually classified pavements into groups and treat all the pavement sections within the group as a whole, predictions on the pavements future conditions are also based on the whole group’s
overall condition. For example, through a network-level prediction, it will be more likely to get results on the percentage of pavements in a group that will drop to a certain quality level. On the contrary, project-level prediction will focus on the forecasting down to each pavement section, and the obtained results will be specific to certain projects. Through project-level prediction, pavement manager will be able to “preview” the future condition of different pavement sections with a given year and a given pavement section ID (which correlates to the characteristics of the pavement section). By comparison, network-level prediction has the advantage on providing forecasting results in terms of a pavement group’s overall condition distributions, so that the costs on the maintenance, rehabilitation or re-constructions of a whole group can be generated without too much effort. In project-level prediction, each pavement section will be assigned for a model based on its characteristics such as pavement type and the exterior pressure such as weather or traffic loading that being imposed on it. The major advantage of project-level prediction is that it is easier for a pavement system manager to make comparison among different pavement sections’ future conditions, which will be more helpful for decision makers on establishing future treatment plans at project-level.

1.1 Objective of the Study

The objective of this study is to develop regression models for project-level pavement performance prediction. Traditionally, regression methods on pavement
performance prediction were usually applied through designating a single model for all the pavement sections within the same pavement family. This manner is simple and straightforward when applying, but usually overlook the variations created by differences (e.g., traffic loads, weather records) within the same pavement family. The deviations on forecasting through using single family regression model should not be ignored. Although designating specific model for each pavement section might eliminate the concerns of variations, the efforts required on applying specific model for each pavement section will be enormous and it is obviously impractical when taking all the pavement sections into account at the state level. This study explores a new method on developing regression models which on one hand keeps the number of regression models relatively small, while on the other hand takes the variations among different pavement sections into consideration in the process of formulating the models.
Chapter 2

Literature Review

2.1 Pavement Performance

Pavement performance, according to Lytton (1987), is a general term describing how pavement condition, or its ability to serve their intended functions, change with accumulating use. Since there isn’t a specific universal definition on pavement performance, or pavement condition, this term usually represent different meanings, depending upon how it is being measured or studied by different organizations or under different research objectives. As a result, different indices have been developed to distinguish those different measurements in terms of representing, or monitoring pavement performance.

Some monitoring measures emphasis on the riding quality of pavement, such as International Roughness Index (IRI), Profile Index (PI) and Present Serviceability Rating (PSR), these indices were developed by measures or calculations of pavement surface smoothness. Some monitoring measures give their emphasis on specific distress levels, for example the rutting level, cracking level or raveling level, these measures focus on one
type of distress index, as their damage levels will severely affect the pavement performance when compared with other types of distress indices. The third type of measurement indices can be concluded as a type of composite index, which represent the overall condition of pavement with regard to several kinds of distress indices, these type of indices are usually generated through a calculation of different distresses’ deducted values. Some of the common composite indices include Present Serviceability Index (PSI), Pavement Condition Index (PCI) or Pavement Condition Rating (PCR). PSI is based on various pavement measures such as slope variance, cracking, rutting, and patching. PCR and PCI are similar measurements which deduct the sum of a series of distress values from the scale 100, with 100 being the perfect condition while 0 as the worst. A fixed ranges of the composite indices values are usually applied in splitting the scale into several condition states because saying a pavement section is in fair condition is often more meaningful than providing a composite index value of 72 (Hedfi and Stephanos, 2001).

It should be noted that, these above indices are not being used exclusively; their values can also be used together on monitoring the pavement performance, or be used as a trigger of applying certain treatments upon pavements. For instance, some transportation departments will use a composite index and several distress indices to monitor their pavement performance. When the composite index value of a pavement section drop to a minimum allowance level, say 60 points, a treatment might be applied on the that pavement section. However, this is not always the case, when a certain distress index value,
such as cracking, of a pavement section drops to an intolerable level, even the overall composite index value still remains acceptable, certain treatment will be applied to that pavement section as well.

In the state of Ohio, PCR is a major value that being applied by The Ohio Department of Transportation as the pavement performance measuring index for its highway systems (Morse and Miller, 2004). A table is presented as follows indicating the condition states that are defined according to PCR values. Therefore, in this study, PCR will be used as the pavement performance index and all the performance of pavement segments or pavement sections mentioned in this paper will be directed to the corresponding PCR values.

<table>
<thead>
<tr>
<th>Index</th>
<th>Condition State</th>
<th>Values Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCR</td>
<td>Very Good</td>
<td>91 to 100</td>
</tr>
<tr>
<td></td>
<td>Good</td>
<td>76 to 90</td>
</tr>
<tr>
<td></td>
<td>Fair</td>
<td>66 to 75</td>
</tr>
<tr>
<td></td>
<td>Fair to Poor</td>
<td>56 to 65</td>
</tr>
<tr>
<td></td>
<td>Poor</td>
<td>41 to 55</td>
</tr>
<tr>
<td></td>
<td>Very Poor</td>
<td>0 to 40</td>
</tr>
</tbody>
</table>

(Adapted from ODOT, 2006)

2.1.1 Factors That Could Affect Pavement Performance

There are a number of factors that could affect pavement performance. Through previous studies, factors that are commonly considered include pavement materials, traffic
loads, climate data, and pavement thickness. Besides these factors, the treatment type that the pavement section received before the starting point of study data should draw attention as well. For example, in a case when two pavement sections have a similar current condition, the pavement section which just received a major type treatment usually has a larger chance to survive for a longer period than the pavement section which just only received a routine maintenance treatment. The difference of effect on the pavement performance after receiving different treatment types is shown in the following figure. Since some of the factors can create considerable variations on pavement performance, how to use different factors (e.g., which factors should be used as pavement group indictors and which factors should be used as performance prediction variables) needs great amount of consideration. Details about these factors and the ways on using these factors will be presented in the later chapters.
2.2 Performance Prediction Models

It is known to drivers and pavement engineers that, the pavement performance will change as the time goes by. The term deterioration is usually being adopted on describing this change. The reasons for deterioration are usually complicated; it could be reasons of traffic loads, the type of maintenance or rehabilitation that were previously selected, climatic influence, structural characteristics or even some effects that haven’t been discovered by engineers. Therefore, accurate deterioration prediction will help transportation agencies make proper management, such as adequate activity plans, or
budget allocation (Prozzi and Madanat, 2002). As it is mentioned in the introduction, there are mainly two types of performance prediction models, which are network-level and project-level, based on pavement management needs. By the approaches of modeling, it can also be classified as deterministic model and probabilistic model.

2.2.1 Deterministic Models

Deterministic models are usually being used on a database of smaller scale level, such as state level, district level or project level. According to different prediction objects and different situations, deterministic models can be further categorized into four types of deterministic models, which are Mechanistic models, Mechanistic-Empirical models, Regression models and Subjective Models.

2.2.1.1 Mechanistic Models

Mechanistic models employed in pavement research field are usually used for calculating pavement stress and deformation attributes in terms of traffic loadings or other pavement characteristics based on mechanical theories. These types of models were initially developed for pavement response analyses, which could contribute to pavement design. Typical examples of mechanistic models should probably trace back to the fatigue and fracture models that derived from the results of AASHO Road Test.

It should be noted that the calculated results from mechanistic models can be used
as input for empirical models.

### 2.2.1.2 Mechanistic-Empirical Models

Since mechanistic models are formed based on pavement response parameters, then the models that combined both field data and response parameters are called empirical-mechanistic models. In other words, mechanistic-empirical models are built based on both mechanistic theories and empirical findings. In general, the empirical base of the model usually includes time-series pavement condition data compiled on pavements exposed to different environmental and loading conditions.

Mechanistic-empirical approach has been employed in various pavement condition forecasting studies. Since the models are not strictly formed based upon existing physical theories or already proven findings such as mechanistic models, the formations of mechanistic-empirical models are usually represented in different forms, as well as the variables that are chosen for the models. The models adopted or obtained might be varied from study to study. In a pavement condition forecasting study by George K. P. et al. (1989) for 2,000 lane-miles of road in northern Mississippi, parameters such as modified structural number (SNC), age, thickness of surface and traffic loads are employed in forming regression models. However, it should be noted that within this study, different type of pavement is represented by different models which are not only distinguished from the value of regression coefficient, but also from the forming of regression model. This
example indicates that, even in the same study, the models that finally selected by the authors could be different when they are being used under different conditions.

2.2.1.3 Regression Models

Regression models are applied on determining the relationship between variables, the pavement condition or specific distress will usually be modeled as a regression function of other variables such as traffic loads, climatic conditions or structural characteristics. Since mechanistic models and mechanistic-empirical models are also built through making connections between variables, above two models can be treated as regression models as well. Regression approach can be used at project-level, as well as network-level. When it is employed at a large scale prediction, say at network-level, or under the condition that variations among the pavement sections cannot be ignored, a pre-defined classification (or grouping) of pavement sections according to their similar characteristics will usually be taken place. For the regression techniques, it can either be linear or non-linear. In order to improve the prediction accuracy, large number of historical data will generally be required on regression modeling.

For linear regression approaches, the fitting curves are usually straight lines. An example of linear regression models application can be found in the obsolete ODOT Pavement Design and Rehabilitation Manual Section 101, where the pavement sections under ODOT management were classified into fifteen pavement groups so that fifteen
linear regression models were generated for the pavement performance prediction. In another research by Yu J. et al. (2007), linear mixed effects model (LMEM) was proposed on forecasting flexible overlay pavements for ODOT’s urban and general pavement network.

For various non-linear regression approaches, they can be further categorized as polynomial (B-spline curve) and exponential (S-shaped curve) according to their fitting techniques. In a research of developing deterioration models for the primary highway network of the province of Alberta, Canada (Salem et al., 2003), sigmoidal prediction, which can be classified as a type of exponential regression approach, was employed on International Roughness Index (IRI) prediction for 8 defined pavement groups. The three types of regression approach that mentioned ahead are represented in the following figure.
2.2.1.4 Subjective Models

For the cases that do not have historical data or only have little data, to take an example, when a pavement section that is newly built or designed, the data such as its previous condition values and traffic loadings are not available. Subjective models, which are set up based upon experience and judgments from the engineers or experts of the field, can be adopted in performance prediction when a large proportion of pavement condition data that are required for common prediction approaches are vacant or invalid.
2.2.2 Probabilistic Models

Probabilistic models are usually being used on a large scale database, such as national level or state level. Instead of predicting the pavement performance of a specific pavement section at a given time as deterministic models do, probabilistic models are employed in predicting the probability distribution of pavement condition of an entire pavement network. Therefore, the forecasting results obtained through probabilistic approaches are more likely presented in ways that indicating how many pavement sections or the length of certain type of pavements that need to be maintained in a given year. This type of results can easily be used for budgeting a road-network’s yearly rehabilitation and maintenance costs. In addition, because of the large scale that is being forecasted, pavement sections of a road-network will generally be divided into various groups or families based on the similar characteristics of those pavement sections. The modeling techniques that are commonly being applied called Survivor Curves models and Markov prediction models.

2.2.2.1 Survivor Curves Models

A survivor curve is used on predicting the probability that the pavement will remain in a serviceable condition at a particular time without the need of receiving major maintenance or rehabilitation (Rao et al., 1994). In other words, it serves as a statistical method in calculating the pavement sections “life expectancy” and the “failure” of a
pavement section is determined by the point it needs to receive a major maintenance or rehabilitation (Gharaibeh and Darter, 2003).

As it is mentioned ahead, the “grouping” of pavement sections is an important step when applying probabilistic approach. This expression is also true for survival analysis when the pavement network is at state level. Pavement sections can be categorized into different families based on various characteristics. The “grouping” standards might vary case by case, according to different research objectives or the characteristics of different pavement networks. For example, the pavement families could be classified based on the locations of the pavements (when distinguished variations on climate data can be observed within the study area), or based on pavement overlay types, or even based on the thickness of the overlay of same type of pavement (when the variations on overlay thickness of same pavement type cannot be ignored).

After the “grouping” process is done, factors that used on forming the survive curve models needs to be determined. In general, age (number of years since last major maintenance) and ESAL (cumulative equivalent single axle loads since last major maintenance), and along with some regression coefficients are used on forming the survival models. Optimizations are used on finding the best fit coefficients for improving the accuracy of the models.

As for the result obtained through Survival analysis, percentage of the failure probability will be represented. With the already known total length of a certain pavement
type, or a pavement family in a road-network, the length of that specific pavement type in
the road-network that need to be maintained in a given year can be calculated through
multiplying the failure percentage by the total length. An example of Survivor Curve is
shown as the following figure.

![Figure 2-3 An Example of Survivor Curve](image)

2.2.2.2 Markov Models

The Markov prediction model is based on a stochastic process which utilizes
transition matrix in calculating the probability that an object or system that will “transit”
from one condition state to another condition state (or will remain in the same condition
state) within a one time period. As for being applied in pavement condition forecasting,
Markov models can be used on computing the probability of pavement (usually a pavement group or pavement network) at a certain condition state that will “transit” to another condition state, or will remain at the same condition state (Ortiz-García et al., 2006). For example, assuming all the pavement sections are currently in the good condition within a pavement family. Then after two years, some of the pavement sections, say 32% of them might drop to poor condition, while other sections, which are 68% of all the sections, still remain as good condition. An example of transition matrix that is used in Markov model is shown as follows.

$$ P = \begin{bmatrix}
\text{Very Good} & \text{Good} & \text{Medium} & \text{Poor} & \text{Severe} & \text{Very Severe} \\
0.72 & 0.26 & 0.02 & 0 & 0 & 0 \\
0 & 0.93 & 0.06 & 0.01 & 0 & 0 \\
0 & 0 & 0.88 & 0.09 & 0.03 & 0 \\
0 & 0 & 0 & 0.78 & 0.16 & 0.06 \\
0 & 0 & 0 & 0 & 0.98 & 0.02 \\
0 & 0 & 0 & 0 & 0 & 1.00 
\end{bmatrix} $$

**Figure 2-4 An Example of Markov Model Transition Matrix**

In general, there are three premises when using Markov models. They are listed as follows:

1. The time parameter that used in “transition” process should be a discrete time period, say one year. If the time period of “transition” process is not constant, Markov models cannot be used for analysis.

2. All the condition states, for instance, “perfect”, “great”, “fair”, “poor” and
“very poor” conditions, of an object or a system should be counted and defined when using Markov models.

3. It is assumed that the future condition of a system or an object during the “transition” process only depends on its current condition, which means the effects by those past conditions will be ignored. This assumption is also called state independence assumption.

The assumptions listed above are set up to eliminate the complexity on model computations and make the decision-making process simple (G. Morcous, 2006). However, in reality, some of the assumptions cannot be met. Take the first premise; it is impractical to ensure that the time period is a constant value, some pavement sections are inevitable to be examined earlier while others are examined later within the same year. It is possible that some pavement sections drop from one condition to another condition during this time difference. Thus the accuracy of the data could be questionable. The third assumption, which assuming that the future condition of a system or an object only depends on its current condition, can also be challenged when it is adopted for pavement condition forecasting. Although the assumptions of employing Markov models might not be satisfied in the field of pavement condition forecasting, a research by G. Morcous (2006), which adopts Markov models in bridge deck systems performance prediction, shows that the state independence assumption can be accepted with a 95% level of confidence, and it is reasonable for bridge network level analysis.
As for the results obtained through Markov models, it is similar to the results that obtained using Survivor Curves model. For each pavement group or pavement network, a series of percentages calculated through matrix computations indicates the pavement section occupancy rates of each condition state at a given year. With the percentages and the total pavement length, the pavement lengths of each condition can be computed thus the maintenance budgets can be estimated as well.

### 2.2.3 Comparing Deterministic Models with Probabilistic Models

Which type of model is better usually varies dependent upon the objectives or the needs of a study or a pavement management program. Probabilistic model is usually a better choice when handling network-level programs such as forecasting the total length of pavement in a road network that requires certain maintenance in the future, or making future M&R activities budget for all the pavement within a network. However, when the situation comes down to making performance prediction for a specific pavement section at the project level, a deterministic model is usually a better choice if a series of variables of the pavement can be provided. For example, by using deterministic model, a decision maker can get to know which pavement sections might deteriorate rapider than the others so that precautionary M&R plans can be set up against those pavement sections, and future cost of those projects can be evaluated as well. With respect to different situations, both deterministic models and probabilistic models can be applied for pavement performance
forecasting (Hedfi and Stephanos, 2001). In terms of prediction accuracy, since deterministic approaches are usually applied at the project level predictions while probabilistic approaches are usually used at the network level predictions, the deviations between the observed data and predicted data using deterministic approach are often relatively larger because overestimations can be offset by underestimations at the network level when using probabilistic approach.

It is also worth mentioning that it is not a fixed concept that deterministic models can only be used when applying deterministic approach, or probabilistic models can only be used when applying probabilistic approach in a research. These two major prediction model types can be used together, or be “transformed” from one type into another type as well. For example, regression model can be deduced through Markov model after certain conversions, so that the deduced regression model can also be used at project-level prediction.
Chapter 3

Methodology

3.1 Data and Data Arrangement

The data used in this study are obtained from the database of ODOTPMIS, which is a software package designed for the PMS of ODOT. According to an annual report of 2013 (ODOT, 2013), ODOT maintains 49,250 miles interstate, U.S. and state routes within the state of Ohio, and approximately 70% of its expenditures were spent on highway construction in the year of 2013. The database contains a vast number of data which records information that related to pavement sections which are under management of ODOT. The information of pavements recorded in the database ranged from inventory information, traffic loads, pavement performance, maintenance treatments, to weather records, etc. In view of data analysis, the information of pavements is the variables of those corresponding pavement sections. All the information of the pavement sections can be classified into two categories: categorical variables and quantitative variables. Each variable and its data type will be further expounded in the subsequent paragraphs.

As it is mentioned earlier, pavements that receive different maintenance and
rehabilitation treatments will display different deterioration patterns. Therefore, in this study, the data of pavement sections that are being studied are defined as the data that can meet following criterions: (1) The data of pavement sections should be recorded starting with a minor treatment activity; (2) The data of pavement sections should be ended with either a minor treatment activity, or an unknown treatment activity which cause a more than 10 points improvement on the pavement performance, or with no treatment recorded (meaning that pavement section still have not received any non-preventive treatment since last minor treatment). To put the criterions in another way, this study focus on the “data truncations” starting with a minor treatment activity and ending with either a minor treatment activity, or an unknown treatment activity which raises the pavement performance over 10 points in the following year, or without any treatment activity. The first consideration by setting up these criterions is that major treatment activities are seldom applied. The second consideration is that preventive treatment activities are usually yearly applied, and these treatments will not greatly improve the pavement performance. The “truncation” process can be easily operated by using the software ODOTPMIS (Ohio Department of Transportation Pavement Management Information System), which is a PMS software developed by the Infrastructure Information System Laboratory at the University of Toledo.
3.1.1 Categorical Variable

A categorical variable is usually an independent or predictor variable that records categorical condition of the response (Indiana University, 2014). Categorical variables are often used as indicators that represent the groups or categories the responses or individuals belong to. Commonly used categorical variables include gender (male or female), blood type or marital status. Sometimes, a categorical variable can also be represented in a numerical way. For example the district divisions (Dist. 1, Dist. 2 … Dist.12) in this study, the numbers being used in district divisions do not have any numerical meaning and cannot be treated as a type of measure. Within ODOT’s database, categorical variables include inventory data such as route number, county name, and pavement type, etc. and activities codes. However, not every categorical variable will be used in modeling. Besides the previous mentioned district division, other categorical variables that will be used in this study are pavement type and pavement system.

3.1.1.1 Pavement Type

In ODOT’s database, there are majorly 4 types of pavements in terms of their overlay material proportions or structure characteristics. The 4 pavement types are Joint-Reinforced Concrete (JRC), Continuously-Reinforced Concrete (CRC), Flexible (FLEX) and Composite (COMP). Both JRC and CRC pavements use concrete as their major overlay materials, they can also be classified as rigid pavement. The difference
between JRC and CRC is that JRC uses both reinforcing steel and contraction joints as means of crack control, while CRC only uses reinforcing steel. Flexible pavements use hot mix asphalt (HMA) as its main overlay materials, they are called “flexible” because asphalt bends to accommodate traffic loads. Composite pavements are designed as asphalt overlays on top of concrete layers, which is the reason why they are called “composite”.

Since different types of pavement has different overlay material proportions and structures, distresses developed on different types of pavement can be revealed in a different fashion as well as the deteriorations of pavement performance.

3.1.1.2 Pavement System

In order to have a better management on pavement rehabilitation, ODOT divided the state highway system into three policy systems, which are priority, urban and general systems. The priority system is mainly made up of Interstates and other rural, multi-lane divided highways. The general system is mainly made up of rural, two-lane US and State routes. The urban system is made of those US and State routes within municipalities with more than 5,000 populations. Currently, ODOT is majorly responsible for pavement rehabilitation on priority and general system routes. This means that pavement classified as priority and general system may receive different levels of maintenance and rehabilitation when compared with those pavements that classified as urban system.
3.1.1.3 District Division

ODOT divided the state of Ohio into 12 district divisions in the purpose of pavement management, and every district has its own funding and develops its own M&R strategies in maintaining their pavements quality according to ODOT’s requirements. Within the same district, the materials used in pavement constructions will usually come from the same source, so that a better quality material may results in a better pavement performance. This means that, same type of pavement in one district may generally have a better performance than another district due to their different material sources.

Besides different material sources, climate effect was also taken into account in district division. For example, District 12 is the district that usually receives the largest amount of snowfall every year. Therefore, average annual PCR decreasing rate is usually higher in District 12. The district divisions can be viewed in Figure 3.1.
3.1.2 Quantitative Variable

A quantitative variable is a variable that takes numerical values and arithmetic with those values makes sense (UC Berkeley, 2014). Quantitative variables are often used in representing measurable attributes of the responses or individuals, and they typically have measurement units. Commonly used quantitative variables are temperature, age,
population and income, etc. In regard to the data of this study, besides PCR values that will be used as the dependent variable in the modeling, other quantitative variables that will be used are age, temperature, the amount of snowfall, the amount of rainfall, thickness added and traffic loadings. Illustrations of numerical values distribution of quantitative variables “Temperature records”, “Snowfall records”, “Rainfall records”, “Thickness Added” and “Average ESAL” are presented in Appendix A.

3.1.2.1 Age

The variable age in this study does not represent the ages of pavement sections since they were initially built or constructed. An amount of age represents the years have passed since a pavement section received its last minor treatment activity. For example, a pavement section which receives a minor treatment this year will have the age of 1 in the next year.

3.1.2.2 Temperature

A temperature variable records the annual temperature (in Fahrenheit) of the county that a pavement section is located in. In the database, annual temperatures are being recorded at county level. An ODOT divided district is usually consisted of several Ohio counties. In addition, since it is annual temperatures that are being recorded, it means that the temperature a pavement section encounters with will be assumed to be the
same in each year when temperature variable is included in the forecasting model. The annual temperature records of all the Ohio counties in the current database range from 47 °F to 56 °F. According to Huang (2004), temperature has influence on the elastic and viscoelastic properties of Hot Mix Asphalt pavement.

3.1.2.3 Snowfall and Rainfall

Just as the temperature records, the amount of snowfall (in inches) and rainfall (in inches) are being recorded in the annual way within the database. They are also being recorded at county level as the temperature variable. The annual precipitation records range from 31 inches to 49 inches. For annual snowfall records, the variations among all the Ohio counties are relatively large when compared with the temperature and rainfall variations. Most counties have the annual snowfall around 15 inches to 38 inches, while there are eight counties have the records over 40 inches, and five out of these eight counties have the records range from 60 inches to 99 inches. A larger amount of snowfall records may suggests that those influenced counties have to drop much more salts than other counties on their pavements during every winter, which could greatly increase their pavements’ deterioration rates. As for the influence of rainfall, a study on the pavement of Iowa by Breakah T. M. et al. (2011) suggests that moisture has a significant effect on the rutting distress of pavement overlay that contain asphalt.
3.1.2.4 Thickness Added

According to a previous Federal Highway Administration (FHWA) report, pavement overlay thickness plays an important role in the pavement performance. In general, thicker overlay pavements perform better than thin overlay pavements in terms of deterioration rate and serviceable duration. The interactions between overlay thickness and pavement type, traffic loading or snowfall amount in the roles of pavement performance may also be noticed. Unfortunately, in ODOT’s database, overlay thickness data is not available. There is only a “thickness added” variable, which means the new overlay’s thickness that added on the pavement’s base layer (after the previous overlay is fully removed) in rehabilitation or reconstruction. Therefore, “thickness added” might be treated as a form of overlay thickness as well. However, there is still a deficiency in thickness added data. For example, in a pavement section’s rehabilitation, some part of the section’s overlay might still in a good shape while other parts don’t. In this type of situation, for the purpose of saving costs, those parts that have a doing well overlay might have their surface layers kept or only be partially removed so that the “thickness added” value might not be able to correctly represent the thickness added from the top of the base layer. The actual thickness added on that specific part could be 0 or less than the thickness added values that applied on the parts whose overlays are in a bad shape.
3.1.2.5 Traffic Loads

In the original ODOT’s database, there are three types of traffic loads data: average daily traffic (ADT), truck average daily traffic (Truck ADT) and equivalent single axle load (ESAL). As the names suggest, ADT represents the average daily traffic volume of corresponding pavements, while Truck ADT represents the average daily truck traffic volume of corresponding pavements. ESAL is defined as a unit that represents the amount of pavement damage caused by one axle or a group of axles, dependent upon the loaded weight of the axle or axles. The concept of ESAL is developed from the data of American Association of State Highway Officials (ASSHO) Road Test; and the reference axle load is an 18,000 lb. single axle of dual tires (TxDOT, 2005).

By making a comparison between these three types of traffic data. It can be found that ADT overlook the difference in causing damage on pavement between a private car and a semi-truck. This difference should not be neglected when a pavement section is dominant by heavy vehicles. Truck ADT somehow ignores the damage effect caused by private cars, especially when a pavement is seldom used by heavy vehicles but frequently used by light vehicles. ESAL is an index that makes more sense because it counts the damage caused by any type of vehicles operating on the pavement. Therefore, in this study, ESAL will be used as a quantitative predictor on performance model’s forming trials.

It should also be noticed that it is the Average ESAL that is being used in this study,
which means it is assumed that a pavement section always has the same ESAL value every year. Just the same way as the climate data (temperature, snowfall and rainfall) are being applied in this study.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCR</td>
<td>Quantitative</td>
<td>Pavement performance index, used as a dependent variable</td>
</tr>
<tr>
<td>District</td>
<td>Categorical</td>
<td>12 district divisions defined by ODOT (OH01 = District 1, etc.)</td>
</tr>
<tr>
<td>Age</td>
<td>Quantitative</td>
<td>Number of years after the pavement section has received its last minor treatment</td>
</tr>
<tr>
<td>Pavement Type</td>
<td>Categorical</td>
<td>4 types of pavement used in ODOT pavement design</td>
</tr>
<tr>
<td>Pavement System</td>
<td>Categorical</td>
<td>3 different pavement systems defined by ODOT</td>
</tr>
<tr>
<td>AvgESAL</td>
<td>Quantitative</td>
<td>Annual average equivalent single axle load operating on the pavement section</td>
</tr>
<tr>
<td>Thickness Added</td>
<td>Quantitative</td>
<td>Overlay thickness added on top of base layer in the last minor treatment</td>
</tr>
<tr>
<td>Temperature</td>
<td>Quantitative</td>
<td>Annual average temperature (in Fahrenheit)</td>
</tr>
<tr>
<td>Rainfall Amount</td>
<td>Quantitative</td>
<td>Annual average rainfall amount (in inches)</td>
</tr>
<tr>
<td>Snowfall Amount</td>
<td>Quantitative</td>
<td>Annual average snowfall amount (in inches)</td>
</tr>
</tbody>
</table>

### 3.1.3 Data Arrangement

Although settings with regard to the pavement treatment activities are applied while obtaining database through PMIS, it doesn’t mean the data can be taken into
analysis directly. There are still two major deficiencies in the obtained database: (1) There are some errors due to the database itself; (2) Obtained data does not compatible with the form that the analysis software required. Since the data obtained through PMIS can be stored as a Microsoft Excel format, the second type deficiency can be easily solved with the help of Excel VBA Programming. However, for solving the first type, it requires some rules to be established and then have those defined “abnormal” data removed. In summary, there are three types of abnormality, which are: (a) Absence of PCR data; (b) Absence of variables data; (c) Abnormal PCR values. Details on the “abnormality” and their correspondent removing rules will be presented in the following paragraphs.

3.1.3.1 Absence of PCR Data

As the name of this erroneous type reveals, the lack of PCR data of some pavement sections’ records is the reason for this erroneous type. With no PCR data on a pavement section, the deterioration trend cannot be displayed as well as analyzed. Therefore, in this condition, the data of those pavement sections will be removed from the obtained database.

3.1.3.2 Absence of Variables Data

Just as the previous erroneous type, the lack of certain variables or all the variables data of some pavement sections records is the reason for this erroneous type.
Since the variables are used on finding out the relationships between pavement performance and pavement variables, the absence of variables data creates hindrance in data analysis process. As a result, any pavement section record that doesn’t contain a full set of variables data will be fully removed from the obtained database.

3.1.3.3 Abnormal PCR Values

Abnormal PCR values situations in this study is defined as the situation that the pavement performance do not follow a usual trend. Since this study is focusing on the general deterioration trends of pavement sections, those unusual trends should be removed before analyzing. Abnormal PCR values situations can be further classified into three different situations: (1) Rapid drops on PCR values; (2) Unusual raise on PCR values; (3) PCR value remains the same for “too long”. These three different types of situations will be discussed respectively as follows.

A declination on PCR values year after year is usually being expected when studying the trends of pavement deterioration. However, a rapid drop, for example, from a perfect condition drops to poor condition in just one year, is unusual. One possible explanation is, the pavement section received unusual amount of traffic loads especially heavy vehicles, or encountered with unusual amount of severe weather conditions in the past year. Since the case is unexpected, the corresponding PCR values should not be taken into analysis. In this study, a “rapid drop” is defined as the condition that the PCR
value of a pavement section decreases over 10 points in the following year. When there is a “rapid drop”, the PCR records will be cut off from the rapidly decreasing year. For instance, a pavement section has a PCR records as 98, 96, 93, 91, 80, etc. The PCR records in this case will be removed from the value of 80 till the end. After filtration, the remaining PCR records will be 98, 96, 93 and 91.

Contrary to the “rapid drop” situation, there exists an “unusual raise” situation. An “unusual raise” on PCR values in this study is defined as the situation that the PCR value of a pavement section increases over 3 points in the year followed. In general, the PCR value of a pavement section will decrease as time goes by when no greater than preventive treatment activity is applied in the past year. Sometimes, a preventive treatment might slightly improve the pavement condition or hold the pavement condition for a short time. However, a significantly improvement on the PCR value is usually a result of minor or major treatment. The reason for the existence of an “unusual raise” might be the result of unrecorded minor or major treatment activities. When there is an “unusual raise”, the PCR records will be cut off from the unusual increasing year till the end.

Besides “rapid drop” and “unusual raise” situations, there is also a situation that the PCR value of a pavement section remains the same for “too long”. The PCR value of a pavement section might remain for a year or two after a minor or major treatment since the pavement is still in a perfect condition. In this study, “too long” is defined as the
situation that the PCR value of a pavement section remains the same for over 3 years; this
type of abnormal situation is probably a result of continuous unrecorded preventive or
minor treatments on the pavement section. When the PCR values remains the same for
“too long”, the PCR records will be cut off from the third year of the identical PCR value
till the end. For example, a pavement section has a series of PCR records as 96, 95, 95,
93, 93, 93, 92, etc. In this case, the PCR records will be excluded from the third 93 and
the PCR records after filtering will be 96, 95, 95, 93, and 93.

3.1.3.4 The Amount of Data before and after Filtering

The sample size is always an important feature on data analysis. If the sample size
is relatively small when compared with the group size that is being studied, the obtained
analysis results might not be convincing because the representativeness of the small
sample size can be questionable. Therefore, it is meaningful to make an illustration on the
original data (the data obtained through PMIS) and the data after the previous filtering
process. The comparison on the data amount can be viewed through Table 3.2 in the
following page.

It can be observed from the tables that, before the filtering process, there are
totally 19,466 pavement sections and 148,618 PCR records. While after filtering, there
are totally 18,161 pavement sections and 111,237 PCR records remained. Roughly 93.3%
of pavement sections and 74.8% of PCR records are kept from the original dataset.
Table 3.2 Comparison on the Data before and after Filtering

<table>
<thead>
<tr>
<th>District</th>
<th>Sections before filtering</th>
<th>Sections after filtering</th>
<th>Percentage remains</th>
<th>PCRs before filtering</th>
<th>PCRs after filtering</th>
<th>Percentage remains</th>
</tr>
</thead>
<tbody>
<tr>
<td>01</td>
<td>1,449</td>
<td>1,337</td>
<td>92.27%</td>
<td>10,166</td>
<td>8,146</td>
<td>80.13%</td>
</tr>
<tr>
<td>02</td>
<td>1,361</td>
<td>1,235</td>
<td>90.74%</td>
<td>10,707</td>
<td>8,296</td>
<td>77.48%</td>
</tr>
<tr>
<td>03</td>
<td>1,977</td>
<td>1,812</td>
<td>91.65%</td>
<td>15,443</td>
<td>11,897</td>
<td>77.04%</td>
</tr>
<tr>
<td>04</td>
<td>1,743</td>
<td>1,692</td>
<td>97.07%</td>
<td>12,175</td>
<td>9,615</td>
<td>78.97%</td>
</tr>
<tr>
<td>05</td>
<td>1,696</td>
<td>1,550</td>
<td>91.39%</td>
<td>13,366</td>
<td>10,506</td>
<td>78.60%</td>
</tr>
<tr>
<td>06</td>
<td>2,097</td>
<td>1,953</td>
<td>93.13%</td>
<td>16,158</td>
<td>11,971</td>
<td>74.09%</td>
</tr>
<tr>
<td>07</td>
<td>2,477</td>
<td>2,276</td>
<td>91.89%</td>
<td>17,382</td>
<td>12,812</td>
<td>73.71%</td>
</tr>
<tr>
<td>08</td>
<td>1,624</td>
<td>1,520</td>
<td>93.60%</td>
<td>13,954</td>
<td>9,483</td>
<td>67.96%</td>
</tr>
<tr>
<td>09</td>
<td>1,067</td>
<td>1,037</td>
<td>97.19%</td>
<td>8,558</td>
<td>6,285</td>
<td>73.44%</td>
</tr>
<tr>
<td>10</td>
<td>1,338</td>
<td>1,274</td>
<td>95.22%</td>
<td>10,589</td>
<td>7,859</td>
<td>74.22%</td>
</tr>
<tr>
<td>11</td>
<td>1,588</td>
<td>1,483</td>
<td>93.39%</td>
<td>11,828</td>
<td>8,622</td>
<td>72.89%</td>
</tr>
<tr>
<td>12</td>
<td>1,049</td>
<td>992</td>
<td>94.57%</td>
<td>8,292</td>
<td>5,745</td>
<td>69.28%</td>
</tr>
<tr>
<td>Sum Up</td>
<td>19,466</td>
<td>18,161</td>
<td>93.30%</td>
<td>148,618</td>
<td>111,237</td>
<td>74.85%</td>
</tr>
</tbody>
</table>

3.2 Forecasting Model

As it is introduced in chapter 1, this study is going to use regression approach for pavement performance modeling. With the help of plentiful data that is available in this case, using regression approach to study the relationships between pavement performance and its corresponding characteristics is not unusual.

3.2.1 The Concept and Formulation Development

The basic concept of forecasting model of this study is shown as follows:

\[ PCR_i = PCR_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + ... + \epsilon \]  

(1)

Where:
$PCR_t = $ Predicted PCR value at age $t$,

$PCR_0 = $ Initial PCR value calculated through modeling,

$X_1, X_2, \text{etc.} = \text{Variable 1 and variable 2, etc.,}$

$\beta_1, \beta_2, \text{etc.} = \text{Co-efficient of variable 1, variable 2, etc., and}$

$\epsilon = \text{the error of predicted value.}$

As those previous researches applied, the forecasting model of this study will also adopt the idea that the predicted PCR value is a calculation of an initial PCR value minus the effects of various variables. Plus signs are used in the model because the values of coefficients can be negative. Since the dependent variable is the PCR value at age $t$, the variable “age” will inevitably be a variable that used in the model. Bearing that in mind, when a series of predicted PCR values are plot in 2-dimensional plane. The coefficient value of variable “age” will become the slope (deterioration rate) of the deterioration curve. Therefore, the above model can be expressed in the following way as well:

$$PCR_t = PCR_0 + \beta_1 \times Age + \beta_2 X_2 + \beta_3 X_3 + \ldots + \epsilon$$

(2)

It is introduced in the previous part of this chapter that, all the variables that are going to be used in this study can be classified into categorical variables and quantitative variables. Since categorical variables do not have any numerical meaning, their uses in a regression model are limited. For this study, categorical variables can be used either as weights added/subtracted to the predicted value or as parts of weights of other quantitative variables. According to the previous researches, where pavements are
classified into different groups using categorical variables as indicators, it makes sense to treat variable “age” as the quantitative variable that is under the influence of some of those categorical variables. In other words, the deterioration rate of a pavement section is dependent upon its categorical variables such as pavement type, the district it is located at, or the pavement system it belongs to. Therefore, three different models can be derived from the above model.

3.2.1.1 Fixed Slope and Random Intercept (FSRI)

When categorical variables are only used as weights added/subtracted to the predicted value (predicted PCR), the model is presented as follows:

\[
P CR_i = PCR_0 + \beta_1 \cdot Age + \beta_2 \cdot X_2 + \ldots + \delta_1 \cdot D_1 + \delta_2 \cdot D_2 + \ldots + \epsilon
\]  

(3)

The above model can be illustrated through the following example. In the above equation, assuming that a three-categories (which are A, B and C) categorical variable is included. The categorical variable is represented by \( \delta_1 \cdot D_1 \) and \( \delta_2 \cdot D_2 \), which is consisted of two introduced dummy regressors. The coding of dummy regressors is shown in the Table 3.3 below. When the pavement section belongs to category A, \( \delta_1 \) and \( \delta_2 \) is 0, then nothing is added to the predicted value, category A in this example is served as the baseline; when the pavement section belongs to category B, \( \delta_1 \) is 1, \( \delta_2 \) is 0, then \( D_1 \) is added to the predicted value; when the pavement section belongs to category C, \( \delta_1 \) is 0, \( \delta_2 \) is 1, then \( D_2 \) is added to the predicted value. It can be observed that when a categorical variable is treated as weight
added/subtracted to the predicted value, different category (under a categorical variable) results in difference in intercept of the model.

Table 3.3 Coding of Two Dummy Regressors

<table>
<thead>
<tr>
<th>Category</th>
<th>$D_1$</th>
<th>$D_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>B</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>C</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

3.2.1.2 Random Slope and Fixed Intercept (RSFI)

If categorical variables are used as weights that affect the deterioration rates, the model can be expressed as follows:

$$PCR_i = PCR_0 + (\delta_1 D_1 + \delta_2 D_2 + ...)*\text{Age} + \beta_2 X_2 + \beta_3 X_3 + ... + \epsilon$$  \hspace{1cm} (4)

In the above equation, it can be observed that categorical variables are treated as weights that added or subtracted to the deterioration rate. However, there should be a small modification on the coding of dummy regressors. In the previous example, the coding of category A is always 0 no matter which category the pavement section is belongs to. If the same manner is adopted, the deterioration rate will be 0 when the pavement section is belong to category A and the variable “age” in this situation is ineffective. The modification on the coding of dummy regressors is shown as the Table 3.4 below. It can be observed that three dummy regressors are coded instead of two. In order to avoid the variable “age” becoming ineffective, at least one categorical variable should be coded in this way.
### Table 3.4 Modification of Coding of Three Dummy Regressors

<table>
<thead>
<tr>
<th>Category</th>
<th>$D_1$</th>
<th>$D_2$</th>
<th>$D_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>B</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>C</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

#### 3.2.1.3 Random Slope and Random Intercept (RSRI)

If categorical variables are used not only as weights that affect the deterioration rates but also as weights that added, or subtracted to the predicted value, the model can be expressed as follows:

$$PCR_i = PCR_0 + (\delta_1 D_1 + \delta_2 D_2 + \ldots) \cdot Age + \beta_2 X_2 + \beta_3 X_3 + \ldots + \mu_1 E_1 + \mu_2 E_2 + \ldots + \epsilon$$  \hspace{1cm} (5)

Where:

- $\mu_1, \mu_2$ = Another sets of Boolean values for another categorical variable, which just work like $\delta_1$ and $\delta_2$,
- $E_1, E_2$ = Another sets of dummy variable values for another categorical variable, which just work like $D_1$ and $D_2$.

#### 3.2.1.4 Models through Statistics Consulting

Since the simulation of pavement performance prediction models is mainly about analyzing the pavement performance database. A statistics consulting was held during the process of this study, by the help of the professor and students from the Department of Mathematics and Statistics of the University of Toledo. Advices on the modeling were
provided according to the previous proposed concepts. The advice can be concluded as follows: (1) Generating models under district division level, (2) Introducing interactions between variables into modeling. As a result, another set of prediction models were also simulated according to the advice from the viewpoints of statistics profession.

For generating models under district division level, the reason of this advice is that, in the original proposed models, the effects (coefficient values) of quantitative variables (traffic loads, thickness added and weather records) and some categorical variables (pavement type and pavement system) are universal throughout the whole state. This assumption might lower the accuracy of the prediction models. Simulating models under district division level can be accomplished by removing the district division variable from the proposed models. Therefore, the quantitative variables as well as the categorical variables of each district would be assigned with unique coefficient values, according to its district division.

For introducing interactions between variables into the models, the reason is that, cooperative effects between the predictors of pavement performance that might be ignored in the proposed models should be explored. However, the limitation and rules on the use, or selection of interaction variables should also be satisfied, before confirming the effects of interactions, as well as the effects of main predictors. One of the crucial points is the change on the Adjusted R-squared value (Adjusted $R^2$). Adjusted R-squared is an unbiased measure of the proportion of variance explained, taking into account the
sample size and number of variables (Duke University, 2015). In other words, R-squared might increase as redundant or insufficient variables are added into the models while Adjusted R-squared might remain the same. Adjusted R-squared value works as an indicator of whether an added variable would improve the model. Another crucial point on determining whether a variable or interaction should be included in the model is Variance Inflation Factor (VIF) or Tolerance value. The VIF is 1/Tolerance value; they work as indicators of the collinearity impact among the variables in the models. For example, in most cases, a place located in a higher latitude area has a lower annual temperature and higher snowfall records. In this case, the latitude value is probably highly collinear with the snowfall records. If these two variables (latitude and snowfall records) are used in a model, one of them is probably redundant and should be excluded. As a general rule of thumb, when the VIF (1/Tolerance value) of a variable or interaction is greater 10, the variable or interaction may be redundant and therefore should be excluded from the model.

3.2.2 The Concept of Baseline Group

As described earlier, categorical variables do not serve as arbitrary input for the model, and do not have any basis in terms of measurement. They will be assigned with arbitrary values through dummy coding after modeling. The dummy coding is usually used when there is a control group as a dependent group, or a reference group, for comparing
with other group members of the same categorical variable. Other group members will be assigned with values according to the dependent group’s value. As for the example shown in Table 3.3, Category A is assigned with 0 for each code variable ($D_1$ and $D_2$), and thus it is served as the reference group while $D_1$ and $D_2$ are coded for Category B and Category C respectively. After modeling, $D_1$ and $D_2$ will be assigned with values, and the value of each code variable is equal to the difference between that specific group and the reference group. When the categorical variable being used is greater than 1, the total number of reference group is greater than 1 given that each categorical variable has its own reference group. In this study, the combination of various reference groups is called baseline group. Since different categorical variables are used differently under different modeling concepts, the “baseline group” varies as the concept of the model changes. Therefore, baseline groups are noted specifically in the process of interpreting obtained models within the later chapter.

3.2.3 The Use of Variables and Parameters

Although the variables that are used in this study have been introduced before, additional considerations are necessary. There are two major issues that need to be pondered on using of parameters: (1) The number of variables that need to be exploited in the model; (2) The accumulative property of some variables.
3.2.3.1 Variables that Applied in the Model

From the viewpoint of increasing prediction accuracy, it might be a good idea to include as many variables in a model as possible. However, a model with too many variables or parameters might not be practical when it is taken into mass usage or further development. In order to keep the obtained model in this study as “simple” as possible, the variables that are concluded as statistical insignificant during the data analysis process will be excluded from the proposed model.

3.2.3.2 Variables that Have Accumulative Property

It is mentioned earlier that among the quantitative variables, traffic loads and weather conditions variables are recorded in the annual manner. However, these variables have an accumulative property. For example, supposing two pavement sections have the same annual ESAL records. One pavement section withstands 5 years before receiving another minor treatment, while another section lasts 10 years. In this case, if the annual ESAL record is used, both these two pavement sections endure same account of traffic loads, which is not correct. The pavement section that lasts for 10 years should endure two times amount of traffic loads when compared with the one that only lasts for 5 years. For the purpose of using higher accuracy data, some variable data might need to be transformed to better reflect the actual situation. The transforming process can be done by multiplying the “age” by the annual amount. Take the snowfall as an example, a pavement section has
an annual record of 28 inches. When at the “age” of 1, the accumulative snowfall record of
that pavement section should be 28, when at the “age” of 3, and then the accumulative
record should be 84, and so forth.

It should also be noted that, the “transformation” might be, or might not be greatly
influence the results on the modeling. The obtained models by using transforming data
could be dramatically different from the obtained models that using annual data or only
have slightly difference. Therefore, in order to keep the procedure simple, if the use of
transforming data cannot greatly improve the forecasting models accuracy after
comparisons, then the transforming process should be dismissed.

3.3 Softwares Implementation

3.3.1 Models Generating

In the process of applying filtered data on developing pavement condition
prediction models, several statistical analysis softwares such as R, SPSS, SPLUS and even
Microsoft Excel can be chosen. In this study, SPSS is employed on generating forecasting
models. SPSS Statistics is a commonly used statistical software package which can
perform complex data manipulation and analysis with relatively simple instructions.
Unlike inputting programming language to perform data analysis functions as R requires
users do, most of tasks in SPSS can be performed by clicking drop-down menus and setting
up through dialog boxes. SPSS also provides a convenient data checking and editing
function through a spreadsheet interface, and data that used for analysis can easily be imported through Excel spreadsheet file, text file or comma-separated values (CSV) file. The statistical models fitting results, as well as related plots can be automatically outputted upon user’s command in SPSS. In addition, R was also applied during this study, for its expeditious executive ability on statistics analysis, and serving as a method on examining the results obtained through SPSS.

### 3.3.2 Results Comparison

Although the proposed models and corresponding fitness indicators can be obtained through SPSS, it is still worthwhile to make comparison between the predicted results from the models that obtained through this study and the predictions through other prediction method. Since the ODOTPMIS has a comprehensive pavement condition forecasting program (which is mainly composed of Regression and Markov prediction methods) that is currently in effect, and also shares the same database with this study, it is more convenient and meaningful to perform a predictions’ comparison.

To predict future pavement conditions through the obtained models from this study, Microsoft Excel can be employed by setting up algorithm between cells which parameter data are being stored. As for the predictions by ODOTPMIS, they can be acquired from the Access database which ODOTPMIS is employing.
Chapter 4

Results and Findings

4.1 Obtained Models

With the help of SPSS, various model concepts that described in chapter 3 were simulated and tested. Under the settings of SPSS, the obtained models came with a series of reports such as coefficient values, standard error of coefficients, statistical significances of predictors, R square, collinearity diagnosis and so on, of the obtained model. Among these reports, the coefficient of determination, which is also called $R^2$, is generally used as an index to indicate how well the data fit the generated model in statistics. Statistical significance value is used to indicate whether the predictor is an important predictor within the forecasting model. Collinearity diagnosis is used on checking if some predictors included in the model are highly collinear. A first step on choosing the models of this study started from comparing the R square value of the models. Therefore, two models that achieved the highest R square value were selected for model selection candidates. A further model selection, as well as the related analysis was provided in the later sections.

Besides the models that obtained through the concepts of chapter 3. Another set of
models were also simulated under the professional advice from the professor and students of the Department of Mathematics and Statistics. A detailed introduction of the concepts of this set of models and the related results comparison are also provided in the later sections.

4.1.1 Models and Interpretations

After several model simulation trials, there were two models that achieved the highest R square value. Both these two models were built based upon the Random Slope and Fixed Intercept (RSFI) concept that introduced in chapter 3 and their specifications are presented in the later sections. The only difference between them was one model treated the weather data and traffic data as constant quantitative variables and the other treated these variables as cumulative quantitative variables. In another word, one model used average weather data and average traffic data as input while the other considered the total amount of weathering and traffic loadings that had landed on the pavement based on the amount of time that pavement had served. As was mentioned earlier, the values of the cumulative variables could be obtained through multiplying the “age” by the average amount data. In the following paragraphs, the model that treated weather and traffic data as constant quantitative variables is called Random Slope and Fixed Intercept Model with Average Weather and Traffic Data, the other model that treated weather and traffic data as cumulative quantitative variables is named as Random Slope and Fixed Intercept Model
with Cumulative Weather and Traffic Data. In addition, the obtained models after statistics consulting would be introduced in the later section as well.

4.1.1.1 Random Slope and Fixed Intercept Models with Average Weather and Traffic Data

The obtained Random Slope and Fixed Intercept Model with average weather records and traffic records are presented through the following equation, as well as a chart. According to the report provided by SPSS, this model achieved an R square value of 0.733.

\[
PCR_i = 83.68 + (\beta_{Dist} + \beta_{PaveType} + \beta_{PaveSys}) \times Age + 0.019 \times Thickness_{Added} \\
+ 0.257 \times Average_{Temp} + 0.034 \times Average_{Rain} - 0.018 \times Average_{Snow} \\
+ 2.636 \times 10^{-4} \times Average_{ESAL}
\]

In the above equation, \(\beta_{Dist}, \beta_{PaveType}, \beta_{PaveSys}\) are representing the coefficients of the variables of district division, pavement type and pavement system respectively. Since these three variables were designated as categorical variables in the model, the values of the coefficients varied as the predictors switched. \(Thickness_{Added}\) is the values of the quantitative variable of overlay thickness added on the pavement. \(Average_{Temp}\), \(Average_{Rain}\) and \(Average_{Snow}\) are representing the values of the quantitative variables of average temperature, average rainfall and average snowfall records respectively. A specific coefficient values of all the predictors, as well as their standard errors and
significances are provided in the following Table 4.1. The baseline categorical variables parameter group for this proposed model is consisted of general pavement system (General) and Composite pavement type (COMP).

### Table 4.1 Coefficients of RSFI Models with Average Weather and Traffic Data

<table>
<thead>
<tr>
<th>Variables</th>
<th>Coefficient Values</th>
<th>Std. Error</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>83.680</td>
<td>0.738</td>
<td>0.000</td>
</tr>
<tr>
<td>Dist. 1</td>
<td>-2.030</td>
<td>0.013</td>
<td>0.000</td>
</tr>
<tr>
<td>Dist. 2</td>
<td>-2.335</td>
<td>0.012</td>
<td>0.000</td>
</tr>
<tr>
<td>Dist. 3</td>
<td>-3.155</td>
<td>0.011</td>
<td>0.000</td>
</tr>
<tr>
<td>Dist. 4</td>
<td>-3.026</td>
<td>0.014</td>
<td>0.000</td>
</tr>
<tr>
<td>Dist. 5</td>
<td>-2.861</td>
<td>0.012</td>
<td>0.000</td>
</tr>
<tr>
<td>Dist. 6</td>
<td>-2.300</td>
<td>0.012</td>
<td>0.000</td>
</tr>
<tr>
<td>Dist. 7</td>
<td>-2.270</td>
<td>0.012</td>
<td>0.000</td>
</tr>
<tr>
<td>Dist. 8</td>
<td>-2.067</td>
<td>0.013</td>
<td>0.000</td>
</tr>
<tr>
<td>Dist. 9</td>
<td>-1.988</td>
<td>0.015</td>
<td>0.000</td>
</tr>
<tr>
<td>Dist. 10</td>
<td>-2.386</td>
<td>0.013</td>
<td>0.000</td>
</tr>
<tr>
<td>Dist. 11</td>
<td>-2.822</td>
<td>0.013</td>
<td>0.000</td>
</tr>
<tr>
<td>Dist. 12</td>
<td>-2.723</td>
<td>0.019</td>
<td>0.000</td>
</tr>
<tr>
<td>PaveSys</td>
<td>Priority</td>
<td>0.449</td>
<td>0.010</td>
</tr>
<tr>
<td></td>
<td>Urban</td>
<td>0.011</td>
<td>0.088</td>
</tr>
<tr>
<td>PaveType</td>
<td>JRC</td>
<td>0.357</td>
<td>0.051</td>
</tr>
<tr>
<td></td>
<td>FLEX</td>
<td>-0.042</td>
<td>0.007</td>
</tr>
<tr>
<td>Thickness</td>
<td>Added</td>
<td>0.019</td>
<td>0.002</td>
</tr>
<tr>
<td>Average ESAL</td>
<td>-2.636 × 10^{-4}</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>Average Temp</td>
<td>0.257</td>
<td>0.014</td>
<td></td>
</tr>
<tr>
<td>Average Rain</td>
<td>0.034</td>
<td>0.007</td>
<td></td>
</tr>
<tr>
<td>Average Snow</td>
<td>-0.018</td>
<td>0.002</td>
<td></td>
</tr>
</tbody>
</table>

In the above table, the variable called “Constant” represents the PCR₀ in the proposed model thus the PCR₀ is 83.68 for the prediction model, no matter which district
that pavement section is located in or what type of pavement overlay that pavement section is being applied. “Dist” represents the categorical variable district division of the model, and it can be observed from the table that each district division is assigned with a different value. The specific value for each district division serves as the slope, which is the value of $\beta_{Dist}$ in the equation (6). For example, if a pavement section is located in the 9th district division, then the value of $\beta_{Dist}$ for calculation is -1.988. Likewise, “PaveSys” and “PaveType” represent the categorical variables pavement system and pavement type of the model. Since the baseline variables of this model belongs to the variables of pavement system and pavement type, the meaning of the coefficients that corresponding to these two variables are a bit different from the “Dist” variable. For instance, the baseline group for pavement system variable is general pavement system (General). Therefore, if a pavement section belongs to General, then the value $\beta_{PaveType}$ used in the predicting model is 0 because General serves as a baseline variables group member. If a pavement section belongs to urban pavement system (Urban), the value $\beta_{PaveType}$ used in the predicting model is 0.011 according to Table 4.1. Take another example; if a pavement section locates in Dist. 8, belongs to urban pavement system and uses Composite pavement type as its overlay, then the slope of the prediction model is -2.056.

For the rest of the variables shown in Table 4.1, which are quantitative variables, one of them needs a further illustration. The variable “Thickness Added” in prediction model is determined to be simplified for two reasons. First, 99% of pavement sections within this
study have their “Thickness Added” records fallen between 0 inches and 5 inches. Second, “Thickness Added” is a variable that related to specific projects and it is only included in project records, the variable values could vary after a pavement section has received another project. It is also for this reason that, the variable “Thickness Added” is also not available within the PMIS’s prediction dataset. As a consequence, it is assumed that the “Thickness Added” for all the pavement sections is 3 inches, so that the effect of “Thickness Added” used in the model is replaced by a constant of 0.057, which is obtained through multiplying 0.019 by 3. Thus the equation (6) can be rewritten as follows:

\[
P_{CR_i} = 83.737 + (\beta_{Dist} + \beta_{PaveType} + \beta_{PaveSys}) \times \text{Age} + 0.257 \times \text{Average Temp} + 0.034 \times \text{Average Rain} - 0.018 \times \text{Average Snow} + 2.636 \times 10^{-4} \times \text{Average ESAL}
\]

(7)

4.1.1.2 Random Slope and Fixed Intercept Models with Cumulative Weather and Traffic Data

When the cumulative property of the variables weather and traffic were taken into consideration on modeling, the obtained Random Slope and Fixed Intercept Model are presented as the following equation and Table 4.2.

\[
P_{CR_i} = 97.43 + (\beta_{Dist} + \beta_{PaveType} + \beta_{PaveSys}) \times \text{Age} + 0.019 \times \text{Thickness Added} - 6.258 \times 10^{-5} \times \text{Cumulative ESAL} + 0.078 \times \text{Cumulative Temp} + 0.018 \times \text{Cumulative Rain} - 4.889 \times 10^{-4} \times \text{Cumulative Snow}
\]

(8)
The baseline categorical variables parameter group for this proposed model has the same combination as the previous model, which is consisted of general pavement system (General) and Composite pavement type (COMP).

Table 4.2 Coefficients of RSFI Models with Cumulative Weather and Traffic Data

<table>
<thead>
<tr>
<th>Variables</th>
<th>Coefficient Values</th>
<th>Std. Error</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>97.427</td>
<td>0.024</td>
<td>0.000</td>
</tr>
<tr>
<td>Dist</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dist. 1</td>
<td>-6.616</td>
<td>0.175</td>
<td>0.000</td>
</tr>
<tr>
<td>Dist. 2</td>
<td>-6.867</td>
<td>0.173</td>
<td>0.000</td>
</tr>
<tr>
<td>Dist. 3</td>
<td>-7.766</td>
<td>0.175</td>
<td>0.000</td>
</tr>
<tr>
<td>Dist. 4</td>
<td>-7.676</td>
<td>0.177</td>
<td>0.000</td>
</tr>
<tr>
<td>Dist. 5</td>
<td>-7.535</td>
<td>0.179</td>
<td>0.000</td>
</tr>
<tr>
<td>Dist. 6</td>
<td>-6.963</td>
<td>0.179</td>
<td>0.000</td>
</tr>
<tr>
<td>Dist. 7</td>
<td>-6.905</td>
<td>0.178</td>
<td>0.000</td>
</tr>
<tr>
<td>Dist. 8</td>
<td>-6.793</td>
<td>0.183</td>
<td>0.000</td>
</tr>
<tr>
<td>Dist. 9</td>
<td>-6.716</td>
<td>0.183</td>
<td>0.000</td>
</tr>
<tr>
<td>Dist. 10</td>
<td>-7.095</td>
<td>0.182</td>
<td>0.000</td>
</tr>
<tr>
<td>Dist. 11</td>
<td>-7.467</td>
<td>0.177</td>
<td>0.000</td>
</tr>
<tr>
<td>Dist. 12</td>
<td>-7.465</td>
<td>0.183</td>
<td>0.000</td>
</tr>
<tr>
<td>PaveSys</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Priority</td>
<td>0.465</td>
<td>0.011</td>
<td>0.000</td>
</tr>
<tr>
<td>Urban</td>
<td>0.006</td>
<td>0.008</td>
<td>0.472</td>
</tr>
<tr>
<td>PaveType</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>JRC</td>
<td>0.353</td>
<td>0.051</td>
<td>0.000</td>
</tr>
<tr>
<td>FLEX</td>
<td>-0.044</td>
<td>0.007</td>
<td>0.000</td>
</tr>
<tr>
<td>Thickness Added</td>
<td>0.019</td>
<td>0.002</td>
<td>0.000</td>
</tr>
<tr>
<td>Cumulative ESAL</td>
<td>-6.258 × 10⁻⁵</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Cumulative Temp</td>
<td>0.078</td>
<td>0.003</td>
<td>0.000</td>
</tr>
<tr>
<td>Cumulative Rain</td>
<td>0.018</td>
<td>0.002</td>
<td>0.000</td>
</tr>
<tr>
<td>Cumulative Snow</td>
<td>-4.889 × 10⁻⁴</td>
<td>0.001</td>
<td>0.430</td>
</tr>
</tbody>
</table>

As for this model, the variable “Cumulative Snow” is removed because the coefficient value computed by the software is 0. Likewise, the effect of variable...
“Thickness Added” in the model is determined to be simplified as the previous model, and it is replaced by a constant of 0.057 as well. Thus, equation (8) can be rewritten as follows:

\[
PCR_i = 97.487 + (\beta_{Dist} + \beta_{PaveType} + \beta_{PaveSys}) \times Age + 0.078 \times Cumulative\_Temp \\
- 6.258 \times 10^{-5} \times Cumulative\_ESAL + 0.018 \times Cumulative\_Rain
\]  

(9)

4.1.2 Models Selection, Further Modifications and Considerations

Apart from the Significance Level, there are generally two other indicators (Adjusted R-square and VIF) on determining whether a main effect or interaction variable should be included in a statistical model. The two sets of the obtained models presented in section 4.1.1.1 and section 4.1.1.2 shares the same Adjusted R-square, which is 0.733. Therefore, Significance Level check and VIF check were adopted in the further analysis step.

4.1.2.1 Significance Level

In statistics, the p-value is used as a continuous measure of divergence of the fit of a model. The use of p-value is always accompanied by a confidence interval, which serves as a threshold on indicating whether a variable in the model is “statistically significant”. For example, when the confidence interval is set as 90%, the p-value of a “statistically significant” variable should lesser than 0.10, and it can also be concluded that: “the variable is statistically significant at the 10% level, or there is a 90% probability that the
variable has an effect in the model.” Otherwise, the variable should be treated as a baseline group variable. In this study, the confidence interval is set as 95%. In other words, the $p$-value of a “statistically significant” variable should be less than 0.05. The $p$-values of the variables of models presented by Table 4.1 and Table 4.2 are represented by the column named “Significance”, which is the rightmost column within these two tables.

It can be observed that in the Table 4.1, which average weather and traffic records were used as quantitative variables, the categorical variable parameter “Urban pavement system” has a $p$-value that greater than 0.05. Therefore, it should be treated as a baseline group variables parameter in this model. Since “General pavement system” is a baseline group variables parameter, it can be also expressed that: “there is no significant difference in the pavement performance predictions when treating a pavement section of an urban pavement system (Urban) as a pavement section of a general pavement system (General)”.

As for the results in the Table 4.2, it can be observed that a variable parameter “Urban pavement system” and the variable “Cumulative snowfall records” have the $p$-value that greater than 0.05. Therefore, the “Urban pavement system” would also be treated as a baseline group variables parameter. The “Cumulative snowfall records” variable should be excluded from the model, it can also be expressed that: “there is no significant difference in the performance predictions when the cumulative snowfall records are not taken into account.”
4.1.2.2 VIF and Tolerance

As described in the previous chapter, a general rule of thumb in collinearity diagnosis of the main effects or interactions of a model is determined by a threshold of 10 when applying the Variance Inflation Factor (VIF), or a threshold of 0.1 when applying Tolerance value (since Tolerance value equals to 1/VIF). For example, if the VIF of a variable is greater than 10 (or less than 0.1 when using Tolerance value as the indicator), it implies that the variable is likely to be deduced from another variable that included in the model which means the variable is probably redundant in the model; and the redundant variable should be removed from the proposed model in statistical analysis. The corresponding VIF and Tolerance values of the variables shown in Table 4.1 and Table 4.2 are presented in the following Table 4.3 and Table 4.4 respectively.
<table>
<thead>
<tr>
<th>Variables</th>
<th>Coefficient Values</th>
<th>VIF</th>
<th>Tolerance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>83.680</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dist. 1</td>
<td>-2.030</td>
<td>1.373</td>
<td>0.728</td>
</tr>
<tr>
<td>Dist. 2</td>
<td>-2.335</td>
<td>1.454</td>
<td>0.688</td>
</tr>
<tr>
<td>Dist. 3</td>
<td>-3.155</td>
<td>1.447</td>
<td>0.691</td>
</tr>
<tr>
<td>Dist. 4</td>
<td>-3.026</td>
<td>1.454</td>
<td>0.688</td>
</tr>
<tr>
<td>Dist. 5</td>
<td>-2.861</td>
<td>1.354</td>
<td>0.739</td>
</tr>
<tr>
<td>Dist. 6</td>
<td>-2.300</td>
<td>1.485</td>
<td>0.674</td>
</tr>
<tr>
<td>Dist. 7</td>
<td>-2.270</td>
<td>1.377</td>
<td>0.726</td>
</tr>
<tr>
<td>Dist. 8</td>
<td>-2.067</td>
<td>1.529</td>
<td>0.654</td>
</tr>
<tr>
<td>Dist. 9</td>
<td>-1.988</td>
<td>1.300</td>
<td>0.769</td>
</tr>
<tr>
<td>Dist. 10</td>
<td>-2.386</td>
<td>1.324</td>
<td>0.755</td>
</tr>
<tr>
<td>Dist. 11</td>
<td>-2.822</td>
<td>1.302</td>
<td>0.768</td>
</tr>
<tr>
<td>Dist. 12</td>
<td>-2.723</td>
<td>1.735</td>
<td>0.577</td>
</tr>
<tr>
<td>PaveSys</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Priority</td>
<td>0.449</td>
<td>1.732</td>
<td>0.577</td>
</tr>
<tr>
<td>Urban</td>
<td>0.011</td>
<td>1.339</td>
<td>0.747</td>
</tr>
<tr>
<td>PaveType</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>JRC</td>
<td>0.357</td>
<td>1.008</td>
<td>0.992</td>
</tr>
<tr>
<td>FLEX</td>
<td>-0.042</td>
<td>1.845</td>
<td>0.542</td>
</tr>
<tr>
<td>Thickness Added</td>
<td>0.019</td>
<td>1.404</td>
<td>0.712</td>
</tr>
<tr>
<td>Average ESAL</td>
<td>$-2.636 \times 10^{-4}$</td>
<td>1.003</td>
<td>0.997</td>
</tr>
<tr>
<td>Average Temp</td>
<td>0.257</td>
<td>2.130</td>
<td>0.469</td>
</tr>
<tr>
<td>Average Rain</td>
<td>0.034</td>
<td>1.497</td>
<td>0.668</td>
</tr>
<tr>
<td>Average Snow</td>
<td>-0.018</td>
<td>2.680</td>
<td>0.373</td>
</tr>
</tbody>
</table>

It can be observed from Table 4.3 that, all the variables shown in the table meet the collinearity check because none of variable achieves a VIF that is greater than 10. In other words, none of them should be dropped from the proposed model based on the collinearity diagnosis.
Table 4.4 VIF and Tolerance Values of Variables of Table 4.2

<table>
<thead>
<tr>
<th>Variables</th>
<th>Coefficient Values</th>
<th>VIF</th>
<th>Tolerance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>97.427</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dist</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dist. 1</td>
<td>-6.616</td>
<td>250.121</td>
<td>0.004</td>
</tr>
<tr>
<td>Dist. 2</td>
<td>-6.867</td>
<td>318.329</td>
<td>0.003</td>
</tr>
<tr>
<td>Dist. 3</td>
<td>-7.766</td>
<td>369.158</td>
<td>0.003</td>
</tr>
<tr>
<td>Dist. 4</td>
<td>-7.676</td>
<td>234.037</td>
<td>0.004</td>
</tr>
<tr>
<td>Dist. 5</td>
<td>-7.535</td>
<td>310.863</td>
<td>0.003</td>
</tr>
<tr>
<td>Dist. 6</td>
<td>-6.963</td>
<td>357.449</td>
<td>0.003</td>
</tr>
<tr>
<td>Dist. 7</td>
<td>-6.905</td>
<td>314.525</td>
<td>0.003</td>
</tr>
<tr>
<td>Dist. 8</td>
<td>-6.793</td>
<td>293.436</td>
<td>0.003</td>
</tr>
<tr>
<td>Dist. 9</td>
<td>-6.716</td>
<td>201.449</td>
<td>0.005</td>
</tr>
<tr>
<td>Dist. 10</td>
<td>-7.095</td>
<td>244.926</td>
<td>0.004</td>
</tr>
<tr>
<td>Dist. 11</td>
<td>-7.467</td>
<td>237.769</td>
<td>0.004</td>
</tr>
<tr>
<td>Dist. 12</td>
<td>-7.465</td>
<td>163.233</td>
<td>0.006</td>
</tr>
<tr>
<td>PaveSys</td>
<td>Priority</td>
<td>0.465</td>
<td>2.168</td>
</tr>
<tr>
<td></td>
<td>Urban</td>
<td>0.006</td>
<td>1.354</td>
</tr>
<tr>
<td>PaveType</td>
<td>JRC</td>
<td>0.353</td>
<td>1.008</td>
</tr>
<tr>
<td></td>
<td>FLEX</td>
<td>-0.044</td>
<td>1.889</td>
</tr>
<tr>
<td>Thickness Added</td>
<td>0.019</td>
<td>1.003</td>
<td>0.997</td>
</tr>
<tr>
<td>Cumulative ESAL</td>
<td>-6.258 × 10⁻⁵</td>
<td>2.139</td>
<td>0.468</td>
</tr>
<tr>
<td>Cumulative Temp</td>
<td>0.078</td>
<td>1295.784</td>
<td>0.001</td>
</tr>
<tr>
<td>Cumulative Rain</td>
<td>0.018</td>
<td>170.584</td>
<td>0.006</td>
</tr>
<tr>
<td>Cumulative Snow</td>
<td>-4.889 × 10⁻⁴</td>
<td>23.799</td>
<td>0.042</td>
</tr>
</tbody>
</table>

As can be observed from Table 4.4, the district division category, cumulative temperature records, cumulative rainfall records as well as cumulative snowfall records fail to meet the VIF or Tolerance requirement. It should be noted that the difference between the models of Table 4.1 and Table 4.2 is the use of cumulative records of traffic loads, temperature, rainfall and snowfall. A deduction of the existence of high VIF value is that
the cumulative climate records, especially the cumulative temperature records, and the
district division variable are highly collinear. In other words, the cumulative climate
records can be deduced according to the district division. The guess makes sense given that
the climate character of a district is generally determined by its geographic character (e.g.
latitude and topography); and the VIF of cumulative ESAL (less than 10) also supports the
supposal to some extent since traffic loads of pavement sections are unlikely to be
determined by natural character.

4.1.2.3 Summary and Updated Models

As a consequence, the Random Slope and Fixed Intercept (RSFI) models with
cumulative weather and traffic data should be dismissed in this study given that the concept
brings up relatively high collinearity between variables in the proposed models.

After variables or parameters removal (including the removal of variable
“Thickness Added” since it is not available in the prediction sheets of the PMIS database)
based on $p$-values and VIF, the updated equation and coefficient values of Random Slope
and Fixed Intercept models with average weather and traffic data concept are re-simulated
in SPSS, the obtained results are presented as the following equation and Table 4.5.

\[
PCR_i = 83.674 + (\beta_{Dist} + \beta_{PaveType} + \beta_{PaveSys}) \times Age + 0.258 \times Average\_Temp \\
+ 0.034 \times Average\_Rain - 0.018 \times Average\_Snow + 2.636 \times 10^{-4} \times Average\_ESAL
\]  (10)
Table 4.5 Coefficients of Updated RSFI Models of Equation No. 10

<table>
<thead>
<tr>
<th>Variables</th>
<th>Coefficient Values</th>
<th>Std. Error</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>83.674</td>
<td>0.737</td>
<td>0.000</td>
</tr>
<tr>
<td>Dist. 1</td>
<td>-2.028</td>
<td>0.013</td>
<td>0.000</td>
</tr>
<tr>
<td>Dist. 2</td>
<td>-2.332</td>
<td>0.012</td>
<td>0.000</td>
</tr>
<tr>
<td>Dist. 3</td>
<td>-3.151</td>
<td>0.010</td>
<td>0.000</td>
</tr>
<tr>
<td>Dist. 4</td>
<td>-3.023</td>
<td>0.014</td>
<td>0.000</td>
</tr>
<tr>
<td>Dist. 5</td>
<td>-2.858</td>
<td>0.012</td>
<td>0.000</td>
</tr>
<tr>
<td>Dist. 6</td>
<td>-2.297</td>
<td>0.011</td>
<td>0.000</td>
</tr>
<tr>
<td>Dist. 7</td>
<td>-2.267</td>
<td>0.011</td>
<td>0.000</td>
</tr>
<tr>
<td>Dist. 8</td>
<td>-2.062</td>
<td>0.013</td>
<td>0.000</td>
</tr>
<tr>
<td>Dist. 9</td>
<td>-1.987</td>
<td>0.015</td>
<td>0.000</td>
</tr>
<tr>
<td>Dist. 10</td>
<td>-2.385</td>
<td>0.013</td>
<td>0.000</td>
</tr>
<tr>
<td>Dist. 11</td>
<td>-2.822</td>
<td>0.013</td>
<td>0.000</td>
</tr>
<tr>
<td>Dist. 12</td>
<td>-2.716</td>
<td>0.018</td>
<td>0.000</td>
</tr>
<tr>
<td>PaveSys</td>
<td>Priority</td>
<td>0.448</td>
<td>0.010</td>
</tr>
<tr>
<td>PaveType</td>
<td>JRC</td>
<td>0.358</td>
<td>0.051</td>
</tr>
<tr>
<td></td>
<td>FLEX</td>
<td>-0.044</td>
<td>0.007</td>
</tr>
<tr>
<td>Average ESAL</td>
<td>-0.00026</td>
<td>0.000168</td>
<td>0.000</td>
</tr>
<tr>
<td>Average Temp</td>
<td>0.258</td>
<td>0.014</td>
<td>0.000</td>
</tr>
<tr>
<td>Average Rain</td>
<td>0.034</td>
<td>0.007</td>
<td>0.000</td>
</tr>
<tr>
<td>Average Snow</td>
<td>-0.018</td>
<td>0.002</td>
<td>0.000</td>
</tr>
</tbody>
</table>

The baseline categorical variables parameter group for the updated model is consisted of general pavement system (General), urban pavement system (Urban) and Composite pavement type (COMP). The R-square and the Adjusted R-square value for this updated are both 0.733. By comparing the obtained R-square value before and after variable and variable parameters removal, which are both 0.733 in this case, it is revealed that the removal of the “statistically significant” quantitative variable “Thickness Added”
did not make a distinguished difference on the prediction. This is probably due to the fact that the computed effects of “Thickness Added” on the PCR value are relatively small.

### 4.1.3 Models through Statistics Consulting

A set of models were also simulated and selected under the advice of statistics consulting. Since these models were generated at district level, and the specifications of the models were different after main effects and interaction variables selections. Only the equation and models for District No. 1 is presented in this section and the rest of the obtained models are presented in Appendix C.

\[
PCR_i = 102.529 + (\beta_{\text{PaveType}} + \beta_{\text{PaveSys}}) \cdot \text{Age} - 0.001 \cdot \text{Average}_\text{ESAL} - 0.159 \cdot \text{Average}_\text{Snow} \tag{11}
\]

The baseline categorical variables parameter group for the above proposed model is consisted of Composite pavement type (COMP), the quantitative variable “Average ESAL” was excluded because this variable is not available in the prediction dataset of ODOTPMIS database, another two variables “Average temperature” and “Average rainfall records” were excluded because their existences did not improve the R-square of the model. Since the models simulated under statistics consulting were at district level individually, the variables used in the obtained models after Significance Level and VIF selections of different districts are slightly different. The R-square of the obtained model of District No. 1 is 0.703.
### Table 4.6 Coefficients of Models of District No. 1 after Statistics Consulting

<table>
<thead>
<tr>
<th>Variables</th>
<th>Coefficient Values</th>
<th>Std. Error</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>102.529</td>
<td>0.415</td>
<td>0.000</td>
</tr>
<tr>
<td>PaveSys</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>General</td>
<td>-2.349</td>
<td>0.025</td>
<td>0.000</td>
</tr>
<tr>
<td>Priority</td>
<td>-1.805</td>
<td>0.036</td>
<td>0.000</td>
</tr>
<tr>
<td>Urban</td>
<td>-2.353</td>
<td>0.029</td>
<td>0.000</td>
</tr>
<tr>
<td>PaveType</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FLEX</td>
<td>0.266</td>
<td>0.024</td>
<td>0.000</td>
</tr>
<tr>
<td>JRC</td>
<td>0.822</td>
<td>0.114</td>
<td>0.000</td>
</tr>
<tr>
<td>Average ESAL</td>
<td>-0.001</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Average Temp</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Average Rain</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Average Snow</td>
<td>-0.159</td>
<td>0.015</td>
<td>0.000</td>
</tr>
</tbody>
</table>

It can be observed from the Table 4.6 and Equation No. 11 that no interaction variable is included in the models. The reason is that, after trials during modeling, it was found that any interaction combination variable would inevitably achieve high VIF (greater than 10 according to the rule of thumb of interpreting VIF). By comparing the obtained models of statistics advices and the updated RSFI models, it can be seen that the categorical variable “District Division” all disappeared, that is because the models were simulated at district level so that the effect of district division does not exist and therefore this variable was removed. In addition, it can also be noticed that the general pavement system parameter (General) of categorical variable “Pavement System” is included in the models. That is due to the need of a small modification on the coding of dummy regressors described in the section 3.2.1.3.
4.2 Forecasting Results Comparison

Besides using the coefficient of determination values (R-square) of the obtained models to check the fitness of the models, there is another approach on further judging the model performance. The approach is to making comparison with other existing models’ prediction results. It needs an existing model which should be enduring or has been proven to be somewhat reliable. In the case of this study, the existing model is the currently operating ODOTPMIS program.

It should be stated that both this comparing approach has its disadvantages under the case of this study. One of the disadvantages is that, the comparisons are between two types of predictions and thus it is improper to make judgment on the fitness of the models that are being compared with.

In this study, both the two sets of models obtained (Updated RSFI models and models simulated under statistical advices) were used for computing the PCRs for prediction comparisons. Five-year period forecasted PCRs comparisons were used in this process.

4.2.1 An Introduction of ODOTPMIS Prediction Approaches

It is mentioned in the section 3.3.2 that, the pavement condition forecasting methods used in ODOTPMIS is mainly consisted of Markov approach and regression approach.
4.2.1.1 Regression Approach Used in ODOTPMIS

In terms of the regression approach that is being used, it is based on linear regression models. The difference between the regression models proposed in this study and the regression models in ODOTPMIS is, the regression models used in ODOTPMIS use only “age” as a variable, the *slopes* and *intercepts* are generated based on the existed PCR values of the predicted pavement section. Likewise, each pavement section that is assigned with this prediction approach has its own regression model. The advantages of this prediction approach are: (1) each pavement section’s prediction model is not “determined” by the performance of other pavement sections that sharing the same characters; (2) the approach is easy to be adopted, especially when the PCR records of a newly built/rehabbed pavement section are not sufficient for the sophisticated prediction approaches.

4.2.1.2 Markov Approach Used in ODOTPMIS

It is worthwhile to describe the Markov approach that is being used in ODOTPMIS. As it is mentioned in chapter 2 that, Markov models are generally being applied for large scale pavement network (state-level or city-level pavement network) condition forecasting and the results are usually presented in a way that informing the percentage rates of each condition state’s pavement lengths in proportion to the entire pavement network lengths. It is apparent that this approach cannot provide an exact PCR
of a pavement section but the probability of pavement condition state of an entire pavement network. However, in ODOTPMIS, the concept of Markov approach is used on forecasting the exact PCRs of pavement sections. The computing method is presented in the following example.

Assuming the PCR of a pavement section is 93 in this year, according to the condition state categorization, it should be categorized as in “Very Good” condition. Using the probabilities concept of Markov approach, assuming it is expected that there is 85% that a pavement section in “Very Good” condition will stay in the same condition state in the next year while there is a 15% chance that the pavement section will drop to “Good” condition. Then the current PCR and the intermediate PCR value of the “next” condition state are used on computing the future PCR value. As for this example, the intermediate PCR value of the “Good” condition is 83 according to the PCR ranges of condition state. Therefore, this pavement section’s PCR of next year will be computed as:

\[ 93 \times 85\% + 83 \times 15\% = 91.5. \]

The advantage of this Markov approach is that, when the pavement section represent an abnormal deterioration trend, such as staying in the same condition state for over 5 years due to unrecorded treatments or drop rapidly after a year. The sudden change can be “captured” and therefore the prediction can be “recovered” because its prediction is based upon the latest condition of the pavement section instead of the condition existed before.
4.2.2 The Adoption of Weighted Correlation

In general, correlation coefficient is used as a measure on checking if two sets of data are strongly linked together. Correlation coefficient is usually expressed by $\rho$, and the result range is between -1 and 1. When the correlation is getting closer to 1, it indicates that the two sets of data are positively correlated, or they have a direct linear relationship. While the result is getting closer to -1, it implies that the two sets of data are negatively correlated, or they are in an inverse linear relationship. If the result trends 0, it usually suggests that the two sets of data are not correlated. Supposing there are two sets of data $X$ and $Y$, the correlation between $X$ and $Y$ can be computed through the following formula:

$$
\rho_{X,Y} = \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum(X_i - \bar{X})^2} \sqrt{\sum(Y_i - \bar{Y})^2}}
$$

Where:

- $X_i$ = the $i^{th}$ value of dataset $X$,
- $\bar{X}$ = the mean of dataset $X$.

However, the above equation might not be the best choice to calculate the correlation for this study. Generally, the above equation is being used when each item of the sample is considered as a uniform observation (e.g., A specific type of screws is manufactured in roughly the same size and weight under certain standards). For the case of each pavement section that was recorded in the dataset of this study, the premise of
uniform observation cannot be satisfied if the length of each pavement section is considered. In ODOT’s database, pavement sections are being partitioned in a complicated manner instead of merely using a uniform length. This complicated manner including the factors of similar recent activities history, similar pavement type, or similar current conditions and so on. Thus, with new treatment activities being applied and pavement conditions of pavement sections are in constant change year by year, the partitions of pavement sections in the database also varies with the years. The partition manner also leads to the fact that, the length of each pavement section within the database is not uniform. According to the latest database, the partition length varies from 0.01 miles to 13.55 miles. The ununiformed pavement sections length brings up the issue of the accuracy of the correlation if the general equation is applied on calculating the relationship between the PCR values of two sets of pavement sections. In order to eliminate the inaccuracy issue that might be brought by because of this ununiformed pavement partition manner, weighted correlation is introduced to be applied on correlation calculations. Weighted correlation can be represented by the following formula:

$$\rho_{(X,Y)w} = \frac{\sum w_i (X_i - \bar{X}_w)(Y_i - \bar{Y}_w)}{\sqrt{\sum w_i (X_i - \bar{X}_w)^2} \sqrt{\sum w_i (Y_i - \bar{Y}_w)^2}}$$

Where:

$$w_i = \text{the weight assigned to item } X_i,$$
\[
\bar{X}_w = \frac{\sum w_i X_i}{\sum w_i},
\]
\[
\bar{Y}_w = \frac{\sum w_i Y_i}{\sum w_i}.
\]

The usefulness of weighted correlation is presented through an example in the Appendix B.

4.2.3 Comparison Results and Findings

To perform comparison between the predicted PCR values from the models obtained in this study and predictions of existing models, modifications on the dataset are also needed. In this process, those pavement sections that had received minor or major treatments before 2012 were removed from the dataset, in other words, those consistent pavement condition records that ended before 2012 were removed because they would have new record rows within the database. Besides, those records that did not have an identical first PCR record year for both datasets (the filtered dataset that used in this study and the ODOTPMIS predictions dataset) were removed as well. In addition, it should be noted that the data filtering rules described in section 3.1.3 were adopted again to filter out those undesired PCR records which presented in the database used for future condition forecasting. An upcoming five-year period (based on the last actual PCR records year stored in the database) forecasting results comparison between ODOTPMIS program and the model obtained in this study would be started from the year of 2013 to
the year of 2017.

Likewise, the PCR predictions comparison between obtained models and ODOTPMIS were also compared yearly and the weighted correlation values of yearly comparisons are presented in the following Table 4.7.

<table>
<thead>
<tr>
<th>Year</th>
<th>Updated RSFI vs. ODOTPMIS</th>
<th>Consulting vs. ODOTPMIS</th>
</tr>
</thead>
<tbody>
<tr>
<td>2013</td>
<td>0.814</td>
<td>0.818</td>
</tr>
<tr>
<td>2014</td>
<td>0.793</td>
<td>0.797</td>
</tr>
<tr>
<td>2015</td>
<td>0.771</td>
<td>0.775</td>
</tr>
<tr>
<td>2016</td>
<td>0.750</td>
<td>0.753</td>
</tr>
<tr>
<td>2017</td>
<td>0.729</td>
<td>0.732</td>
</tr>
</tbody>
</table>

In Table 4.7, *Updated RSFI* means the predictions of the obtained models described in section 4.1.2.3, *Consulting* means the predictions of the obtained models generated after statistics consulting and ODOTPMIS means the predictions of ODOTPMIS.

As it can be observed, the weighted correlations between the two sets of obtained models predictions and the ODOTPMIS predictions achieve the highest values in the closest year of 2013; the weighted correlations are 0.814 and 0.818 respectively. It can also be noticed that the weighted correlations decrease as the year of predictions extends, the weighted correlations for two sets of comparisons drops to 0.729 and 0.732 respectively. Besides, the obtained models after statistics consulting achieves slightly higher correlation value than the RSFI models, which indicate that the predictions of
models under statistics advices are closer to the predictions of ODOTPMIS than the predictions by RSFI models.

<table>
<thead>
<tr>
<th>Value of the Correlation Coefficient (+/-)</th>
<th>Strength of Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.90 ~ 1.00</td>
<td>Very High</td>
</tr>
<tr>
<td>0.70 ~ 0.89</td>
<td>High</td>
</tr>
<tr>
<td>0.50 ~ 0.69</td>
<td>Moderate</td>
</tr>
<tr>
<td>0.30 ~ 0.49</td>
<td>Low</td>
</tr>
<tr>
<td>Smaller than 0.29</td>
<td>Little or None</td>
</tr>
</tbody>
</table>

(Adapted from Coffelt, D. P., 2008)

According to the rule of thumb on charactering the value of correlations presented in Table 4.8, the forecasted PCRs between the two sets of models and ODOTPMIS are correlated at the high level in a 5 most recent year’s predicted period. It is expected that, the strength of correlation would drop to moderate level in the prediction comparisons after the 5 most recent year’s period. In other words, the predicted PCRs between the obtained models and ODOTPMIS would have considerable diversions after the 5 most recent years.
Chapter 5

Conclusions and Recommendations

5.1 Conclusions

Accurate predictions on pavement performance is a critical point in a Pavement Management System (PMS) for better facilitate engineers or pavement managers on setting up maintenance and rehabilitation plans and keeping the costs as low as possible. In general, probabilistic approach is used on large scale pavement network (such as all the pavement sections at state level) performance forecasting, while deterministic approach is used on small scale pavement groups (such as specific pavement sections) performance forecasting. This study attempts to develop a linear regression approach which is able to predict pavement performance of specific pavement sections while also able to capture the effects of some pavement characters (such as pavement type, or district division) which might universally exist in a large pavement network.

Different sets of random slope and fixed intercept linear regression models were obtained after data processing, model simulations and variables selections. Through the obtained results and examinations on the database, conclusions can be made as follows:
1. The predictions of the obtained models are highly correlated with the existing ODOTPMIS prediction results at the recent 5 coming year’s prediction term. The difference between the obtained models’ predictions and ODOTPMIS predictions will be greater as the predicted year in the future extends.

2. Through the VIF or Tolerance Value analysis of the obtained models, it is likely that the effect of district division on the pavement performance is collinear with the effects of weather records. It would be worthwhile to verify this hypothesis through other methods in the future and therefore the effect of either district division difference or weather records can be used in a more proper way in the future modeling.

3. It is found through the obtained models after statistics consulting that the prediction model might have better performance, when making comparison between their predictions with ODOTPMIS predictions, if the models can be broken down into different sub-models based on the different pavement characters accordingly. Although the models of some districts do achieve a higher R-square value when compared with the models obtained before statistics consulting, the improvement of this type of change on modeling still needs to be further examined and tested as the improvement on the overall prediction is not yet known.

4. The interactive effects of different main predictors of pavement performance
are not applicable as the collinearity diagnosis suggested. This finding should be further examined via other modeling approaches.

5. The Markov prediction approach used in ODOTPMIS is more suitable on performance prediction after an “abnormal” change occurs in the deterioration trend of a pavement section. The regression approach developed in this study assumes that the pavement condition of a pavement section is deteriorating in a same pattern. However, this assumption cannot be satisfied in the reality. During the process of examining the database, it was found that the Markov approach had better performance on the pavement condition prediction after an “abnormal” deterioration trend occurs because of its prediction concept. Although it is impossible to forecast the occurrence of an “abnormal” deterioration pattern, the capability on prediction amendment is still practical.

### 5.2 Future Study Recommendations

As for future study on pavement condition forecasting, based on the findings and research experience of this study, there are mainly three recommendations for the future study of pavement condition forecasting.

First, more reliable data would help on the prediction model simulation. During this study, quite a lot of time was spent on filtering the data that did not meet the standards of this study. Data filtering had been operated during both the process of data
arrangement and the process of results validations. There were several reasons for the needs of data filtering: (1) Not every pavement section was being recorded in a uniform way within the database, some pavement sections do not have Average ESAL records and the PCR records of some pavement sections are missing, etc.; (2) Some pavement sections might have received unrecorded treatments, which led to “abnormal” deterioration trends, such as a more than 10 points’ improvement on the PCR after a year, or presenting a steadily performance improvement over years; (3) There might still be some performance deterioration patterns which haven’t been discovered or focused on, for example, the synergy effects of heavy traffic and high temperature during the summer might bring about a sudden drop on the pavement performance. It should be stated that, although a perfect database might never exist or impractical to achieve as in this case, a relatively comprehensive or complete database will provide a better basis for obtaining more reliable prediction models.

Second, there is a need for a further level prediction performance evaluation by comparing the prediction results with actual PCR values of the matched pavement sections. In this study, only the predictions between different forecasting models were performed and the actual performance of the prediction models might still need to be evaluated by making comparison between the predicted results and actual PCR values. Apart from the fact that the comparison between predicted results and actual PCRs might take several years (at least 5 years since the comparisons between different prediction
models were set as 5 years in this study), another long term attention on the predicted results is also needed. Some attention needs to be paid to the change of pavement section when trying to make comparison between predicted results and actual future PCR values. During the process of observing and rearranging the database of this study, it was found that the partitions of pavement sections vary as the years go by. For example, pavement sections partitions in the 2012’s database could not match with the partitions of the 2013’s database. This is partially due to different partition manner as the year changes, and partially due to a setting of the ODOTPMIS which merges the nearby pavement sections when they show a similarity on the performance (for simplifying the prediction process of the program). Therefore, a comparison and modification on the pavement sections may be needed for performing prediction comparisons in the future.

Third, it is worthwhile to attempt to use polynomial regression or exponential regression on prediction models formulating in the future. In this study, linear regression fitting techniques were adopted on formulating the prediction models at state and district division level. It would be a valuable attempt to use other regression fitting techniques on simulating the prediction models and then to check if there is significant difference on the predictions between other fitting techniques and linear fitting technique, or to examine which regression technique would fit the actual PCR records better.
Reference


Appendix A

Distributions of Quantitative Variables Values

1. Average Annual Rainfall Records Distribution

Table A.1 Frequency Distribution of Average Annual Rainfall Amount

<table>
<thead>
<tr>
<th>Rainfall Amount (inches)</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>30-35</td>
<td>8</td>
</tr>
<tr>
<td>36-40</td>
<td>56</td>
</tr>
<tr>
<td>41-45</td>
<td>22</td>
</tr>
<tr>
<td>46-50</td>
<td>2</td>
</tr>
<tr>
<td>50+</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>88</td>
</tr>
</tbody>
</table>
Figure A-1 Average Annual Rainfall Amount Distribution (in percentage)
2. Average Annual Temperature Records Distribution

Table A.2 Frequency Distribution of Average Annual Temperature

<table>
<thead>
<tr>
<th>Temperature (°F)</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>46-48</td>
<td>1</td>
</tr>
<tr>
<td>49-50</td>
<td>20</td>
</tr>
<tr>
<td>51-52</td>
<td>42</td>
</tr>
<tr>
<td>53-54</td>
<td>21</td>
</tr>
<tr>
<td>54-56</td>
<td>4</td>
</tr>
<tr>
<td>56+</td>
<td>0</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>88</strong></td>
</tr>
</tbody>
</table>

Figure A-2 Average Annual Temperature Distribution (in percentage)
3. Average Annual Snowfall Records Distribution

Table A.3 Frequency Distribution of Average Annual Snowfall Amount

<table>
<thead>
<tr>
<th>Snowfall Amount (inches)</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-15</td>
<td>1</td>
</tr>
<tr>
<td>16-30</td>
<td>62</td>
</tr>
<tr>
<td>31-45</td>
<td>18</td>
</tr>
<tr>
<td>46-60</td>
<td>3</td>
</tr>
<tr>
<td>61-75</td>
<td>2</td>
</tr>
<tr>
<td>76-90</td>
<td>1</td>
</tr>
<tr>
<td>90+</td>
<td>1</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>88</strong></td>
</tr>
</tbody>
</table>

Figure A-3 Average Annual Snowfall Amount Distribution (in percentage)
4. Average ESAL Distribution

Table A.4 Frequency Distribution of Average ESAL

<table>
<thead>
<tr>
<th>Average ESAL</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>0~1k</td>
<td>15508</td>
</tr>
<tr>
<td>1k~2k</td>
<td>1223</td>
</tr>
<tr>
<td>2k~4k</td>
<td>870</td>
</tr>
<tr>
<td>4k~6k</td>
<td>391</td>
</tr>
<tr>
<td>6k~8k</td>
<td>90</td>
</tr>
<tr>
<td>8k~9k</td>
<td>1</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>18083</strong></td>
</tr>
</tbody>
</table>

Figure A-4 Average ESAL Distribution (in percentage)
5. Thickness Added Distribution

Table A.5 Frequency Distribution of Thickness Added

<table>
<thead>
<tr>
<th>Thickness Added (inches)</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>0~5</td>
<td>17259</td>
</tr>
<tr>
<td>5.1~10</td>
<td>296</td>
</tr>
<tr>
<td>10.1~15</td>
<td>37</td>
</tr>
<tr>
<td>15.1~20</td>
<td>78</td>
</tr>
<tr>
<td>20.1~25</td>
<td>9</td>
</tr>
<tr>
<td>25+</td>
<td>6</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>17685</strong></td>
</tr>
</tbody>
</table>

* NOTE: The total frequency number of variable “Thickness Added” is not equal to the number of variable “Average ESAL”, this is due to the missing of “Thickness Added” records in the original dataset before data filtering.

Figure A-5 Thickness Added Distribution (in percentage)
Appendix B

An Example of the Use of Weighted Correlation

Supposing there are 6 equi-length pavement sections which are called A, B, C, D, E and F. The expected PCR values of these sections are \( X = \{66, 99, 72, 81, 69, 86\} \), the actual PCR values of these sections are \( Y = \{68, 83, 63, 88, 68, 93\} \). Using the above equation, the correlation of the expected values and actual values is computed as 0.73.

If pavement section A is divided into 5 equi-length subsections while other sections remain the same, the dataset for expected values, actual values and the correlation are as follows:

\[
X' = \{66, 66, 66, 66, 99, 72, 81, 69, 86\}
\]
\[
Y' = \{68, 68, 68, 68, 68, 83, 63, 88, 68, 93\}
\]
\[
Corr(X', Y') = 0.79
\]

If pavement section B is divided into 5 equi-length subsections with other sections remains the same, the dataset for expected values, actual values and the correlation are as follow:
\[ X'' = \{66, 99, 99, 99, 72, 81, 69, 86\} \]
\[ Y'' = \{68, 83, 83, 83, 83, 63, 88, 68, 93\} \]
\[ Corr(X'', Y'') = 0.69 \]

It can be noticed through the above example that, the correlation value varies as the partitioning manner varies. When weighted correlation is applied by assigning weights to the items or variables in the calculation of correlation coefficient between the items or variables of two datasets, the previous correlation coefficient formula can be expressed by the following formula:

\[ \rho_{(X,Y)_{w}} = \frac{\sum w_i (X_i - \bar{X}_w)(Y_i - \bar{Y}_w)}{\sqrt{\sum w_i (X_i - \bar{X}_w)^2} \sqrt{\sum w_i (Y_i - \bar{Y}_w)^2}} \]

Where:

- \( w_i \) = the weight assigned to item \( X_i \),
- \( \bar{X}_w = \frac{\sum w_i X_i}{\sum w_i} \),
- \( \bar{Y}_w = \frac{\sum w_i Y_i}{\sum w_i} \).

Following the example, if all the pavement sections are equally divided, the \( w_i \) used in the weighted correlation formula would be 1, then the general correlation equation can be obtained. Back to the example, assuming the original length for each pavement section is 1. When section A, section B are divided into 5 equi-length subsections respectively, the weights for these two conditions would be shown as follows respectively:
$w' = \{0.2, 0.2, 0.2, 0.2, 0.2, 1, 1, 1, 1\}$

$w'' = \{1, 0.2, 0.2, 0.2, 0.2, 1, 1, 1, 1\}$

Using the above weights, the weighted correlations for those two conditions are computed as follows:

$Corr(X', Y')w = 0.73$

$Corr(X'', Y'')w = 0.73$

The above results coincided with the calculated correlation coefficient before partitioning. It indicates that the adoption of weighted correlation is suitable for the case of this study when the pavement sections are not uniformly divided by their lengths.
Appendix C

Another 11 District Divisions Prediction Models under Statistics Consulting

Table C.1 Coefficients of Models of District No. 2 after Statistics Consulting

<table>
<thead>
<tr>
<th>Variables</th>
<th>Coefficient Values</th>
<th>Std. Error</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>99.497</td>
<td>0.516</td>
<td>0.000</td>
</tr>
<tr>
<td>PaveSys</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>General</td>
<td>-2.44</td>
<td>0.022</td>
<td>0.000</td>
</tr>
<tr>
<td>Priority</td>
<td>-2.483</td>
<td>0.032</td>
<td>0.000</td>
</tr>
<tr>
<td>Urban</td>
<td>-2.798</td>
<td>0.033</td>
<td>0.000</td>
</tr>
<tr>
<td>PaveType</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FLEX</td>
<td>0.121</td>
<td>0.022</td>
<td>0.000</td>
</tr>
<tr>
<td>JRC</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Singularity*</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Average ESAL</td>
<td>NA**</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Average Temp</td>
<td>NA**</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Average Rain</td>
<td>NA**</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Average Snow</td>
<td>-0.043</td>
<td>0.015</td>
<td>0.005</td>
</tr>
<tr>
<td>R-square</td>
<td></td>
<td>0.741</td>
<td></td>
</tr>
</tbody>
</table>

* NOTE: There is only one pavement section record under this category

** NOTE: The variable is not taken into account in the prediction model
Table C.2 Coefficients of Models of District No. 3 after Statistics Consulting

<table>
<thead>
<tr>
<th>Variables</th>
<th>Coefficient Values</th>
<th>Std. Error</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>101.354</td>
<td>0.881</td>
<td>0.000</td>
</tr>
<tr>
<td>PaveSys</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>General</td>
<td>-3.144</td>
<td>0.021</td>
<td>0.000</td>
</tr>
<tr>
<td>Priority</td>
<td>-2.198</td>
<td>0.044</td>
<td>0.000</td>
</tr>
<tr>
<td>Urban</td>
<td>-3.032</td>
<td>0.023</td>
<td>0.000</td>
</tr>
<tr>
<td>PaveType</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FLEX</td>
<td>NA*</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>JRC</td>
<td>-1.007</td>
<td>0.154</td>
<td>0.000</td>
</tr>
<tr>
<td>Average ESAL</td>
<td>NA**</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Average Temp</td>
<td>NA**</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Average Rain</td>
<td>NA**</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Average Snow</td>
<td>-0.128</td>
<td>0.023</td>
<td>0.000</td>
</tr>
<tr>
<td>R-square</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* NOTE: “FLEX” of “Pavement type” is computed as a baseline parameter

** NOTE: The variable is not taken into account in the prediction model

Table C.3 Coefficients of Models of District No. 4 after Statistics Consulting

<table>
<thead>
<tr>
<th>Variables</th>
<th>Coefficient Values</th>
<th>Std. Error</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>101.013</td>
<td>4.803</td>
<td>0.000</td>
</tr>
<tr>
<td>PaveSys</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>General</td>
<td>-2.845</td>
<td>0.023</td>
<td>0.000</td>
</tr>
<tr>
<td>Priority</td>
<td>-2.752</td>
<td>0.025</td>
<td>0.000</td>
</tr>
<tr>
<td>Urban</td>
<td>-2.668</td>
<td>0.026</td>
<td>0.000</td>
</tr>
<tr>
<td>PaveType</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FLEX</td>
<td>-0.424</td>
<td>0.024</td>
<td>0.000</td>
</tr>
<tr>
<td>JRC</td>
<td>1.582</td>
<td>0.191</td>
<td>0.000</td>
</tr>
<tr>
<td>Average ESAL</td>
<td>NA*</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Average Temp</td>
<td>-0.184</td>
<td>0.084</td>
<td>0.029</td>
</tr>
<tr>
<td>Average Rain</td>
<td>0.138</td>
<td>0.030</td>
<td>0.000</td>
</tr>
<tr>
<td>Average Snow</td>
<td>-0.010</td>
<td>0.005</td>
<td>0.047</td>
</tr>
<tr>
<td>R-square</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* NOTE: The variable is not taken into account in the prediction model
### Table C.4 Coefficients of Models of District No. 5 after Statistics Consulting

<table>
<thead>
<tr>
<th>Variables</th>
<th>Coefficient Values</th>
<th>Std. Error</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>95.780</td>
<td>0.565</td>
<td>0.000</td>
</tr>
<tr>
<td>PaveSys</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>General</td>
<td>-3.058</td>
<td>0.023</td>
<td>0.000</td>
</tr>
<tr>
<td>Priority</td>
<td>-2.550</td>
<td>0.030</td>
<td>0.000</td>
</tr>
<tr>
<td>Urban</td>
<td>-3.208</td>
<td>0.029</td>
<td>0.000</td>
</tr>
<tr>
<td>PaveType</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FLEX</td>
<td>0.174</td>
<td>0.021</td>
<td>0.000</td>
</tr>
<tr>
<td>JRC</td>
<td>0.749</td>
<td>0.122</td>
<td>0.000</td>
</tr>
<tr>
<td>Average ESAL</td>
<td>NA*</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Average Temp</td>
<td>NA*</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Average Rain</td>
<td>NA*</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Average Snow</td>
<td>0.084</td>
<td>0.020</td>
<td>0.000</td>
</tr>
<tr>
<td>R-square</td>
<td></td>
<td></td>
<td>0.774</td>
</tr>
</tbody>
</table>

* NOTE: The variable is not taken into account in the prediction model

### Table C.5 Coefficients of Models of District No. 6 after Statistics Consulting

<table>
<thead>
<tr>
<th>Variables</th>
<th>Coefficient Values</th>
<th>Std. Error</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>99.822</td>
<td>0.652</td>
<td>0.000</td>
</tr>
<tr>
<td>PaveSys</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>General</td>
<td>-2.440</td>
<td>0.020</td>
<td>0.000</td>
</tr>
<tr>
<td>Priority</td>
<td>-1.737</td>
<td>0.018</td>
<td>0.000</td>
</tr>
<tr>
<td>Urban</td>
<td>-2.246</td>
<td>0.024</td>
<td>0.000</td>
</tr>
<tr>
<td>PaveType</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FLEX</td>
<td>0.120</td>
<td>0.019</td>
<td>0.000</td>
</tr>
<tr>
<td>JRC</td>
<td>1.789</td>
<td>0.375</td>
<td>0.000</td>
</tr>
<tr>
<td>Average ESAL</td>
<td>NA*</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Average Temp</td>
<td>NA*</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Average Rain</td>
<td>NA*</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Average Snow</td>
<td>-0.098</td>
<td>0.024</td>
<td>0.000</td>
</tr>
<tr>
<td>R-square</td>
<td></td>
<td></td>
<td>0.685</td>
</tr>
</tbody>
</table>

* NOTE: The variable is not taken into account in the prediction model
### Table C.6 Coefficients of Models of District No. 7 after Statistics Consulting

<table>
<thead>
<tr>
<th>Variables</th>
<th>Coefficient Values</th>
<th>Std. Error</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>96.777</td>
<td>0.064</td>
<td>0.000</td>
</tr>
<tr>
<td>PaveSys</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>General</td>
<td>-2.052</td>
<td>0.019</td>
<td>0.000</td>
</tr>
<tr>
<td>Priority</td>
<td>-1.852</td>
<td>0.031</td>
<td>0.000</td>
</tr>
<tr>
<td>Urban</td>
<td>-2.035</td>
<td>0.022</td>
<td>0.000</td>
</tr>
<tr>
<td>PaveType</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FLEX</td>
<td>-0.109</td>
<td>0.018</td>
<td>0.000</td>
</tr>
<tr>
<td>JRC</td>
<td>1.104</td>
<td>0.350</td>
<td>0.002</td>
</tr>
<tr>
<td>Average ESAL</td>
<td>NA*</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Average Temp</td>
<td>NA*</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Average Rain</td>
<td>NA*</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Average Snow</td>
<td>NA*</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>R-square</td>
<td></td>
<td></td>
<td>0.664</td>
</tr>
</tbody>
</table>

* NOTE: The variable is not taken into account in the prediction model

### Table C.7 Coefficients of Models of District No. 8 after Statistics Consulting

<table>
<thead>
<tr>
<th>Variables</th>
<th>Coefficient Values</th>
<th>Std. Error</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>95.255</td>
<td>1.009</td>
<td>0.000</td>
</tr>
<tr>
<td>PaveSys</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>General</td>
<td>-2.080</td>
<td>0.023</td>
<td>0.000</td>
</tr>
<tr>
<td>Priority</td>
<td>-1.398</td>
<td>0.024</td>
<td>0.000</td>
</tr>
<tr>
<td>Urban</td>
<td>-2.067</td>
<td>0.024</td>
<td>0.000</td>
</tr>
<tr>
<td>PaveType</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FLEX</td>
<td>-0.174</td>
<td>0.021</td>
<td>0.000</td>
</tr>
<tr>
<td>JRC</td>
<td>0.749</td>
<td>0.193</td>
<td>0.000</td>
</tr>
<tr>
<td>Average ESAL</td>
<td>NA*</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Average Temp</td>
<td>NA*</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Average Rain</td>
<td>0.078</td>
<td>0.024</td>
<td>0.001</td>
</tr>
<tr>
<td>Average Snow</td>
<td>NA*</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>R-square</td>
<td></td>
<td></td>
<td>0.700</td>
</tr>
</tbody>
</table>

* NOTE: The variable is not taken into account in the prediction model
Table C.8 Coefficients of Models of District No. 9 after Statistics Consulting

<table>
<thead>
<tr>
<th>Variables</th>
<th>Coefficient Values</th>
<th>Std. Error</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>95.716</td>
<td>0.812</td>
<td>0.000</td>
</tr>
<tr>
<td>PaveSys</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>General</td>
<td>-1.959</td>
<td>0.026</td>
<td>0.000</td>
</tr>
<tr>
<td>Priority</td>
<td>-1.888</td>
<td>0.038</td>
<td>0.000</td>
</tr>
<tr>
<td>Urban</td>
<td>-2.188</td>
<td>0.032</td>
<td>0.000</td>
</tr>
<tr>
<td>PaveType</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FLEX</td>
<td>NA*</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>JRC</td>
<td>0.618</td>
<td>0.135</td>
<td>0.000</td>
</tr>
<tr>
<td>Average ESAL</td>
<td>-0.001</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Average Temp</td>
<td>NA**</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Average Rain</td>
<td>0.076</td>
<td>0.020</td>
<td>0.000</td>
</tr>
<tr>
<td>Average Snow</td>
<td>NA**</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>R-square</td>
<td></td>
<td></td>
<td>0.721</td>
</tr>
</tbody>
</table>

* NOTE: “FLEX” of “Pavement type” is computed as a baseline parameter

** NOTE: The variable is not taken into account in the prediction model

Table C.9 Coefficients of Models of District No. 10 after Statistics Consulting

<table>
<thead>
<tr>
<th>Variables</th>
<th>Coefficient Values</th>
<th>Std. Error</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>100.565</td>
<td>0.351</td>
<td>0.000</td>
</tr>
<tr>
<td>PaveSys</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>General</td>
<td>-2.290</td>
<td>0.025</td>
<td>0.000</td>
</tr>
<tr>
<td>Priority</td>
<td>-2.128</td>
<td>0.036</td>
<td>0.000</td>
</tr>
<tr>
<td>Urban</td>
<td>-1.896</td>
<td>0.045</td>
<td>0.000</td>
</tr>
<tr>
<td>PaveType</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FLEX</td>
<td>-0.259</td>
<td>0.024</td>
<td>0.000</td>
</tr>
<tr>
<td>JRC</td>
<td>0.676</td>
<td>0.184</td>
<td>0.000</td>
</tr>
<tr>
<td>Average ESAL</td>
<td>NA*</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Average Temp</td>
<td>NA*</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Average Rain</td>
<td>NA*</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Average Snow</td>
<td>-0.098</td>
<td>0.016</td>
<td>0.000</td>
</tr>
<tr>
<td>R-square</td>
<td></td>
<td></td>
<td>0.765</td>
</tr>
</tbody>
</table>

* NOTE: The variable is not taken into account in the prediction model
### Table C.10 Coefficients of Models of District No. 11 after Statistics Consulting

<table>
<thead>
<tr>
<th>Variables</th>
<th>Coefficient Values</th>
<th>Std. Error</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>94.886</td>
<td>0.741</td>
<td>0.000</td>
</tr>
<tr>
<td>PaveSys</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>General</td>
<td>-2.710</td>
<td>0.026</td>
<td>0.000</td>
</tr>
<tr>
<td>Priority</td>
<td>-1.846</td>
<td>0.035</td>
<td>0.000</td>
</tr>
<tr>
<td>Urban</td>
<td>-2.957</td>
<td>0.061</td>
<td>0.000</td>
</tr>
<tr>
<td>PaveType</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FLEX</td>
<td>-0.425</td>
<td>0.026</td>
<td>0.000</td>
</tr>
<tr>
<td>JRC</td>
<td>-0.482</td>
<td>0.175</td>
<td>0.006</td>
</tr>
<tr>
<td>Average ESAL</td>
<td>NA*</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Average Temp</td>
<td>NA*</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Average Rain</td>
<td>0.086</td>
<td>0.019</td>
<td>0.000</td>
</tr>
<tr>
<td>Average Snow</td>
<td>NA*</td>
<td>NA</td>
<td>NA</td>
</tr>
</tbody>
</table>

R-square: 0.760

*NOTE: The variable is not taken into account in the prediction model*

### Table C.11 Coefficients of Models of District No. 12 after Statistics Consulting

<table>
<thead>
<tr>
<th>Variables</th>
<th>Coefficient Values</th>
<th>Std. Error</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>100.627</td>
<td>1.005</td>
<td>0.000</td>
</tr>
<tr>
<td>PaveSys</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>General</td>
<td>-3.008</td>
<td>0.041</td>
<td>0.000</td>
</tr>
<tr>
<td>Priority</td>
<td>-2.227</td>
<td>0.032</td>
<td>0.000</td>
</tr>
<tr>
<td>Urban</td>
<td>-2.657</td>
<td>0.028</td>
<td>0.000</td>
</tr>
<tr>
<td>PaveType</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FLEX</td>
<td>0.147</td>
<td>0.038</td>
<td>0.000</td>
</tr>
<tr>
<td>JRC</td>
<td>0.662</td>
<td>0.168</td>
<td>0.000</td>
</tr>
<tr>
<td>Average ESAL</td>
<td>-0.001</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Average Temp</td>
<td>NA*</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Average Rain</td>
<td>-0.093</td>
<td>0.024</td>
<td>0.000</td>
</tr>
<tr>
<td>Average Snow</td>
<td>NA*</td>
<td>NA</td>
<td>NA</td>
</tr>
</tbody>
</table>

R-square: 0.727

*NOTE: The variable is not taken into account in the prediction model*