

GIS-BASED PROBABILISTIC APPROACH FOR ASSESSING AND ENHANCING
INFRASTRUCTURE DATA QUALITY

A Dissertation

by

SIAMAK SALIMINEJAD

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Approved by:

Chair of Committee, Nasir Gharaibeh
Committee Members, Clifford Spiegelman
Robert Lytton
Roger Smith
Head of Department, John Niedzwecki

December 2012

Major Subject: Civil Engineering

Copyright 2012 Siamak Saliminejad

ABSTRACT

The task of preserving and improving infrastructure systems is becoming extremely challenging because these systems are decaying due to aging and over utilization, have limited funding, and are complex in nature (geographically spread, and affect and are affected by technological, environmental, social, security, political, and economic factors). The infrastructure management paradigm has emerged to assist in the challenging task of managing infrastructure systems in a systematic and cost-effective manner. Infrastructure management is a data-driven process. It relies on large databases that contain information on the system's inventory, condition, maintenance and rehabilitation (M&R) history, utilization, and cost. This data feeds into analytical models that assess infrastructure current conditions, predict future conditions, and develop optimal M&R strategies. Thus, complete and accurate data is essential to a reliable infrastructure management system.

This study contributes to advancing the infrastructure management paradigm (with focus on pavement management) in two primary ways: (a) it provides in-depth understanding of the impact of errors in condition data on the outputs of infrastructure management systems, and (b) it provides efficient computational methods for improving infrastructure data quality. First, this research provides a quantitative assessment of the effects of error magnitude and type (both systematic and random) in pavement condition data on the accuracy of PMS outputs (i.e., forecasted needed budget and M&R activities in a multi-year planning period). Second, a new technique for detecting gross outliers and pseudo outliers in pavement condition data was developed and tested. Gross outliers

are data values that are likely to be erroneous, whereas pseudo outliers are pavement sections performing exceptionally well or poor due to isolated local conditions. Third, a new technique for estimating construction and M&R history data from pavement condition data was developed and tested. This technique is especially beneficial when M&R data and condition data are stored in disparate heterogeneous databases that are difficult to integrate (i.e., legacy databases).

The main merit of the developed techniques is their ability to integrate methods and principles from Bayesian and spatial statistics, GIS, and operations research in an efficient manner. The application of these techniques to a real-world cases study (pavement network in Bryan district) demonstrated the potential benefits of these techniques to infrastructure managers and engineers.

DEDICATION

This thesis is dedicated to my family. Thanks for your endless support.

ACKNOWLEDGEMENTS

I would like to thank my committee chair, Dr. Gharaibeh, and my committee members, Dr. Lytton, Dr. Spiegelman, and Dr. Smith for their guidance and support throughout the course of this research.

Thanks also go to my friends and colleagues and the department faculty and staff for making my time at Texas A&M University a great experience. I also want to extend my gratitude to the Texas Transportation Institute and Texas Department of Transportation, which provided the data and funding required for this research.

Finally, I would like to thank my wife, my mother and my father for their encouragement, patience and love.

NOMENCLATURE

AASHTO	American Association of States Highway and Transportation Officials
AADT	Average Annual Daily Traffic
ACP	Asphalt Concrete Pavement
ANN	Artificial Neural Network
AUPC	Area Under Performance Curve
BCR	Benefit to Cost Ratio
BDP	Budget Deviation Percentage
BCA	Benefit Cost Analysis
CI	Condition Index
DN	Do Nothing
DS	Distress Score
ESAL	Equivalent Single Axle Load
FHWA	Federal Highway Administration
GA	Genetic Algorithm
GIS	Geographic Information System
HERS	Highway Economic Requirements System
HR	Heavy Rehabilitation
IBCA	Incremental Benefit - Cost Analysis
IQR	Interquartile Range
IRI	International Roughness Index

LR	Light Rehabilitation
LTPP	Long-Term Pavement Performance
M&R	Maintenance and Rehabilitation
MR	Medium Rehabilitation
NIST	National Institute of Standards and Technology
PCI	Pavement Condition Index
PCR	Pavement Condition Rating
PM	Preventive Maintenance
PMIS	Pavement Management Information System
PMS	Pavement Management System
POI	Point Outlierness Index
SOI	Section Outlierness Index
TxDOT	Texas Department of Transportation
VMT-A	Vehicle-Miles Traveled on Adequate pavement

TABLE OF CONTENTS

	Page
ABSTRACT	ii
DEDICATION	iv
ACKNOWLEDGEMENTS	v
NOMENCLATURE.....	vi
TABLE OF CONTENTS	viii
LIST OF FIGURES.....	x
LIST OF TABLES	xi
1. INTRODUCTION.....	1
Motivation and Problem Statement.....	1
Research Questions	3
Research Hypothesis	4
Research Objectives and Methodology.....	5
Organization of the Dissertation	7
2. LITERATURE REVIEW.....	8
Overview of Pavement Management	8
Pavement Management Data Quality Issues.....	9
Effects of Data Quality on Complex Systems.....	16
Impact of Pavement Condition Data Quality on PMS Output.....	18
Effect of Data Quality on Pavement Condition Prediction.....	19
Effect of Data Quality on Needs Analysis	21
Existing Outlier Detection Methods.....	23
3. TEST-BENCH PAVEMENT MANAGEMENT SYSTEM.....	26
Condition Assessment.....	26
M&R Assignment	28
Condition Prediction Models	28
Data Storage and Management	33
Pavement Condition Data Quality.....	34

	Page
4. EVALUATING THE IMPACT OF CONDITION DATA QUALITY ON NETWORK-LEVEL PAVEMENT MANAGEMENT SYSTEMS.....	36
Error Propagation Simulation Process	36
Results and Discussion.....	41
Section Summary and Concluding Remarks.....	54
5. A PROXIMITY-BASED TECHNIQUE FOR DETECTING OUTLIERS IN PAVEMENT CONDITION DATA	57
Developed Outlier Detection Method	57
Application and Validation of the Developed Method	71
Impact of Gross Outliers	92
Section Summary and Concluding Remarks.....	93
6. A SPATIAL-BAYESIAN TECHNIQUE FOR IMPUTING PAVEMENT NETWORK REPAIR DATA	95
Developed Data Imputation Method.....	95
Application of the Developed Technique.....	105
Results and Discussion.....	113
Section Summary and Concluding Remarks.....	122
7. SUMMARY, CONTRIBUTION, CONCLUSIONS, AND RECOMMENDATIONS.....	123
Contribution	124
Conclusions	124
Recommendations	127
REFERENCES.....	129
APPENDIX A: CODE OF COMPUTER PROGRAM DEVELOPED FOR DETECTING PROXIMITY-BASED OUTLIERS IN PAVEMENT CONDITION DATA.....	149
APPENDIX B: AUDIT AND ORIGINAL RATINGS OF NORMAL DATA	196
APPENDIX C: CODE OF COMPUTER PROGRAM DEVELOPED FOR IMPUTING PAVEMENT NETWORK REPAIR DATA	228

LIST OF TABLES

	Page
Table 1 Examples of treatment types for ACP	28
Table 2 Descriptive statistics of audited DS data	87
Table 3 Comparison between audit and original ratings of gross outliers	88
Table 4 Mean and standard deviation of Δ DS for the three comparisons (audit versus gross outliers, audit versus normal data, and audit versus original data)	89
Table 5 Possible explanations for detected pseudo outliers	91
Table 6 PMS outputs before and after omitting outliers	93
Table 7 Numerical example of key calculations of the developed technique.	113

LIST OF FIGURES

	Page
Figure 1 Primary components and functions of PMS	3
Figure 2 Data clustering used in a classic outlier detection method	29
Figure 3 Climate and subgrade zones for ACP performance prediction model calibration.....	30
Figure 4 S-shaped TxDOT pavement condition prediction model	32
Figure 5 Condition prediction models for flexile pavement sections with wet and warm climates, poor subgrade condition and medium traffic	33
Figure 6 Pavement sections with an increased DS from 2009 to 2010 (Bryan District).....	35
Figure 7 Process developed to simulate the propagation of pavement condition data error to the output of a PMS needs analysis	37
Figure 8 Concept of area under a performance curve	40
Figure 9 Effects of systematic errors in pavement condition data on errors in the estimated average annual budget.....	43
Figure 10 Effects of systematic errors in pavement condition data on errors in estimated (a) portions of the network that need no M&R, (b) portions of the network that need maintenance, and (c) portions of the network that need rehabilitation	45
Figure 11 Pavement network M&R needs in the presence of different levels of systematic error in conditions data (a) 2011, (b) 2012, (c) 2013, and (d) 2014	47
Figure 12 Effects of random errors in pavement condition data on errors in the estimated average annual budget.....	49
Figure 13 Effects of random errors in pavement condition data on errors in estimated (a) portions of the network in need of “do nothing,” (b) portions of the network in need of maintenance, and (c) portions of the network in need of rehabilitation	50

	Page
Figure 14 Pavement network M&R needs in the presence of different levels of random errors in condition data (a) 2011, (b) 2012, (c) 2013, and (d) 2014.....	52
Figure 15 Percentage of network sections with assigned M&R types changed as a result of a DS random error with $\sigma=5$	53
Figure 16 Clustering a roadway corridor with n sections into k clusters	59
Figure 17 The first clustering scenario.....	60
Figure 18 (a) Condition data points associated with the sections of Clustering Scenario 1 (b) PDFs fitted to the points of each cluster.....	61
Figure 19 The second clustering scenario	61
Figure 20 (a) the data points associated with the sections (b) PDFs fitted to the points of each cluster.....	62
Figure 21 Outlier detection criteria	68
Figure 22 Relative frequency distribution of the length of the M&R projects in 2010 in the Bryan district	72
Figure 23 Example of detected outliers (Highway 36 in Bryan district)	73
Figure 24 Geographic locations of pavement sections with medium traffic, pavement family A, and last treatment type being preventive maintenance, in the Bryan district in 2011 (classic clustering method).....	74
Figure 25 Location of highway 36 on Bryan District map	75
Figure 26 Cluster 1 consisting of eight pavement sections, (a) Historical condition data of sections (b) Location on map	76
Figure 27 Cluster 2 consisting of seven pavement sections, (a) Historical condition data of sections (b) Location on map	76
Figure 28 Cluster 3 consisting of seven pavement sections, (a) Historical condition data of sections (b) Location on map	77

	Page
Figure 29 Cluster 4 consisting of seven pavement sections, (a) Historical condition data of sections (b) Location on map	77
Figure 30 Cluster 5 consisting of nine pavement sections, (a) Historical condition data of sections (b) Location on map	78
Figure 31 Cluster 6 consisting of seven pavement sections, (a) Historical condition data of sections (b) Location on map	78
Figure 32 Cluster 7 consisting of nine pavement sections, (a) Historical condition data of sections (b) Location on map	79
Figure 33 Cluster 8 consisting of seven pavement sections, (a) Historical condition data of sections (b) Location on map	79
Figure 34 Cluster 9 consisting of nine pavement sections, (a) Historical condition data of sections (b) Location on map	80
Figure 35 Cluster 10 consisting of seven pavement sections, (a) Historical condition data of sections (b) Location on map	80
Figure 36 Cluster 11 consisting of eight pavement sections, (a) Historical condition data of sections (b) Location on map	81
Figure 37 Cluster 12 consisting of eight sections, (a) Historical condition data of sections (b) Location on map	81
Figure 38 Cluster 13 consisting of seven sections, (a) Historical condition data of sections (b) Location on map	82
Figure 39 Classic outlier detection method.....	83
Figure 40 Outliers detected by the classic method versus outliers detected by the proposed spatial-temporal method in Bryan District in 2011	85
Figure 41 Examples of false positive outliers detected by the classic method	86
Figure 42 Examples of false negative outliers detected by the classic method	86
Figure 43 Data used to calculate the probability that a repair action has been performed on a group of L adjacent pavement sections.....	97

	Page
Figure 44 Six possible boundaries for a project spanning over 5 sections of a roadway that consists of 10 sections	98
Figure 45 Condition index semivariogram for Highway SH- 79 in 2010.....	102
Figure 46 Condition index semivariogram for Highway I- 45 in 2010.....	103
Figure 47 Sections with improved condition in Bryan District from 2009 to 2010.....	107
Figure 48 Noisy temporal variation in DS of an example pavement section (Brazos County, Roadway FM0060 starting at 0636 +00.5).....	108
Figure 49 P(R L): Percentage of projects with various length values	109
Figure 50 P(DS): Percentage of sections with each DS value	110
Figure 51 P(DS R): Percentage of repaired sections with each DS value.....	111
Figure 52 P(Δ DS,R): Percentage of repaired sections with each Δ value in each DS category	111
Figure 53 P(Δ DS): Percentage of sections with each Δ value in each DS category	112
Figure 54 Estimated M&R projects (dark segments) versus actual M&R projects (thick gray segments) in Bryan District	115
Figure 55 Sensitivity of the accuracy and precision of the developed Method to the NRP input parameter	116
Figure 56 M&R projects estimated solely based on improvement in DS on a section-by-section basis (dark segments) versus actual M&R projects (thick gray segments) in Bryan District.....	118
Figure 57 Estimating the precision of spatial-Bayesian and section-by-section methods versus standard deviation of error in pavement condition.....	119
Figure 58 Predicted versus actual repaired sections for $\sigma=0$: (a) using the spatial-Bayesian method (b) based on DS increases.....	121
Figure 59 Predicted versus actual repaired sections for $\sigma=5$: (a) using the spatial-Bayesian method (b) based on DS increases.....	121

	Page
Figure 60 Predicted versus actual repaired sections for $\sigma=10$: (a) using the spatial-Bayesian method (b) based on DS increases.....	121
Figure 61 Predicted versus actual repaired sections for $\sigma=15$: (a) using the spatial-Bayesian method (b) based on DS increases.....	122

1. INTRODUCTION

Motivation and Problem Statement

The task of preserving and improving infrastructure systems is becoming extremely challenging because these systems are decaying due to aging and over utilization, have limited funding, and are complex in nature (geographically spread, and affect and are affected by technological, environmental, social, security, political, and economic factors) (Heintz et al. 2009). The resources invested to preserve and improve these systems are extraordinary. For example, the Federal Highway Administration SAFETEA-LU program has authorized \$25.2 billion to preserve the interstate highway system during the period of 2005-2009 (FHWA, 2012).

The infrastructure management paradigm has emerged to assist in the challenging task of managing infrastructure systems in a systematic and cost-effective manner. It involves the systematic collection of infrastructure data and the application of engineering, economic, and decision making principles and analytical tools such as life-cycle cost analysis and risk assessment (Hudson et al. 1997). Over the past four decades, many computerized engineered management systems have been developed to apply the concepts and methods of infrastructure management, including pavement management systems (PMS), bridge management systems (BMS), and integrated management systems that address multiple infrastructure asset classes (Hudson 1997).

Infrastructure management is a data-driven process. It relies on large databases that contain information on the system's inventory, condition, treatment history, utilization, and cost. This data feeds into analytical models that assess infrastructure

current conditions (e.g., aggregating distress data into condition indexes), predict future conditions, and develop optimal maintenance and rehabilitation (M&R) strategies (Deshpande et al. 2010; Shahin 2005). Thus, complete and accurate data is essential to a reliable Infrastructure management system.

Generally, the process of implementing an infrastructure management system involves four phases: (1) data collection, (2) data processing and quality management, (3) analysis, and (4) reporting (see Figure 1). Much of past research has focused on data collection (processes, sensing technologies, etc.) as well as data analysis (performance modeling, optimization, etc.) (Shahin 2005); little work has been done to help understand the impact of data quality on the outputs of infrastructure management systems and to develop efficient methods for assessing and improving data quality. This dissertation aims to fill this gap in the literature, with focus on pavement management systems.

While there is a general agreement in the literature that the quality of condition data affects the output of PMS (Buchheit et al. 2005; Livneh 1994; McNeil and Humplick 1991; Ong and Sinha 2008; Shekharan et al. 2007); no major research effort has been done to measure the impact of data quality on the output of infrastructure management systems. Additionally, the assessment of pavement condition data quality has been limited to characterizing the variability in data collection processes (Migliaccio et al. 2011). This dissertation addresses these issues by first assessing the impact of errors in pavement condition data on the output of pavement management systems, and then developing and testing new computational methods for detecting likely errors,

identifying pavement sections with abnormal (yet valid) behavior, and imputing M&R work history data.

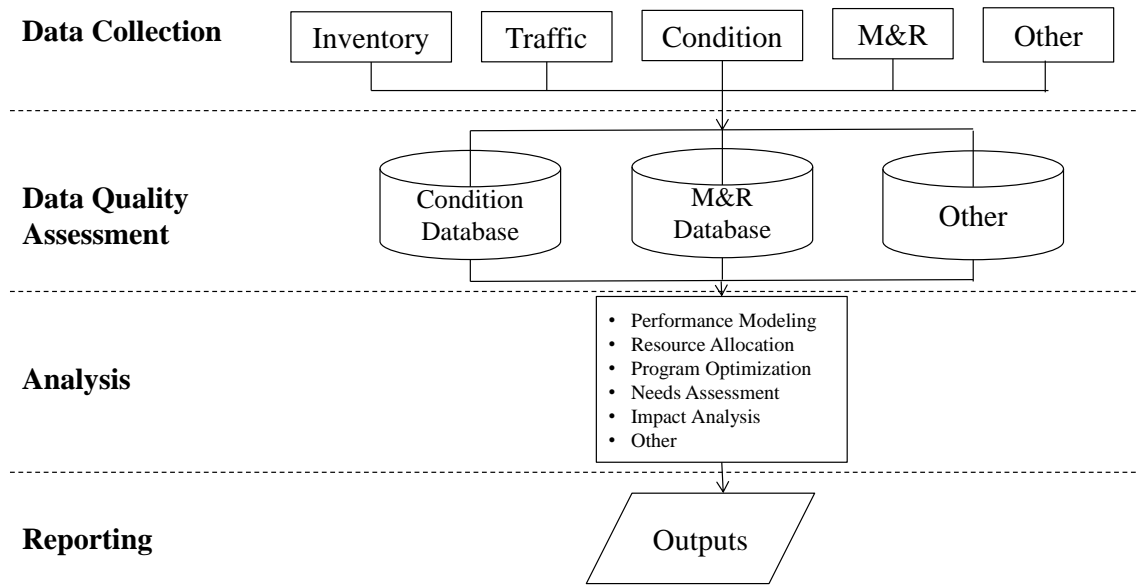


Figure 1 Primary components and functions of PMS

Research Questions

Based on the aforementioned observations, this study aims to address the following specific questions:

- How does the quality of pavement condition data impact PMS output?
- How can we detect likely errors (called gross outliers) in large datasets of pavement network condition data?
- How can we detect statistically dissimilar, but likely true values (called pseudo outliers) in large datasets of pavement network condition data?

- Can we estimate missing treatment history for a pavement network by detecting spatiotemporal patterns in condition data?

Research Hypothesis

The overarching hypothesis used in this thesis to answer the above questions is that incorporating geographic proximity and condition history into outlier detection methods and data estimation methods can lead to more efficient computational methods for assessing and improving the quality of pavement data. This hypothesis is based on the observation that both pavement performance and pavement data collection processes are affected by local factors. Pavement performance is affected by local factors such as soil type, drainage, and construction variations (AASHTO 2002). Pavement condition data collection processes are affected by local factors such as rater, time of inspection, and roadway geometry (Flintsch 2009). To test this hypothesis, a new method for detecting gross outliers using geographic proximity and condition history was developed and compared to traditional methods (see Section 5). In addition, geographic proximity and condition history data were further used for imputing M&R data, which was then compared to classic methods that do not use geographic data (see Section 6).

The research question of how the quality of pavement condition data affects network-level PMS outputs is investigated by simulating a standard pavement management system. Systematic and random errors were then introduced into the condition dataset of a case study (i.e., the Bryan district of TxDOT), and the impact of condition data quality on PMS outputs was investigated. This process is explained in Section 4.

Research Objectives and Methodology

The specific objectives of this research and the tasks completed to fulfill them are defined as follows:

Assess the Impact of Pavement Condition Data Quality on the Accuracy of PMS Needs Estimates

This work attempts to provide a deeper understanding of the impact of data quality on PMS outputs (i.e., the forecasting of necessary budgets and M&R projects for a multi-year planning period), including interactions, propagation of involved errors and the sensitivity of PMS outputs to the levels and types of errors in condition data. Systematic and random errors were introduced into the condition dataset of the case study (i.e., the Bryan district) to perform these investigations. To accomplish this objective, the following tasks have been carried out:

- The process of a PMS needs analysis and project prioritization was computerized.
- Needs analysis and M&R project prioritization were run on condition datasets with varying levels and types of error.

Develop a Technique for Detecting Local Outliers in Pavement Condition Datasets

The outlier detection method is comprised of two steps: first, neighboring pavement sections in close proximity are clustered based on similarities in historical condition data and prior information on M&R project lengths; second, gross and pseudo outliers in each cluster are detected based on the Interquartile Range (IQR) of the pavement condition indicator. Gross and pseudo outlier points are defined as follows:

- Pseudo outlier: If an outlier point belongs to a section that has a different condition history from the other sections within its cluster, the section is likely to be affected by different local factors (e.g., drainage condition, subgrade condition, and M&R histories) and is less likely to be erroneous. This point is classified as a pseudo outlier.
- Gross outlier: If an outlier point belongs to a section with a similar condition history to that of other sections within its cluster that point is likely to be an erroneous observation and is classified as a gross outlier.

To accomplish this objective, the following tasks were carried out:

- A method for clustering pavement sections was developed, based on spatiotemporal patterns in the condition data.
- A method for detecting local outliers was developed within each pavement cluster.
- The segmentation and outlier detection methods were integrated to form a local outlier detection technique for pavement condition data.
- The developed local outlier detection technique was tested and validated using the Bryan district case study.

Develop a Technique for Imputing Missing Pavement Repair History Data

The developed GIS-based technique searches for evidence of repair in the condition data of groups of adjacent pavement sections. Traditionally, the focus has been on temporal patterns in condition data (i.e., pavement conditions versus age relationships) which tend to be highly variable. To accomplish this objective, the following tasks were carried out:

- A technique was developed for imputing pavement construction and M&R history in legacy PMS databases.
- The developed data imputation technique was tested and validated using the Bryan district case study.

Organization of the Dissertation

This dissertation is composed of seven sections. Information from the literature on relevant topics are summarized and provided in Section 2. Then, the test-bench PMS used for evaluating the impact of the quality of pavement condition data on PMS outputs and for testing the techniques developed to improve this data is described in Section 3. A quantitative assessment of the impacts of error magnitude and type (both systematic and random) in pavement condition data on the accuracy of PMS outputs is presented in Section 4. Section 5 explains the development of a new technique for detecting gross and pseudo outliers in pavement condition data. The developed technique integrates concepts and methods from GIS, Bayesian statistics and Genetic Algorithms to identify data values that are likely to be erroneous, as well as pavement sections with unusual pavement performance.

The development of a new technique for estimating construction and M&R history data from pavement condition data is presented in Section 6. The developed technique employs Bayesian and spatial statistics in a GIS environment to search for evidence of repair in the condition data (current and past) of group of adjacent pavement sections. Finally, summary, contributions, conclusions and recommendations are presented in Section 7.

2. LITERATURE REVIEW

This section presents a summary of the relevant information provided from the current literature on this topic. First, an overview of pavement management is provided, and pavement management data quality issues are discussed; then, the factors influencing the quality of pavement condition data are identified. Next, the effects of data quality on complex systems are explained. Later, the more specific effects of pavement condition data quality on PMS outputs, condition prediction and needs analysis are discussed. Finally, an overview of outlier data and existing methods for detecting it are provided.

Overview of Pavement Management

Pavement management is a management approach used by personnel within a given agency to make informed and cost-effective decisions. The set of tools used to assist decision-makers in this process is called PMS (AASHTO 2001).

Pavement management can cover all phases of pavement planning, programming, analysis, design, construction, and research (Smith 2002). Different PMS's often have functionalities in common, such as a network inventory, pavement condition assessment, pavement performance predictions, and management planning methods. However, PMSs may take many different forms, based upon the particular organization's priorities (Mooney et al. 2005).

Pavement management is generally described and used in two levels: the network level and the project level. These two levels differ in terms of both their application and the data collected. A project-level PMS assesses the cause of deterioration, recommends the most cost-effective treatment for individual projects, and performs a site-specific

analysis. Thus, a project-level analysis requires detailed data (such as distresses, materials, and traffic). On the other hand, a network-level PMS analysis deals with the entire pavement network (or a large part of it) at once (Shahin 2005). The main purpose of a network-level analysis is to assess the effects of various funding scenarios on the “health” of the entire system. The basic components of a network-level PMS are: (1) inventory data collection, (2) conditions assessment, (3) a determination of needs, (4) the prioritization of projects needing M&R, (5) a method of determining the impact of funding decisions, and (6) a feedback process (Hudson et al. 1997).

Some agencies use additional levels to support the decision-making process. Some examples of these additional levels include: (1) a strategic level in which pavement information is integrated with the information from other management systems in conjunction with the use of asset management concepts (Flintsch 2004a); (2) a research level in which information provided by the PMS affects the undergoing research on current PMS practices, such as performance predictions and project prioritization methods (AASHTO 2001); and (3) a project-selection level used for identifying constraints, refining alternative projects, and improving cost estimates (Gurganus and Gharaibeh 2012). This thesis addresses issues of data quality for network-level pavement management.

Pavement Management Data Quality Issues

Pavement management is a data-driven process that involves inventory and monitoring data collection (e.g., pavement structure, construction history, traffic, etc.), condition assessment (e.g., aggregation of distress data into condition indexes), condition

prediction, and planning of optimal M&R strategies (Deshpande et al. 2010; Shahin 2005). Thus, complete and accurate data is essential for a reliable PMS. Unfortunately, the process of preparing and integrating accurate and complete pavement management data sets historically has been a major challenge for most highway agencies around the world.

In general, no standard and rigorous definition of data quality exists. In fact, the quality of data generally depends upon its application (Flintsch and McGhee 2009). However, accuracy, precision, and completeness have all been mentioned as characteristics of data quality (Flintsch and McGhee 2009; Ong et al. 2010; Wand and Wang 1996). The Transportation Research Board (TRB) defines accuracy as “the degree to which a measurement, or the mean of a distribution of measurements, tends to coincide with the true population mean” and precision as “the degree to which tests or measurements on identical samples tend to produce the same results” (TRB 2009). National Institute of Standards and Technology (NIST) on the other hand defines accuracy as "the difference between a measurement reading and the true value of that measurement" and precision as "the ability of a process to repeat the same accurate measurement over time (NIST, 2003). Completeness is often defined as “a measurable error of omission observed between the database and the specification” (Brassel et al., 1995).

Variability in Pavement Condition Data

There is inherent variability in pavement condition data; regardless of the data collection method [i.e., automated (Lajnef et al. 2011; Ying and Salari 2010) and manual

(Migliaccio et al. 2011)]. There are at least two reasons for this inherent variability in pavement condition data. First, in contrast to many other types of data, there is no unique and objective definition of “true value” with regards to pavement distress measurements. Since the distress definitions and, as a result, the measurement processes are partially subjective and somewhat ambiguous; most agencies opt to use the ratings of trained and experienced raters as reference values (Flintsch and McGhee 2009). This variability in true value contributes to the distress data variability. Rada (1997), for example, reports the ambiguity between longitudinal and low severity alligator cracking as one of the largest sources of variability in Long-Term Pavement Performance (LTPP) data.

The second reason for such variability is errors in measurement. Generally, measurement errors are categorized as either random or systematic (though it is sometimes difficult to distinguish between the two), and many errors are a combination of both (Grabe 2010). Random errors are caused by irregular causes where the laws of action are either unknown or too complex to be investigated. They are disordered in incidence, variable in magnitude, and positive and negative values occur in repeated measurements in no detectable order. Systematic errors can originate from the rater, the instrument, or the technology used. Systematic errors may be constant or may vary in some regular way (e.g., in proportion or by a percentage) to the actual value of the measured quantity, or even to the value of a different quantity (Shekharan 1998). Systematic errors have a higher impact on the PMS output than random errors because they can affect a larger number of sections (Flintsch and McGhee 2009; Shekharan et al. 2007).

Random errors, as well as variable systematic errors (such as drift), contribute to the variability of pavement distress data. This variability can be expressed in terms of repeatability (i.e., “the capability of producing identical results by several rounds of measurement done by the same raters”) or reproducibility (i.e., “the capability of producing identical results by different raters”) (Livneh 1994), and can be quantified by standard deviation, variance and level of precision (Daleiden and Simpson 1998; Larson et al. 2000; Rada 1997), coefficient of variation (Larson et al. 2000; Rada et al. 1998), average deviation index (Migliaccio et al. 2011), kappa statistics (Ong et al. 2010), and Kendall’s correlation coefficient (Bogus et al. 2010a). Minimizing this variability can improve the accuracy and precision of the data, and is central to most quality assessments, controls, and management processes [e.g. (Bogus et al. 2010a; Larson et al. 2000; Migliaccio et al. 2011; Morian et al. 2002; Ong et al. 2009; Rada 1997)].

Constant systematic measurement errors affect the entire network in the same manner, and do not create variability in the data over time and place. Gauss (1777 to 1855) believed that it is up to the experimenter to eliminate the causes of such errors and to free the measured values from the errors' influence (Grabe 2010). However, even after the invention of many modern technologies designed to alleviate these types of issues, systematic measurement errors still exist. These errors can be detected by studying the difference between measurements of the same object by alternative means, then examining the form that the difference takes, testing its significance, and calibrating the measuring instruments (Humplick 1992). However, some systematic errors affect all measurements, regardless of the technology and the method used. Examples include a

bias towards detecting higher-severity distresses (while missing lower-severity distresses) (McQueen and Timm 2005), and the ambiguity often found among distress definitions (Morian et al. 2002). Since this kind of error does not contribute to variability in time, place or even ratings obtained by different methods, it is very difficult to detect (and thus often ignored).

Maximizing precision (and therefore minimizing variability) over time and among raters is the central focus of most of data quality studies (Bogus et al. 2010b; Larson et al. 2000; Migliaccio et al. 2011; Morian et al. 2002; Ong et al. 2009; Rada et al. 1998). Agencies also review the quality of the collected pavement condition data for consistency, completeness, and reasonableness. In 2008, 71% of the DOTs used software routines to review if the collected data were within the expected ranges, and 50% of the DOTs used statistical/software routines to compare the collected data with existing time-series data (Flintsch and McGhee 2009). New Mexico DOT, for example, compared their results with the previous years' data to identify locations where large changes occurred. If there were areas where large changes occurred, the data were then checked for reasonableness and consistency (NmDOT 2007).

Completeness of Pavement Data

Based on a study of 27 transportation agencies in the United States, Vandervalk-Ostrander et al. (2003) concluded that “most agencies are dealing with disparate data sources in mainframe flat files, redundant data, stovepipe management systems, and functional area barriers.” A recent peer exchange of six state and local transportation agencies reported that transportation agencies struggle with numerous challenges

regarding the integration of their disparate data (Hall, 2006). Adams (2008) has suggested that data integration problems are both technical and organizational, since many of the existing information systems and databases were developed separately, with one application in mind, and using a wide variety of hardware and software platforms. Thus, it is extremely difficult to locate and aggregate information across these many systems (Adams, 2008).

M&R history is necessary for several pavement management functions such as developing pavement condition prediction models [see for example, (Lytton 1987; Wang and Li 2011)], conducting survival analysis, and developing M&R trigger values based on past repair frequency. Several techniques have been used for developing deterministic and probabilistic pavement condition prediction models. In most of these modeling techniques, pavement age and construction history are key variables. The problem is that even though the M&R data may be available, it is not easily accessible or ready for integration. M&R data is usually gathered by construction personnel and are stored in construction databases (separate from PMS database). For example, at TxDOT, pavement condition data is collected and stored by a state-wide data collection unit; whereas M&R data is collected and stored by maintenance and construction units at the district level. Additionally, these databases use incompatible location referencing systems. Integrating these data has always been a major challenge for TxDOT (Zhang et al. 2001). Similar challenges exist in many other highway agencies. For small and new pavement networks, it may be practical to determine the missing M&R information through manual methods such as interviewing construction personnel or manually

searching project documents. However, as the network expands and ages, manual solutions turn very time and labor consuming.

Researchers have estimated pavement age based on current condition and assumed performance curves (Dewan and Smith 2003; Lee et al. 1993; McNinch et al. 2008). While the utilized performance curve is different in each case, all of these methods estimate construction and M&R history (i.e., pavement age) on a section-by-section basis. Estimating the age of a pavement section solely based on its own condition has some limitation. First, pavement projects often span over multiple pavement sections. Thus, a pavement section that might be in good condition may receive an overlay (for instance) if its adjacent sections are in poor condition and need to be repaired. The second, and perhaps more important limitation, is that pavement condition data tend to be highly variable. The technique developed in this thesis helps to address these limitations by recognizing spatial and temporal patterns in pavement condition data.

Spatial data analysis has been used widely in various areas of civil engineering, such as construction engineering and management (Cheng et al. 2005; Jia and Wang 2010; Jie and Caldas 2008; Lee and Adams 2004), transportation engineering (Li et al. 2007; Vlahogianni et al. 2007; Wang and Kockelman 2009), water pipeline condition assessment (Adachi and Ellingwood 2009; De Oliveira et al. 2011), and Hydrology (Cho and Olivera 2009; Olivera 2001; Olivera et al. 2006; Olivera and Maidment 1999). In PMS, spatial data analysis has been used to cluster pavement sections based on condition uniformity for project segmentation (Yang et al. 2009) and contracting (Kim et al. 2010)

purposes, and for identifying uniform regions for performance modeling purposes (Mishalani and Koutsopoulos 2002). Many agencies use GIS for generating maps and reports and linking multiple data layers. This thesis combines spatial data analysis and Bayesian statistics to develop a computational technique for the imputation of missing or inaccessible PMS data (specifically construction and M&R history).

Effects of Data Quality on Complex Systems

Organizational databases consistently have been reported to have a significant amount of errors (Laudon 1986; Redman and Godfrey 1996; Rosenberg 1974; Waldo 2005). The effects of data errors on the outputs of complex systems have been investigated by a number of researchers [e.g., (Ballou et al. 1987; Bansal et al. 1993; Coleman and Gardiner 2009; Klein and Rossin 1999; Rizk et al. 2000)]. The method used, however, depends upon the type of problem.

In problems where measurement errors (ΔX_1 to ΔX_n) are statistically independent and insignificant as compared to the overall measurements (X_1 to X_n), an error associated with the estimation of $Y=f(X_1, X_2, \dots, X_n)$ can be expressed by the use of the local partial derivative of an output variable (Y) in response to an input variable of the researcher's choice (X_i) (Saltelli et al. 2004). Estimating the model variance and the contribution of variables' variances to that model variance is another applicable approach (Saltelli et al. 2008). A more computationally efficient method is to fit a simple function (e.g., a line) to the model and evaluate the effects of the data quality on the simplified function using local derivatives (Chen 2005). In the case of complex engineering systems, however, large numbers of functional elements are involved, and

each functional element can exhibit its own complex behavior. The relationships among the outputs and inputs in such systems cannot be expressed in terms of a closed form function or reasonably be approximated by a simple function. Therefore, the use of any of the aforementioned methods becomes very difficult.

Monte Carlo simulation has been used extensively for simulating complex systems (Hutcheson 2009). Monte Carlo simulation includes defining a domain of possible inputs, generating inputs randomly from a probability distribution over the domain, and performing a deterministic computation on the inputs to compute an output (Hu et al. 2011). Decomposing the complex systems into functional subsystems allows for the performing of local analyses within each subsystem, as well as for a global analysis of the entire system (Hutcheson 2009).

A needs estimating process for a network-level PMS functioning as a complex system can be decomposed into the three subsystems: (1) condition assessment, (2) condition prediction, and (3) needs analysis (i.e., identifying appropriate M&R strategies, evaluating different alternatives, and prioritizing or optimizing the allocation of resources) (Flintsch 2004a). The data collected in the conditions assessment subsystem are then inputted into the other two systems. The impact of the errors associated with the collected data on the system's output depends upon the methods used in the associated subsystems. Computer simulation has been used successfully to evaluate the effects of various factors on PMS's, such as uncertainty in pavement condition prediction models (Madanat 1993), different policies (De la Garza et al. 1998), variations in agents' behaviors, different analytical techniques, variability, random

failures (Manzella and McNeil 2006), and advances in conditions sampling (Mishalani and Gong 2008). Different subsystems and the impact of errors are discussed in the following sections.

Impact of Pavement Condition Data Quality on PMS Output

Pavement condition data is a major component of network-level PMS (Mishalani and Gong 2008). The quality of this data is important not only in assessing the current condition of the network but also in the prediction of future condition and the planning of future M&R activities (AASHTO 1993; Flintsch and McGhee 2009). While there is a general agreement in the literature that the quality of condition data affects PMS outputs (Buchheit et al. 2005; Livneh 1994; McNeil and Humplnick 1991; Ong and Sinha 2008; Shekharan et al. 2007), little work has been done to quantify these effects.

It is also believed that systematic errors have a more significant impact on PMS output than random errors because they can affect a larger number of sections (Flintsch and McGhee 2009; Shekharan et al. 2007). There are also some case studies reported in the literature that show the impact of data quality on PMS's. For example, Shekharan et al. (2007) reported that a quality assurance/quality control plan for pavement condition data collection used by the Virginia DOT has resulted in an adjustment of the predicted maintenance treatments by 21%. However, little work has been done to quantify the benefits that high-quality pavement condition data provides in economic terms. This thesis describes an attempt to fill in the benefits determination gaps and to present a process that can be utilized to quantify the benefits of improving the quality of pavement condition data.

Effect of Data Quality on Pavement Condition Prediction

Modeling can be defined as “applying mathematical and statistical techniques to obtain answers to real problems” (Edwards and Hamson 1990). Several different kinds of prediction modeling techniques have been applied to pavement condition data. These techniques range from simple deterministic methods to complex probabilistic techniques (Lytton 1987). Most of these models are based on the assumption that the observational data are error free (Wang and Zaniewski 1995). However, pavement condition datasets almost always contain errors of varying types and magnitudes (Buchheit et al. 2005). Generally, correlations between measured values (i.e., pavement CI and age data) are weakened by measurement errors. However, the effects of these measurement errors depend upon the predictive method used.

Regression using the least-squares method is the most frequently used technique for modeling pavement condition data. In this technique, pavement CI is modeled as a function of pavement age and, the parameters of the model are estimated using the least-squares technique. Predictions from regression models are subjected to errors propagated from the estimated parameters in the model (Sulieman et al. 2001). In the case of linear regression, random errors in CI data do not cause bias in the model within large datasets. However, random errors in age data bias the prediction by causing the line slope to be underestimated. That is, the magnitude of the predicted CI is usually smaller than the true CI (Fuller 2006). In the case of non-linear regression, closed form solutions are available for calculating the sensitivity of the objective function and the regression coefficients of the most common parameter estimation methods (e.g., least squares and

min-max) (Castillo 2009). However, in general, no analytical method is currently available for computing the magnitude or direction of errors in predicted values without using extraneous sources of information (i.e., repeated observations). Alternatively, through the use of simulations it has been shown that if both the dependent and independent variables are exposed to measurement errors, the results of the regression can be poor (Valsami 2000).

Markov models have been widely used for modeling pavement conditions [e.g., (Golroo and Tighe 2009; Li 1996; Mao and Yu 2009; Takeyama et al. 1990)]. Markovian predictions are based on the assumption that the individual movements in condition are independent and governed by transition probabilities that depend only on current condition (Bartholomew 1975). Markovian predictions are exposed to three types of error. First, random errors result from the inherent randomness that results from the stochastic assumptions of the model. Second, estimation errors occur in instances where probability values are rarely known, and thus are estimated using past data. Third, specification errors occur when the assumptions of the model may not hold true in practice (Hoem 1973). In the absence of exact closed form solutions, the effects of change in the transition probability matrix on the steady state performance of a Markov process have been studied using computer simulation [e.g., (Cao and Xi-Ren 1998; Dai 1997; Fu 1992)] and by approximating these effects through a Poisson process (Cao 2000).

Artificial Neural Network (ANN) has been widely used for pavement condition prediction [e.g., (Lou 2001; OwusuAbabio and Owusu 1998; Yang 2003)]. Typically, a

neural network learns about a problem by being trained with examples. Training algorithms search for a set of weights that offer the best fit to the given examples. Once trained, a network can be used to make predictions (Adeli 2001). Data errors may affect the predictive accuracy of neural network models in two ways. First, the data used to train the network during the model construction may contain errors. Second, even if training data are free from errors, a user may input data containing errors into the model after a neural network is put into production (Klein and Rossin 1999). Shekharan (1998) examined an ANN by inputting erroneous data into it. The prediction accuracy of the ANN was found to degrade, but it still remained acceptable and as robust as a regression.

Effect of Data Quality on Needs Analysis

The needs analysis report is perhaps the most commonly generated report in pavement management. Typically, it includes detailed information about pavement needs in terms of overall M&R needed, corresponding funds required, and information regarding which pavement sections requires treatment, what treatments are needed, and when to apply these treatments (AASHTO 2001). The effects of data quality on PMS needs analysis outputs depend upon the method used in the needs analysis.

Benefit Cost Analysis (BCA) is the most common method of economic evaluation. BCA generally refers to the tools and procedures required to establish whether a proposed public investment will be worthwhile for meeting the public's needs (Cheneviere 2006). BCA, as a process, can be summarized as determining the scope and objectives of the project, identifying any constraints, alternatives, costs, and benefits, quantifying the associated costs and benefits, and finally calculating the net present

value of any alternatives (Kingston 2001). The reliability of a condition prediction model does not highly affect the needs analysis outputs as long as the model is used to predicting a reasonable time horizon (i.e., 10 years, and certainly no more than 20 years). Spencer (1994) investigated the effects of data quality on the outputs of CBA and concluded that if the benefits and costs were both large, but their differences were small, the results would be highly sensitive to data quality. Otherwise, the effects of data quality depend upon the sharpness of the benefit versus the costs graph near the optimum point. Hadley (2011) suggested analyzing the sensitivity of the CBA to its input data via a computer simulation, and developed a framework for such an analysis.

The objective of the BCA is to maximize the total of the net benefits, whether or not funding or performance are constrained. In the presence of a funding constraint, however, a lesser improvement with a higher BCR may be selected in order to implement an improvement with a higher BCR on another section. This procedure applies a methodology referred to as Incremental Benefit-Cost Analysis (IBCA), which compares a given improvement to a less aggressive improvement (FHWA 2005). The IBCA method determines the best M&R alternative for each pavement section by ensuring that the maximum benefit is obtained within the available budget (Boardman et al. 2005). IBCA is a well-established technique (Shahin et al. 1985) and is used in existing PMS's such as TxDOT's PMIS (Stampley et al. 1993), as well as HERS (FHWA 2005) which is used for prioritizing pavement M&R projects.

ANN has been widely used [e.g., (Al Qudrah and Alsugair 1998; Flintsch 1996; Hajek and Hurdal 1993; Taha and Hanna 1995)] in PMS needs analyses performed after

training, by using knowledge obtained from the histories of project selections. ANN's are attractive in PMS needs analysis because they can be used in a way similar to the decision-making process of a human being, and ANN's do not require predefined mathematical equations. Chan and Fwa (1993) investigated the performance of ANN's using computer simulations and showed that ANN's can perform well despite the presence of errors in the pavement condition data.

Fuzzy logic has also been used in PMS needs analysis [e.g., (Grivas and Shen 1995; Liang 2011; Prechaverakul 1995)] as an extension of the traditional rule-based reasoning (expert systems) capable of incorporating qualitative data into the decision-making process (Flintsch 2004b). To the best of knowledge of the authors, however, the effects of condition data quality on fuzzy systems have not yet been investigated.

Existing Outlier Detection Methods

The most common outlier detection processes can be classified as parametric and non-parametric. Parametric methods assume a distribution, estimate the distribution's parameters, and finally identify any outliers based on deviations in the data from that distribution. These methods are useful for low-dimensional data and for cases where a sufficient amount of prior knowledge of the distribution of the data has been obtained (Hodge and Austin 2004).

In pavement management, classic outlier detection methods are commonly used to filter data for modeling purposes. In these methods, conditions data are grouped into broad clusters (e.g., structure, usage level, climatic zone, etc.). Then, the data within each cluster is plotted against age. Outlier values within each age group are detected

based on various statistical criteria such as points deviating more significantly than three standard deviations from the mean (Nunez and Shahin 1986), studentized residuals greater than 2 (Lee and Darter 1994), points deviating larger than the 1.5 inter-quartile range from the first and third quartile values (Haider et al. 2007), residuals of a constrained least-square estimation model (Shahin et al. 1987), and general engineering judgment. These data filtering methods cannot differentiate between gross outliers and pseudo outliers, and thus are not suitable for data quality assessment purposes. Therefore, a new outlier detection method for pavement conditions data is needed.

Clustering is a non-parametric outlier detection method. It is a data mining method that tries to arrange the data into natural groupings (Ming-Syan et al. 1996). The data that remain out of the cluster are detected as outliers. The main objective of this method is the clustering, and the method is not optimized for outlier detection (Knorr and Ng 1998).

Distance-based methods identify the outliers by defining the ‘closeness’ of vectors using some suitable distance metric (Hodge and Austin 2004). For example, Knorr and Ng (1998) suggested the following criteria for detecting outliers: “An object 0 in a dataset T is a $DB(p, D)$ -outlier if at least fraction p of the objects in T lies greater than distance D from 0 .” These researchers also showed that the distance-based outlier detection method generalizes other statistical outlier detection methods. These methods come from computer science and work well in large datasets, with high dimensional data, and especially when the observed information does not fit any standard distribution (Papadimitriou et al. 2003).

Pattern-based methods use existing patterns in the data set to identify records that do not conform to the pattern. Pattern-based detection methods can also be used successfully as clustering and as a classic method (i.e., using measures like standard deviation or quartile ranges) for detecting outliers in pavement condition datasets (Buchheit et al. 2005).

Proximity-based methods are another category of outlier detection methods. These methods detect spatial outliers by searching for extreme observations or local instabilities with respect to adjacent values. These observations, however, may not appear extreme when compared to the entire population (Ben-Gal 2005). Proximity-based statistical methods form the basis for the outlier detection method developed in this research.

3. TEST-BENCH PAVEMENT MANAGEMENT SYSTEM

This section describes the PMS used in this research as a test bench for evaluating the impact of condition data quality on PMS outputs and for evaluating and testing the techniques developed for improving this data. The models and components of this test-bench PMS were developed for and are used by TxDOT; however, they are fairly representative of most pavement management systems currently in practice.

Condition Assessment

Pavement surface distress data is collected annually on the TxDOT network and used for programming and scheduling M&R actions. To evaluate the accuracy of the collected data, another independent set of data (called audit data) is collected on approximately 5% of the network. The collected condition data of a county is rejected if more than 15% of the data within that county shows a difference of 10 points or higher in the Distress Score (a composite pavement condition index, discussed next in this Section). For data collection purposes, TxDOT divides each highway into a number of sections; the sections have an average length of half a mile.

For asphalt pavement, distress quantity is measured in terms of percent of wheel path length (e.g., shallow and deep rutting, alligator cracking), percent of wheel path area (e.g., patching, block cracking), number of distress occurrences (e.g., failures), and length per 100-ft station (e.g., longitudinal and transverse cracking). Measured distress quantities were converted to distress densities (L). L values can range anywhere from 0 to 100 percent. L values are converted to utility values using Eq. (1), and then

aggregated into a 0-100 condition index called a DS using Eq. (2). This condition assessment process is described in further detail in the literature (Gharaibeh et al. 2010).

$$U_i = \begin{cases} \max(0, 1 - \alpha e^{-\left(\frac{\rho}{L_i}\right)^\beta}) & L_i > 0 \\ 1 & L_i = 0 \end{cases} \quad (1)$$

where

U_i : Utility value of distress type i

L_i : Density of distress type i

and α , β , and ρ : Utility factor coefficients defined based on pavement and distress types.

These factors are always positive.

$$DS = 100 \prod_{i=1}^n U_i \quad (2)$$

where

n : Number of distress types considered

U_i : Utility value for distress type i

DS is a 0-100 distress score of the pavement section (with 100 representing ideal condition). The DS of each section plays an important role in both the condition prediction and needs analysis phases (discussed in the following section).

Since α , β , and ρ are positive, U_i values will always be in the range of [0,1], according to Eq. (1). As a result, the multiplication of any number of U_i values will also be in the range of [0,1]. Consequently, the DS is guaranteed to be within the range of [0,100], according to Eq. (2).

M&R Assignment

M&R actions that are applied to pavement sections are classified into four categories: PM, LR, MR and HR. Table 1 shows examples of the treatment types associated with each sub-type of ACP.

Table 1 Examples of treatment types for ACP

M&R Type	Thick ACP (Type 4)	Intermediate ACP (Type 5)	Thin AC (Type 6)	Composite (Type 7)	Concrete overlaid (Type 8)	Flexible overlaid (Type 9)	Thin-surfaced flexible base (Type 10)
PM	Crack seal, Surface seal	Crack seal, Surface seal	Crack seal, Surface seal	Crack seal, Surface seal	Crack seal, Surface seal	Crack seal, Surface seal	Surface seal, no patching
LR	Thin asphalt overlay	Thin asphalt overlay	Thin asphalt overlay	Thin asphalt overlay	Thin asphalt overlay	Thin asphalt overlay	Surface seal, Light/medium patching
MR	Thick asphalt overlay	Thick asphalt overlay	Mill and asphalt overlay	Mill and asphalt overlay	Mill and asphalt overlay	Thick asphalt overlay	Surface seal, Heavy patching
HR	Remove asphalt surface, Replace and rework base	Remove asphalt surface, Replace and rework base	Reconstruct	Remove asphalt surface, Replace and rework base	Remove asphalt surface, Replace and rework base	Remove asphalt surface, Replace and rework base	Rework base & surface seal

Condition Prediction Models

For modeling pavement performance and detecting outliers, TxDOT clusters pavement sections together based on climate, subgrade quality, pavement type, last M&R type, and traffic loading level. Clustering the data according to these characteristics creates a tree-like division, as shown in Figure 2.

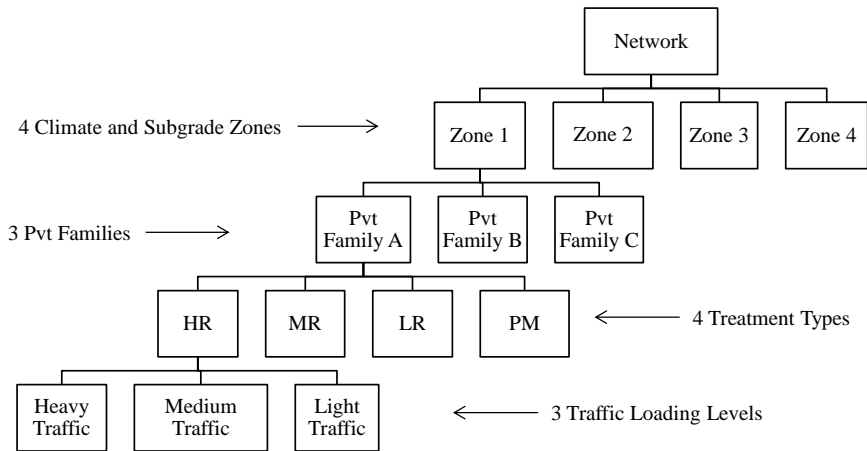


Figure 2 Data clustering used in a classic outlier detection method

Climate and Subgrade Zones

Since temperature, moisture, and subgrade quality all affect pavement performance, TxDOT classifies the state of Texas into four climate zones. Each zone represents the mixed effects of subgrade and climate.

- Zone 1: This zone represents a wet-cold climate, and poor, very poor, or mixed subgrade.
- Zone 2: This zone represents a wet-warm climate, and poor, very poor, or mixed subgrade.
- Zone 3: This zone represents a dry-cold climate, and good, very good, or mixed subgrade.
- Zone 4: This zone represents a dry-warm climate, and good, very good, or mixed subgrade.

Each county in Texas is assigned one of the above zones. These zones are depicted in the color-coded map shown in Figure 3.

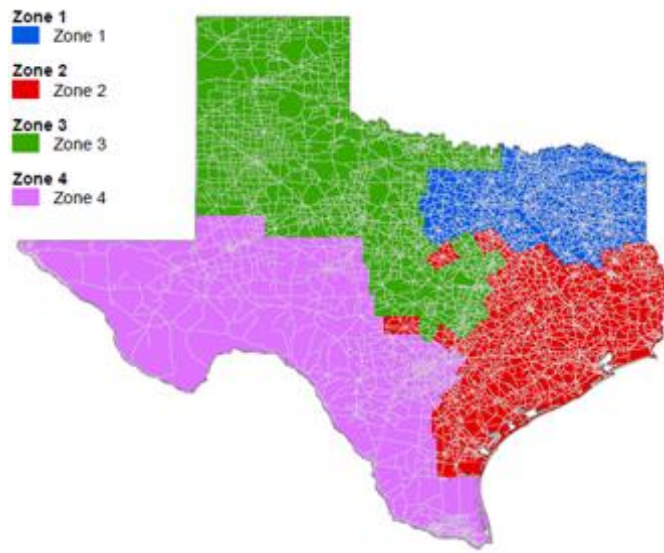


Figure 3 Climate and subgrade zones for ACP performance prediction model calibration

Pavement Families

PMIS divides ACP into seven types. These seven ACP types are grouped into three broader families, as follows.

- Pavement Family A: This pavement family includes thick ACP (PMIS Pavement Type 4), Intermediate ACP (PMIS Pavement Type 5), and overlaid ACP (PMIS Pavement Type 9).

- Pavement Family B: This pavement family includes composite pavement (PMIS Pavement Type 7) and concrete pavement overlaid with ACP (PMIS Pavement Type 8).
- Pavement Family C: This pavement family includes thin ACP (PMIS Pavement Type 6) and thin-surfaced ACP (PMIS Pavement Type 10).

Treatment Types

M&R actions that are applied to pavement sections are classified into four categories: PM, LR, MR and HR. Table 1 shows examples of the treatment types associated with each sub-type of ACP.

Traffic Loading Levels

This division includes three loading levels, as follows:

- Low Traffic Loading: This level includes pavement sections that have a 20-year projected cumulative Equivalent Single Axle Load (ESAL) of less than 1.0 million ESALs.
- Medium Traffic Loading: This level includes pavement sections that have a 20-year projected cumulative ESAL greater than or equal to 1.0 million ESALs and less than 10 million ESALs.
- Heavy Traffic Loading: This level includes pavement sections that have a 20-year projected cumulative ESAL greater than or equal to 10 million ESALs.

Within each cluster, TxDOT's condition prediction models are used to predict the quantity (L) of key individual distresses in the pavement section as a function of its type,

latest treatment category, age, subgrade type, traffic loading, and climatic region. As shown in Figure 4, TxDOT models are s-shaped, with the following form [Eq. (3)]:

$$L = \alpha e^{-\left(\frac{A}{age}\right)^\beta} \quad (3)$$

where age: number of years since last construction of the pavement section; α : (horizontal asymptote factor): controls the maximum amount of utility that can be lost; β : (slope factor): controls how steeply utility is lost in the middle of the curve; and A: (prolongation factor): controls how long the utility curve will "last" above a certain value. This equation reflects the effects of subgrade strength, climate, rainfall, freeze-thaw cycles, and traffic level; α , β , and A are provided by TxDOT for all combinations of pavement type, distress type, subgrade type, climate and traffic level in (Gharaibeh et al. 2012).

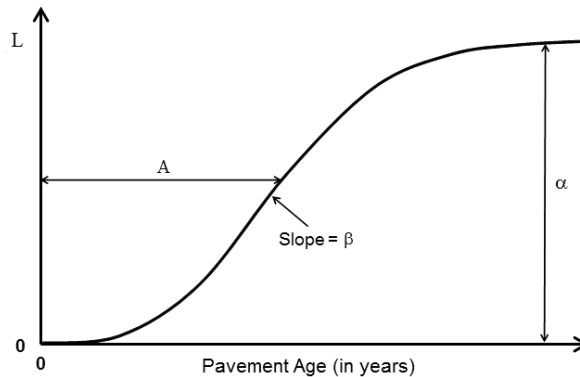


Figure 4 S-shaped TxDOT pavement condition prediction model

Examples of conditions prediction models of different treatment types applied to flexible pavement sections with wet and warm climates, poor subgrade conditions, and medium traffic loading (between 1 and 10 million cumulative ESALs projected over 20 years) are shown in Figure 5.

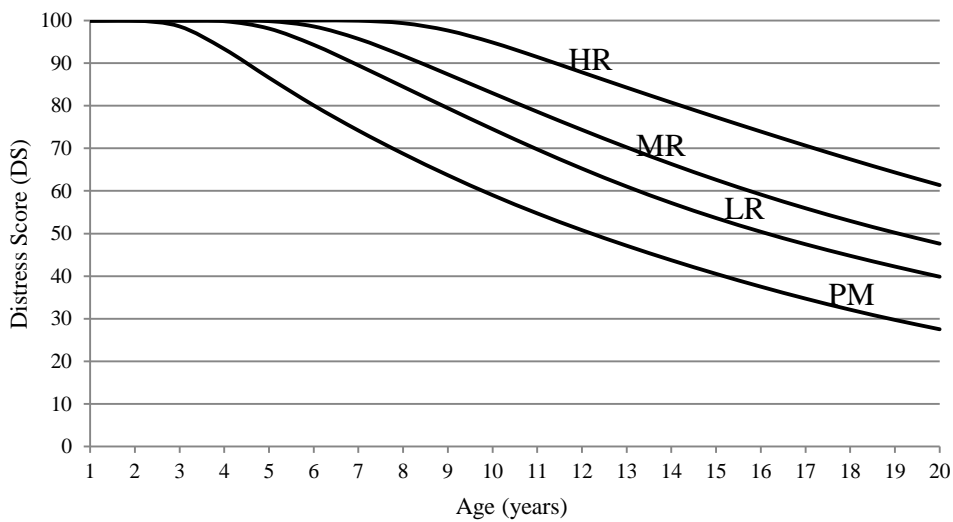


Figure 5 Condition prediction models for flexile pavement sections with wet and warm climates, poor subgrade condition and medium traffic

Data Storage and Management

Pavement conditions data are stored directly in the PMIS database while maintenance and repair project data are stored in separate databases. In the PMIS database, pavement sections are identified (or referenced) by a unique address which is a combination of district name, county name, highway name, and beginning and end reference mile markers. On the other hand, M&R and construction databases identify projects by a code

(called control-section-job or CSJ) and an approximate location on the roadway. In many cases, the repair year and specific work type are missing. Additionally, some routine maintenance projects are performed in-house, and thus are not recorded in the database at all. Hence, it is very difficult to integrate construction and repair history with condition data from these databases.

Pavement Condition Data Quality

To control the quality of collected data, another independent set of data, called audit data, is collected on approximately 5% of the network. The collected DS dataset of a county is rejected if more than 15% of the dataset within that county has a difference of 10 points or higher (Button et al. 2001). Audit data are collected using the same definitions and technology and are subjected to the same systematic and random errors.

The quality of the data also can be judged indirectly. As explained earlier, due to the economy-of-scale as well as practical reasons, M&R actions are normally applied to a group of sections and not to isolated individual sections. Figure 6 shows the sections that, based on the collected condition data, were improved in condition from 2009 to 2010 in the Bryan district. These sections are expected to represent M&R projects. However, many are single, isolated sections. One explanation is the dataset is erroneous and the error represents more than the year to year deterioration of the pavement.



Figure 6 Pavement sections with an increased DS from 2009 to 2010 (Bryan District)

TxDOT PMIS as explained in this section is later used to validate the research presented in next sections.

4. EVALUATING THE IMPACT OF CONDITION DATA QUALITY ON NETWORK-LEVEL PAVEMENT MANAGEMENT SYSTEMS

This section presents a quantitative assessment of the impacts of error magnitude and type (both systematic and random) in pavement condition data on the accuracy of PMS outputs (i.e., the forecasted needed budget and M&R activities throughout a multi-year planning period). Systematic and random errors were introduced into the condition dataset of the Bryan district of TxDOT and allowed to propagate throughout the components of the test-bench PMS.

The results of this investigation will help to optimize pavement condition data collection processes by focusing on error levels and types that cause the greatest impact on PMS outputs. Also, these results will help highway agencies to establish quality acceptance criteria for pavement condition data based on the impact of the data quality on the accuracy of PMS outputs. Currently, these criteria are set based on agencies past experiences and the capabilities of data collection technologies (Landers et al. 2001; Larson et al. 2000).

Error Propagation Simulation Process

Figure 7 illustrates the overall process that was developed to simulate the propagation of pavement condition data errors to the output of a PMS needs analysis. This process has five key components: (1) condition data generation, (2) error perturbation, (3) condition prediction, (4) M&R prioritization, and (5) output generation.

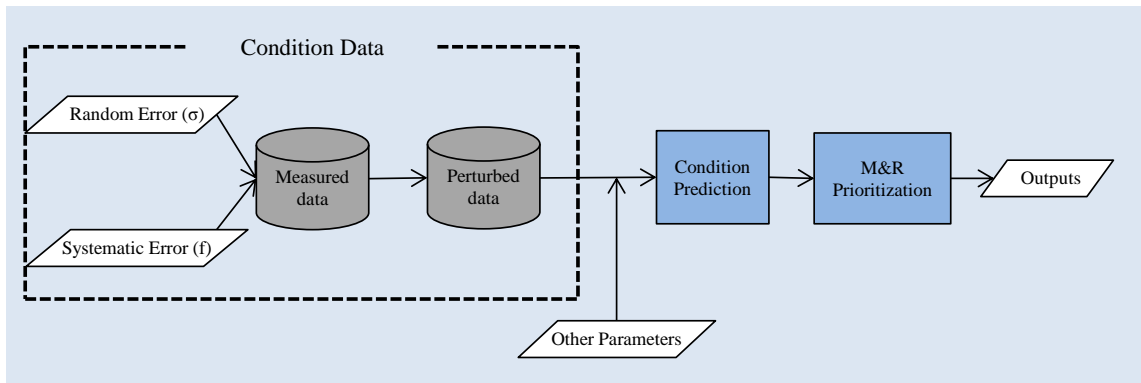


Figure 7 Process developed to simulate the propagation of pavement condition data error to the output of a PMS needs analysis

Computation burden is a major consideration when carrying out the above process. The computation time is influenced by the size of the condition dataset (i.e., the number of pavement sections in the PMS database), the number of parameters being varied, the range in which the parameters are being varied, the output of interest (e.g., point-estimate vs. interval estimate or confidence interval), and the computation computer's power.

Condition Dataset

The 2011 pavement condition dataset for TxDOT's Bryan district was used for the analysis presented here. For pavement condition data collection purposes, TxDOT divides each highway into a number of sections with an average length of half a mile (Stampley et al. 1993). Pavement distress and roughness data are collected every year for these sections. In 2011, the Bryan network consisted of 7,299 0.5-mile sections (approximately 3,650 centerline miles).

As discussed earlier (see Section 3), distress quantities are aggregated into a 0-100 condition index called a DS (or distress score). The DS plays an important role in both the condition prediction and needs analysis phases and thus is used as a basis for quantifying the impact of condition data error on PMS output.

Error Perturbation

First, the 2011 pavement condition dataset for the Bryan district was cleansed of gross outliers to form a so-called parent “clean” dataset. Both random and systematic errors were then introduced into this parent dataset. In this research, a systematic error is represented as an additive constant. A random error is modeled as a normally-distributed random variable with a mean of zero and a standard deviation of σ . Additive systematic errors (f) and random errors (e) were added to the DS_{parent} values in the parent dataset to form a perturbed DS dataset, as shown in Eq. (4). In cases the calculated DS becomes greater than 100 or less than 0 after adding the errors, it is set to 100 or 0 respectively.

$$DS = DS_{\text{parent}} + f + e \quad (4)$$

where e : Additive random errors distributed normally with a mean of zero and a standard deviation of σ [Mathematically expressed: $e \sim N(0, \sigma^2)$].

To keep the computation burden affordable, the analysis was conducted by varying the DS systematic errors (f) and the standard deviation of random errors (σ) within the ranges of $[-10, 10]$ and $[0, 10]$, respectively. The range for σ is equivalent to having a random error in DS fall within the range of $[-10, 10]$ approximately 70% of the time, and within the range of $[-20, 20]$ approximately 95% of the time. The simulation

was performed by changing the error parameters (f and σ) one at a time, while keeping all other PMS parameters fixed.

Condition Prediction

The condition prediction models were discussed earlier (see Section 3). These models were calibrated to pavement condition field data that represent the various pavement types used by TxDOT and the diverse climatic, subgrade, and traffic conditions in Texas.

M&R Project Prioritization

Needs analysis in PMS involves estimating the amount of funding needed (i.e., the budget) to achieve a target condition level for the pavement network over a multi-year planning period. TxDOT defines its target as having 90% of the pavement network in good (or better) condition (TxDOT 2005). This target is expressed here as 90% of the pavement sections having $DS \geq 80$.

BCA was performed to select the most cost-effective M&R category for each pavement section, and to prioritize the pavement sections for M&R throughout a 10-year planning period. Four M&R categories were considered in this analysis: PM, LR, MR, and HR. Each section received one of these four M&R categories or the DN option in each year of the planning period.

VMT-A was used as a measure of the benefit. VMT-A represents the AUPC, weighted by project length and traffic volume. This concept is depicted in Figure 8 and is computed using Eq. (5). The prediction curves shown in this graph represent flexible pavement sections with a wet and warm climate, poor subgrade condition, and under

medium traffic loading (between 1 and 10 million cumulative ESALs projected over 20 years). For example, a heavy rehabilitation treatment deteriorates at a slower rate than a medium rehabilitation treatment, and thus has a greater AUPC than a medium rehabilitation treatment (in Figure 8, $AUPC-H > AUPC-M$).

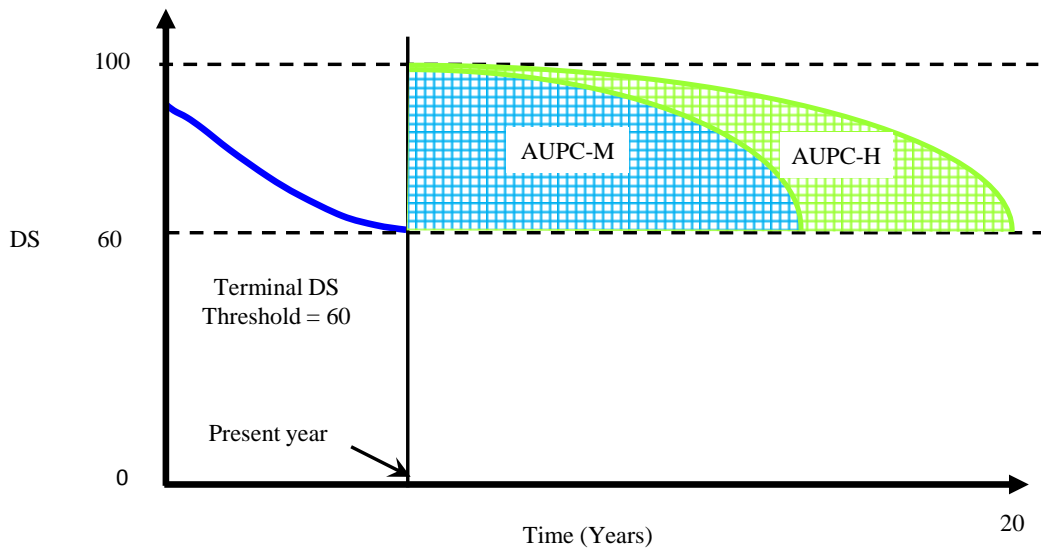


Figure 8 Concept of area under a performance curve

The VMT-A is computed for all viable M&R treatment categories for every section in the network, as follows:

$$VMT-A = AUPC \times Length \times AADT \times 365 \quad (5)$$

where AADT: Annual Average Daily Traffic as vehicles per day; Length: Pavement section length in miles; and AUPC: Area under the performance curve.

The cost of the M&R treatments is computed as follows:

$$\text{Cost} = \text{Length} \times C \times N \quad (6)$$

where Length: Length of the pavement section (in miles); C: Unit cost of the M&R treatment (\$ per lane-mile); and N: Number of lanes of the pavement section.

Results and Discussion

The simulation process discussed earlier was applied to the roadway pavement network in the Bryan district of TxDOT. The results are presented in terms of errors in the computation of the average annual M&R budget [Eq. (7) and (8)], the portion of the network requiring that the repair officials “do nothing” [Eq. (9)], the portion of the network needing maintenance [Eq. (10)], and the portion of network needing rehabilitation [Eq. (11)].

$$e_B = \text{Budget}_{\text{Perturbed}} - \text{Budget}_{\text{Parent}} \quad (7)$$

where e_B : Error in the estimated average annual budget; $\text{Budget}_{\text{Perturbed}}$: Estimated needed average annual budget using perturbed condition data; and $\text{Budget}_{\text{Parent}}$: Estimated needed average annual budget using the parent condition dataset (i.e., the dataset without any introduced errors).

$$\text{Proportional } e_B = \frac{e_B}{\text{Budget}_{\text{Parent}}} \quad (8)$$

where Proportional e_B : Error in the estimated average annual budget expressed as a ratio of the needed average annual budget using the parent condition dataset.

$$e_{DN} = \text{DN}_{\text{Perturbed}} - \text{DN}_{\text{Parent}} \quad (9)$$

where e_{DN} : Error in the estimated portion of the network requiring that repair officials “do nothing”; $\text{DN}_{\text{Perturbed}}$: Estimated portion of the network requiring that the repair officials “do nothing” using the perturbed condition dataset; and $\text{DN}_{\text{Parent}}$: Estimated

portion of the network requiring that the repair officials "do nothing" using the parent condition dataset.

$$e_M = M_{Perturbed} - M_{Parent} \quad (10)$$

where e_M : Error in the estimated portion of the network needing preventive maintenance; $M_{Perturbed}$: Estimated portion of the network needing preventive maintenance using the perturbed condition dataset; and M_{Parent} : Estimated portion of the network needing preventive maintenance using the parent condition dataset.

$$e_R = R_{Perturbed} - R_{Parent} \quad (11)$$

where e_R : Error in the estimated portion of the network needing rehabilitation; $R_{Perturbed}$: Estimated portion of the network needing rehabilitation using the perturbed condition dataset; and R_{Parent} : Estimated portion of the network needing rehabilitation using the parent condition dataset.

Effect of Systematic Error

Systematic error was modeled as a constant added to the measured DS value of each section in the network. This error was introduced into the condition data in the base year (i.e., 2011) only, and was allowed to propagate to future years through the condition prediction models and the M&R project prioritization process.

Figure 9 shows the effect of the introduced systematic error on errors in the estimated average annual budget. Positive systematic errors falsely improve the condition of the network and make the pavement sections look better than they actually are. Thus, as a consequence, no repair is often assigned to certain sections already in poor conditions, and consequently the necessary budget is underestimated. Negative

systematic errors make the network falsely look worse than it actually is. As a result, M&R is often assigned to sections that do not actually need it, and consequently the needed budget is overestimated.

Figure 9 suggests that the needed budget is more sensitive to negative systematic errors than to positive systematic errors. This conclusion can be explained by the fact that most of the pavement sections in the network have a DS higher than the repair trigger value, and thus are unaffected by the positive values of systematic errors. However, negative systematic errors can bring these pavement sections under the repair trigger value and falsely make them candidates for M&R. Figure 9 illustrates that even a systematic error as low as -2 DS points can distort the estimated annual budget value by as much as 50%.

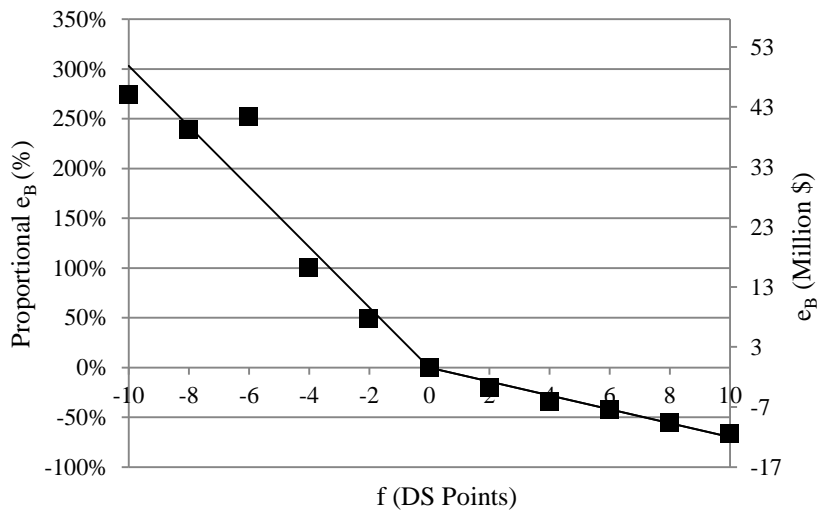


Figure 9 Effects of systematic errors in pavement condition data on errors in the estimated average annual budget

Figure 10 shows the effects of the introduced systematic errors in condition data on the error in the estimated portions of the network that require maintenance and rehabilitation, as well as on any portions that do not require any M&R. e_M and e_R follow a trend similar to e_B (as the percentage of repaired section increases, the needed budget also increases), while e_{DN} follows an opposite trend (as the percentage of repaired sections increases, the percentage of the sections that do not need M&R decreases). Also, Figure 10 suggests that the estimated M&R portions of the network are more sensitive to negative systematic errors than positive systematic errors in the condition data.

Figure 11 shows thematic maps of the M&R types assigned to each pavement section during the planning period, in the presence of additive systematic errors of -5 and 5, compared to the parent case of no introduced error (i.e., $f=0$). These maps show that in any given year of the planning period, a positive error leads to falsely assigning less extensive M&R (changes from rehabilitation to maintenance and from maintenance to do nothing). Conversely, a negative error leads to falsely assigning more extensive M&R (changes from maintenance to rehabilitation and from “do nothing” to maintenance). The densities of the M&R sections on these maps show the effects of positive and negative systematic errors in a visual manner.

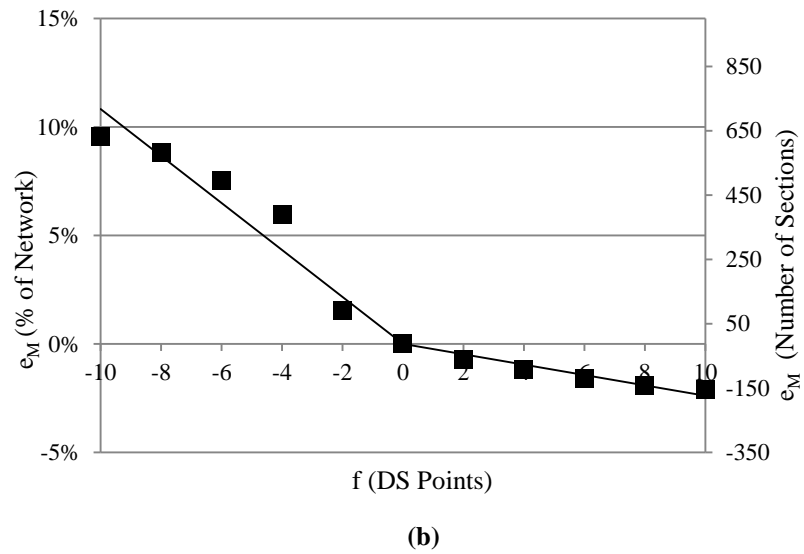
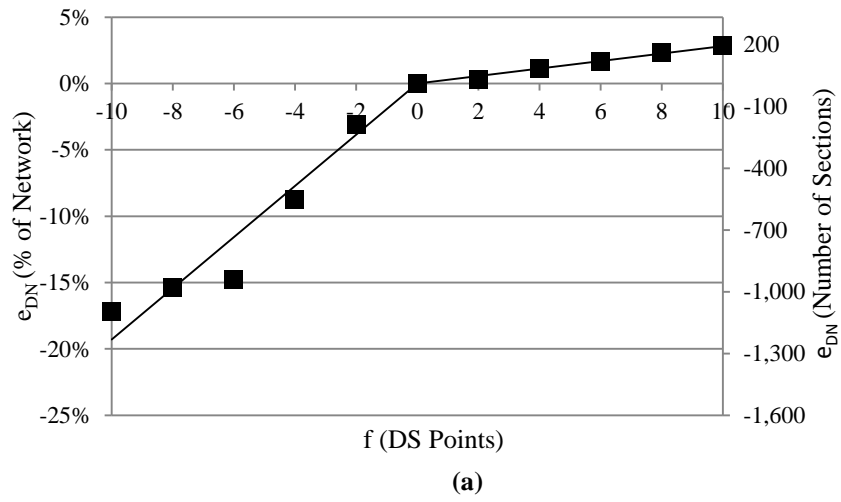
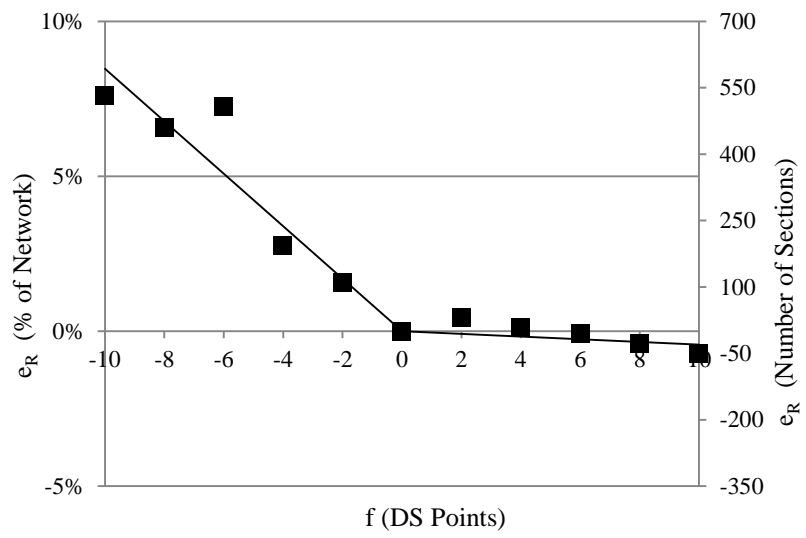


Figure 10 Effects of systematic errors in pavement condition data on errors in estimated (a) portions of the network that need no M&R, (b) portions of the network that need maintenance, and (c) portions of the network that need rehabilitation



(c)

Figure 10 continued

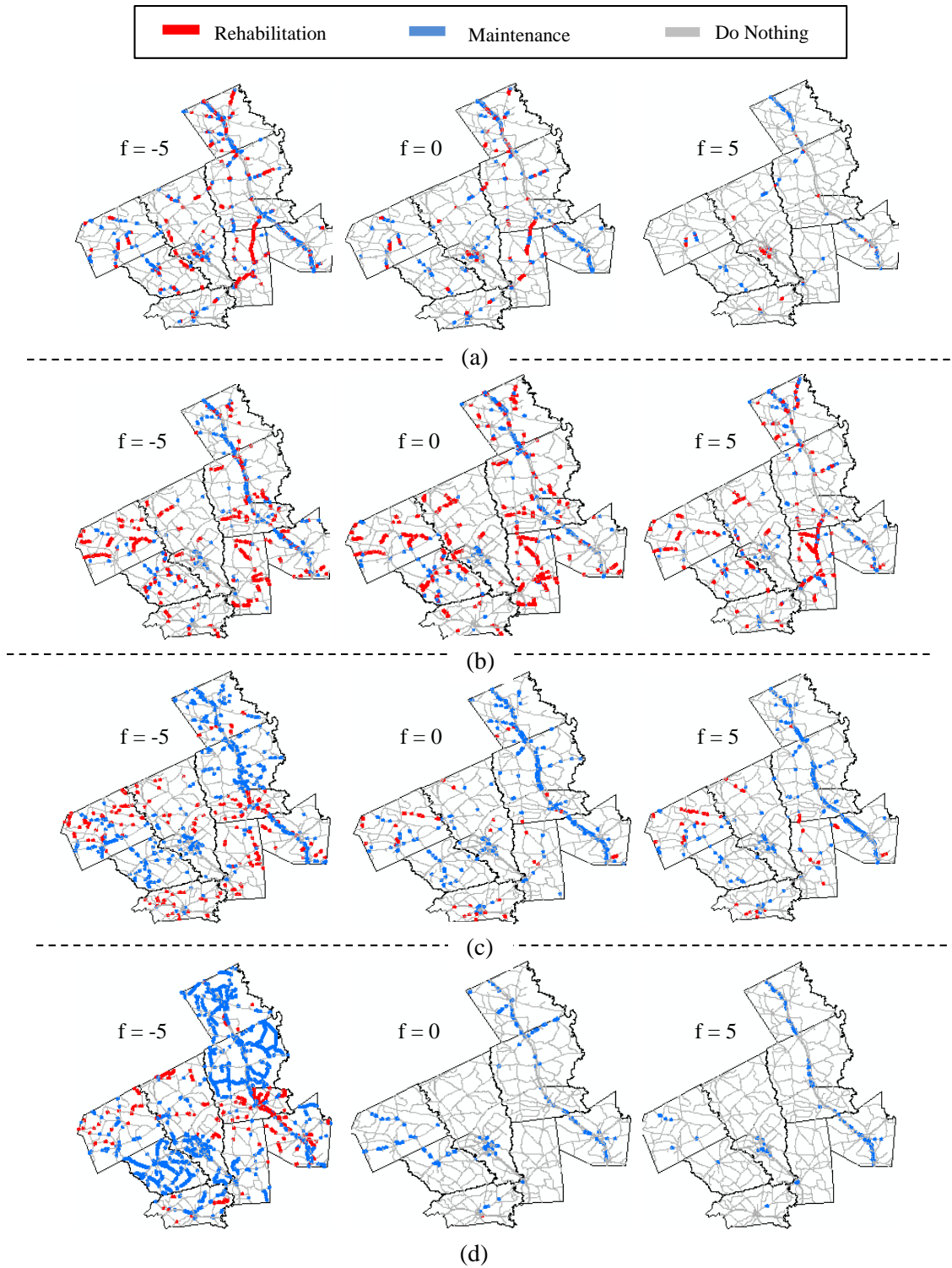


Figure 11 Pavement network M&R needs in the presence of different levels of systematic error in conditions data (a) 2011, (b) 2012, (c) 2013, and (d) 2014

Effects of Random Errors

Random error in the DS is modeled as a normally distributed random variable with a mean of zero and a standard deviation of σ . In each run of the needs analysis simulation, the randomly generated DS error values were added to the DS values in the parent database. These errors were then introduced to the condition data of the base year (i.e., 2011) only, and were allowed to propagate to future years through the condition prediction models and the M&R project prioritization process.

As can be seen from figure 13 and figure 14, the introduction of random errors into the DS led to an overestimation of the network's needs. Although the introduced random errors were simulated from a normal distribution with an average of zero, their effects on the PMS outputs were not neutral. In most networks, a large majority of sections are in a good condition and well above the repair trigger value. Positive error values only affected sections with a condition level below the repair trigger value (which made up only a small portion of the network) and did not affect pavement sections with a condition level above the repair trigger value (which were a large majority of the network). However, negative error values can affect sections with a condition level above the repair trigger value and bring their condition down below the repair trigger value. Therefore, although positive and negative random error values had an equal probability of occurrence, they were not balanced in terms of their impact on the estimated network needs.

Figure 12 and Figure 13 suggest that the sensitivity to random error of the estimated needed budget is higher than that of the estimated network portions needing

“do nothing,” maintenance, or rehabilitation. For example, a DS random error with $\sigma = 4$, which is equivalent to having 95% of the error values within the range of $[-8, 8]$, can lead to an error of 50% in the estimation of the average annual budget needed, and only 1% - 4% in the estimated M&R network portions. This difference can be attributed to the fact that these computed portions are blind to the identities of the pavement sections included in them. To demonstrate the effects of random errors on the estimated M&R types assigned to individual pavement sections, thematic maps of these M&R types are included in Figure 14.

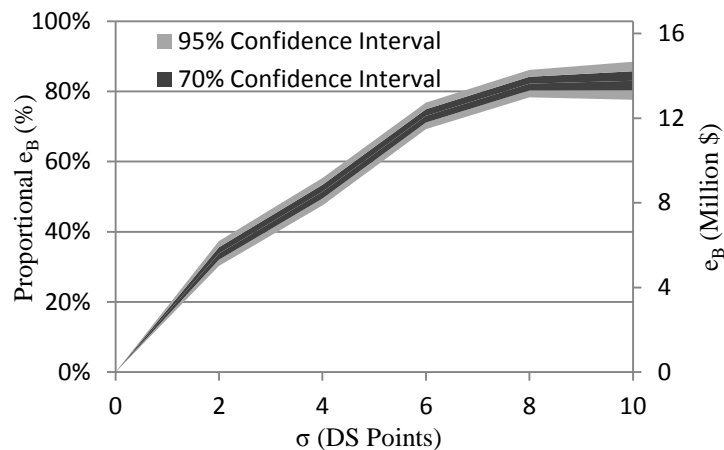
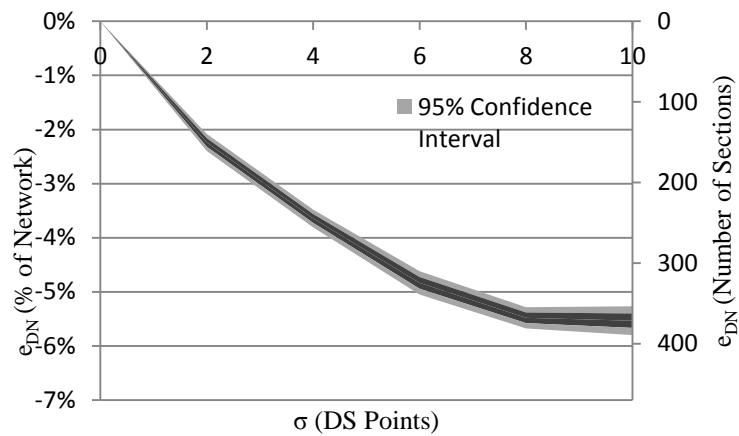


Figure 12 Effects of random errors in pavement condition data on errors in the estimated average annual budget

Figure 14 shows thematic maps of the M&R types assigned to each pavement section in the presence of introduced, normally-distributed random errors with a standard deviation of 5 and 10 DS points, in comparison to the parent condition dataset (i.e., no introduced

random errors). As explained before, with an increase in the standard deviation of error (σ), the portion of the network that does not need any repair action (i.e., the "do nothing" category) decreases, and more pavement sections are assigned rehabilitation and maintenance. It can be observed that for many sections, the assigned M&R type changes in the presence of random errors in the condition data.



(a)

Figure 13 Effects of random errors in pavement condition data on errors in estimated (a) portions of the network in need of "do nothing," (b) portions of the network in need of maintenance, and (c) portions of the network in need of rehabilitation

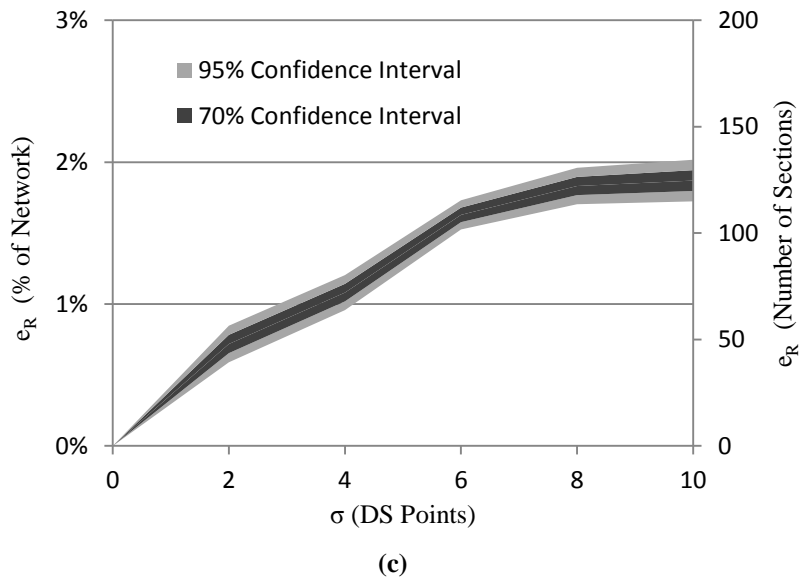
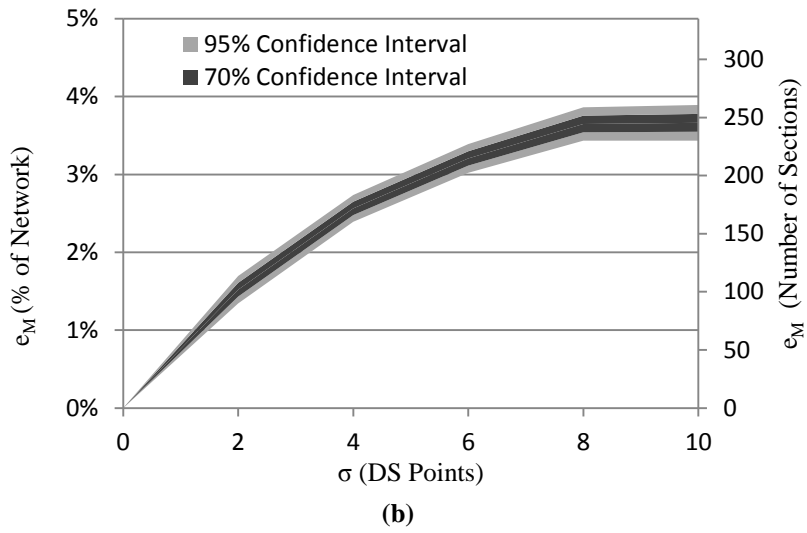


Figure 13 continued

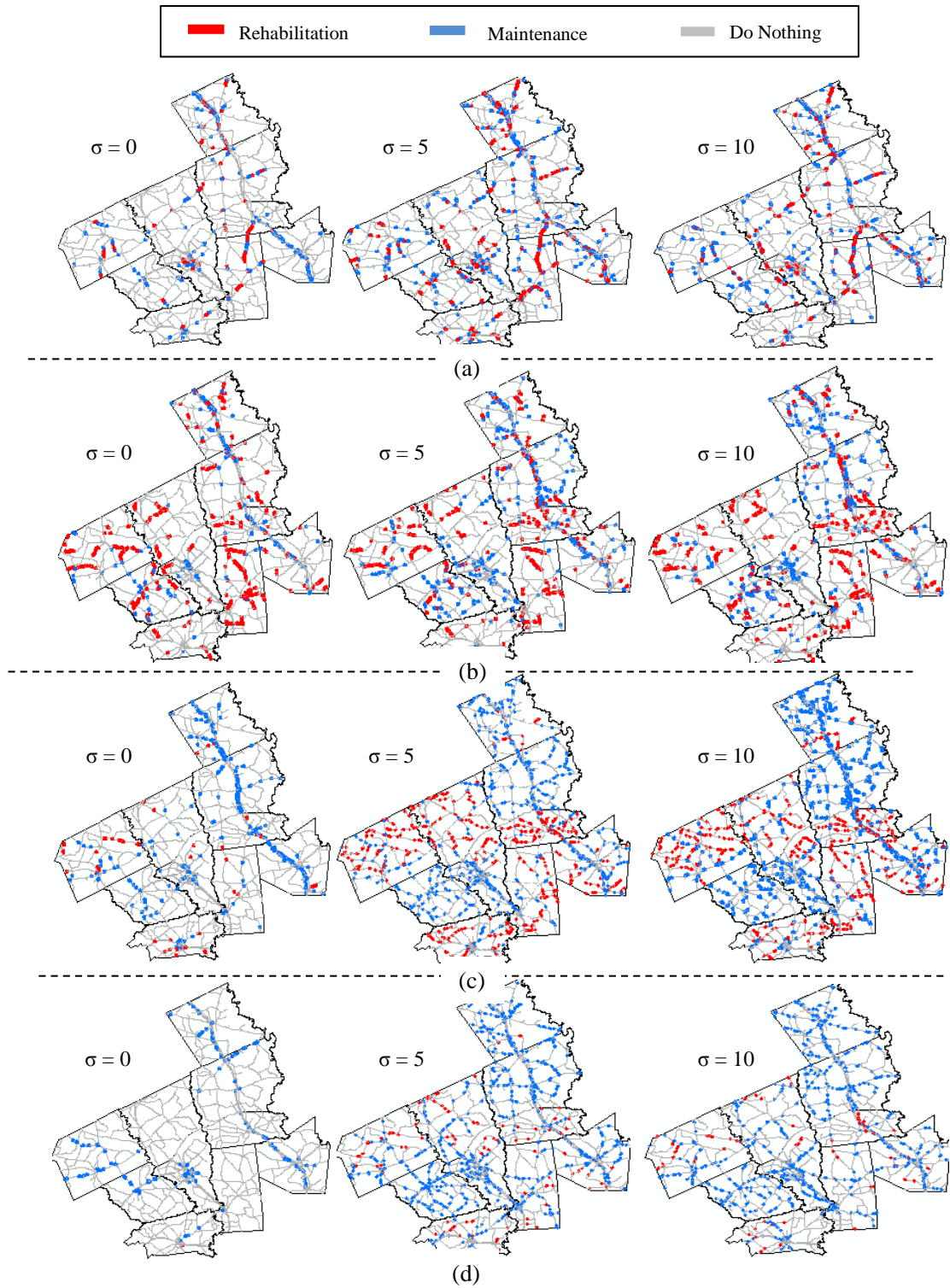


Figure 14 Pavement network M&R needs in the presence of different levels of random errors in condition data (a) 2011, (b) 2012, (c) 2013, and (d) 2014

The above results suggest that the sensitivity of a PMS needs analysis at the section level (i.e., M&R types assigned to individual sections) to random errors in the condition data is higher than that of the aggregated level (i.e., portions of the network assigned different M&R types). This finding is demonstrated through an example. Figure 15 shows the percentage of network sections that have their assigned M&R types changed as a result of a DS random error with $\sigma=5$. Aggregate PMS outputs (i.e., average annual budget and portions of the network in need of each M&R type) only reflect a 2.91% increase in the maintenance category and a 1.43% increase in the rehabilitation category, while 10.92% of the network sections had their assigned M&R types changed. In the case of systematic errors, however, the changes were only in one direction. For example, a negative systematic error made the pavement sections look worse than they actually were, and as a result, the sections received a more extensive M&R type. Therefore, the aggregate PMS outputs reflect the true extent of the impact of systematic errors on the network.

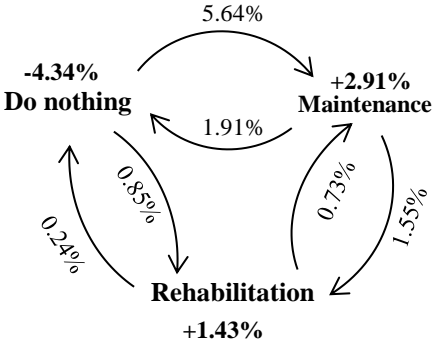


Figure 15 Percentage of network sections with assigned M&R types changed as a result of a DS random error with $\sigma=5$

Practical Implications of Study Results

A common way of assessing and controlling the quality of pavement condition data is to recollect data samples from the network (called audit data) and compare such data to the data collected by the vendor (called original data). For example, TxDOT accepts the data of each county only if more than 90% of the audited sections have an absolute deviation of less than 15.0 DS points from the original data (Griffith and Chamman 2000). Similarly, Virginia DOT's acceptance criteria require that 95% of the data have an absolute deviation of less than 10 points (Flintsch and McGhee 2009). TxDOT's criteria can be translated to $1.65 \times \sqrt{2} \times \sigma < 15$ or $\sigma < 6.42$ and Virginia DOT criteria can be translated to $2 \times \sqrt{2} \times \sigma < 10$ or $\sigma < 3.53$. Based on Figure 12, this level of condition data errors can, in the worst case scenario, result in an overestimation of the annual budget by as much as 80% for TxDOT, and as much as 50% for Virginia DOT, with a 95% level of confidence. These calculations suggest that these acceptable criteria still result in a highly distorted estimated needed budget. However, these levels of error in condition data are less distorting to the estimated M&R network portions (1% - 5% error, with a 95% level of confidence).

Section Summary and Concluding Remarks

This section provides a quantitative assessment of the effects of error magnitude and type (both systematic and random) in pavement condition data on the accuracy of PMS outputs (i.e., forecasted needed budget and M&R activities in a multi-year planning period). The process used to simulate the propagation of pavement condition data errors to the output of a PMS needs analysis consists of five key components: (1) condition

data generation, (2) error perturbation, (3) condition prediction, (4) M&R prioritization, and (5) output generation. This process was applied to the 2011 pavement condition dataset of TxDOT's Bryan district. In 2011, this roadway network consisted of approximately 3,650 centerline miles.

Key conclusions of this section are as follows:

- Forecasted average annual budget needed is highly sensitive to both systematic and random errors in pavement condition data. This output parameter of PMS needs analysis can be highly distorted by usually considered “acceptable” or inevitable random error in pavement condition data. For example, a ± 10 standard error in DS can result in overestimation of annual budget by as much as 85%, with 95% confidence.
- Predicted portions of network in need of rehabilitation, maintenance, and “do nothing” are less sensitive to error in pavement condition data compared to forecasted needed budget. For example, a ± 10 standard error in DS can result in overestimation of rehabilitation portion by as much as 2%, overestimation of maintenance portion by as much as 3.8%, and underestimation of do nothing portion by 5.8%, with 95% confidence.
- The sensitivity of PMS needs analysis at the section level (i.e., M&R type assigned to individual sections) to random error in pavement condition data is higher than that at aggregated level (i.e., portions of network assigned different M&R types).

- PMS outputs are more sensitive to negative errors in condition data (i.e., overestimation of distress and underestimation of condition index) than positive errors (underestimation of distress and overestimation of condition index).
- Systematic error in pavement condition data has a higher impact on PMS outputs than random error because it affects the entire network in the same manner (i.e., either increase or decrease the condition of the whole network).
- The effects of pavement condition data error on PMS needs analysis outputs persist throughout the planning period.

5. A PROXIMITY-BASED TECHNIQUE FOR DETECTING OUTLIERS IN PAVEMENT CONDITION DATA

This section presents the development of a new technique for detecting gross and pseudo outliers in pavement condition data. The developed technique integrates concepts and methods from GIS, Bayesian statistics and Genetic Algorithms to cluster homogenous neighboring pavement sections and then identify gross outliers (data values that are likely to be erroneous), as well as pseudo outliers (pavement sections affected by isolated local conditions that caused unusual pavement performance).

Hawkins (1980) defines outlier as: "an observation which deviates so much from other observations as to arouse suspicions that it was generated by a different mechanism." In the case of a pavement section, for example, the different mechanism could be a defective drainage condition causing a faster deterioration rate, or a data error caused by the rater's distraction. Therefore, outliers can be truly dissimilar values (pseudo outliers) or erroneous values (gross outliers).

The proximity of the sections and their chain-like form is devised in this work, and the outlier detection method developed is applicable to the condition data of all large linear infrastructures (e.g., roadway networks, pipelines, power transmission lines, etc.).

Developed Outlier Detection Method

Sections in close proximity are influenced by similar local factors, such as drainage condition, subgrade, and construction and repair history. Also, the condition data for sections in close proximity are normally collected by the same rater and equipment over the same period of time. Therefore, spatial trends can play an important role in detecting

outliers in pavement condition data. The method developed and presented in this thesis verifies whether the collected condition data of a given section is reasonable in relation to the previously collected data for that section, as well as for the data of its neighboring sections.

Several phenomena (e.g., air temperature and air pollution) that occur in nature have been analyzed successfully by modeling their gradual changes over space [see for example, (Matejcek 2011; Zhang et al. 2009)]. However, in the case of pavement condition data, repair actions cause sudden changes in spatial trends. In other words, the spatial trends of pavement condition data can be considered only within clusters of adjacent sections that have similar repair histories. This research demonstrates how the consideration of geographic proximity and historical patterns can lead to more informative outlier analyses of pavement condition data.

Multivariate outliers can be categorized into gross outliers and pseudo outliers (Zhang et al. 2007). In the case of spatio-temporal analyses, gross outliers are those that deviate significantly both from their neighbors' data and from their own history. These data are likely to be erroneous (Byrne et al. 2011). Data that are spatial outliers but not outliers with respect to historical trends are called pseudo outliers. Pseudo outliers are likely to be roadway sections influenced by different factors over their service lives. The outlier detection method developed for this research is comprised of two steps: first, the clustering of neighboring sections based on similarities in historical condition data and prior information on M&R project lengths; second, the detection of outlier data and the

classification of that data into those instances that are likely errors (i.e., gross outliers) and those instances that are likely dissimilar data (i.e., pseudo outliers).

Step 1: Clustering Sections Based on their Spatiotemporal Patterns in Condition Data

For this research, each roadway corridor was segmented into uniform clusters of contiguous sections so that the sections within each cluster shared the most similarities in terms of their deterioration patterns. Figure 16 illustrates this concept of clustering the roadway sections.

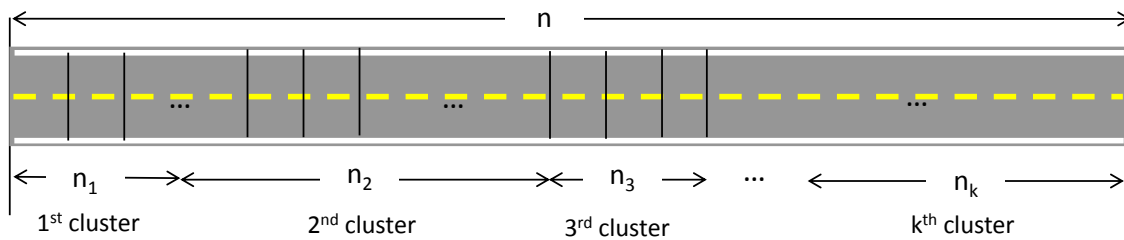


Figure 16 Clustering a roadway corridor with n sections into k clusters

The algorithm used to cluster the neighboring sections is illustrated by an example consisting of a 5.5-mile portion (i.e., 11 0.5-mile pavement sections) of Highway 50 in the Bryan district of TxDOT. For simplicity, only two years (2009 and 2010) of condition data are considered in this example. Two scenarios for clustering these 11 pavement sections are discussed below.

Scenario # 1: This clustering scenario is illustrated in Figure 17. The first five sections represent one cluster (i.e., Cluster 1) and the remaining six sections represent a

second cluster (i.e., Cluster 2). As shown in Figure 18, the 2009 and 2010 CI data for Cluster 1 are clustered closely, indicating that the sections within this cluster had similar condition index (CI) data in both 2009 and 2010 (i.e., the sections in this cluster are similar in terms of the CI patterns and values). However, the 2009 and 2010 CI data for Cluster 2 are scattered, indicating that the sections within this cluster were different in terms of the CI patterns and values in 2009 and 2010.

To measure the uniformity of each cluster, a multivariate Gaussian probability density function (PDF) was fitted to the points of each cluster. As shown in Figure 18, Cluster 1 has a tight multivariate Gaussian PDF and Cluster 2 has a widespread Gaussian PDF, indicating that Cluster 1 has a lower CI variability as compared to Cluster 2.

In both cases, the multivariate Gaussian PDF was computed using Eq. (12).

$$Q_i = f(X) = \frac{1}{(2\pi)^{\frac{k}{2}} |\Sigma|^{\frac{1}{2}}} e^{-\frac{1}{2}(X-\bar{X})' \Sigma^{-1} (X-\bar{X})} \quad (12)$$

where \bar{X} = the mean vector and Σ = the covariance matrix.

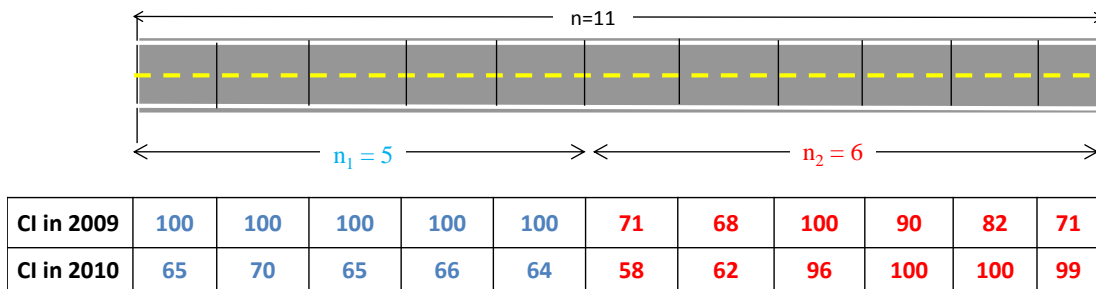


Figure 17 The first clustering scenario

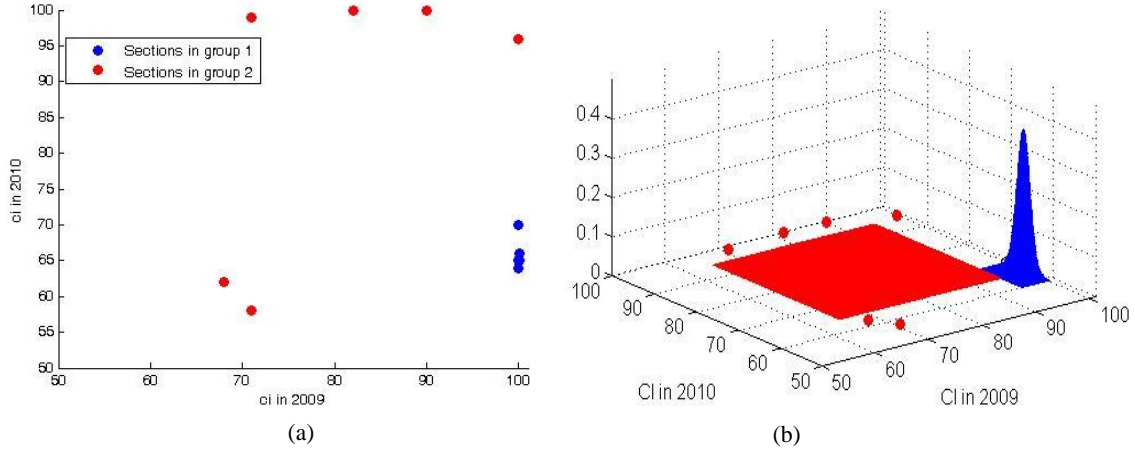


Figure 18 (a) Condition data points associated with the sections of Clustering Scenario 1 (b) PDFs fitted to the points of each cluster

Scenario # 2: In this clustering scenario, Cluster 2 of Scenario 1 is divided into two clusters (called Cluster 2 and Cluster 3) (see Figure 19). It can be seen from Figure 20 that the multivariate Gaussian PDF of Cluster 3 is tighter than that of Cluster 2. This indicates that Scenario 2 (as a whole) has more uniform clusters of pavement sections as compared to Scenario 1.

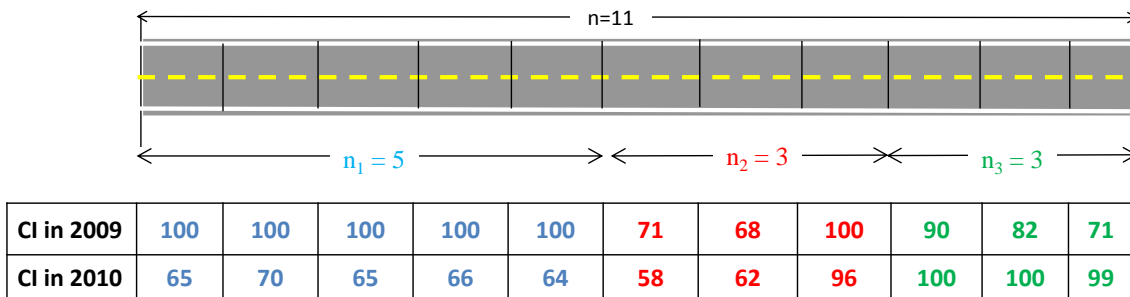


Figure 19 The second clustering scenario

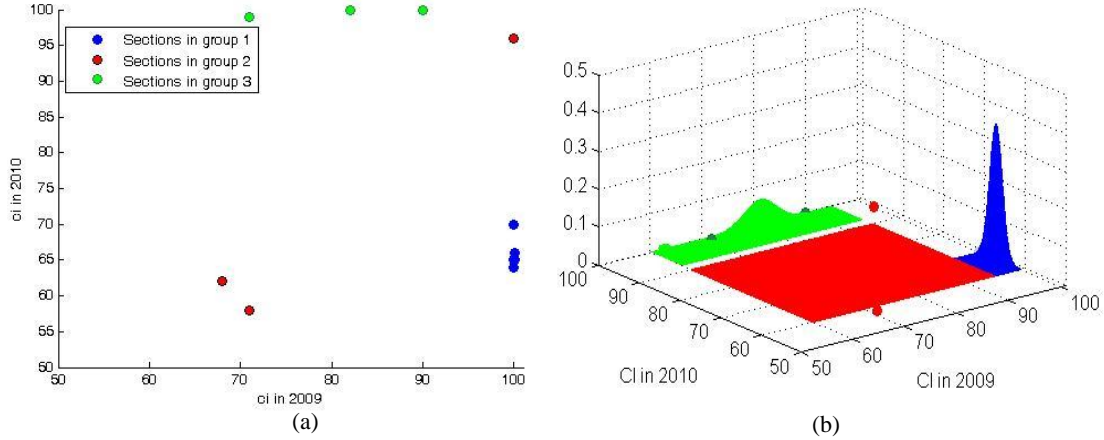


Figure 20 (a) the data points associated with the sections (b) PDFs fitted to the points of each cluster

While the above example uses only two years of condition data, the approach is applicable to any number of inspection cycles (i.e., years) of condition data. The condition history of a pavement section over m cycles can be represented by the m -tuple of $CI=(ci_1, \dots, ci_m)$, where m is the number of inspection cycles in the past for which condition data is available and CI is the condition index (or measurement).

This method seeks to determine the clustering scenario that best fits the data by segmenting each roadway in a pavement network in a way that maximizes the overall probability of the sections belonging to the clusters, considering prior knowledge of the distribution of M&R project lengths.

Most clustering methods such as k-means and distribution-based methods require the number of clusters as an input either directly or indirectly (Everitt et al. 2011). Similarly, the developed method requires some knowledge of the number of clusters per roadway corridor. The frequency distribution of past M&R project lengths is used as

prior knowledge (or a guide) for determining the cluster boundaries. Ignoring any prior knowledge of M&R project length would lead to a segmentation scheme that consists of one pavement section per cluster, which is unrealistic considering that M&R actions are often applied on a group of neighboring 0.5-mile sections and not just to a single segment. While the full history of annual M&R projects (along with their locations) would be very difficult to obtain from existing databases (Saliminejad and Gharaibeh 2012), the frequency distribution of project lengths can be determined from a sampling of past projects.

Since historical condition data generally is available for multiple inspection cycles, the developed method uses Euclidian spaces of more than three dimensions. The historical condition data of n sections for m cycles can be represented mathematically as: CI_i = vector of condition history for section i , where i is 1 to n .

For a roadway corridor with n sections, this research focused on finding the most likely clustering scheme (i.e., roadway segmenting scenario) that fit the historical condition data of these sections (CI_i). In mathematical terms, the probability that a corridor consists of a clustering scheme (g_1 to g_k), given that the condition data for the sections in this roadway corridor is CI_1 to CI_n , is computed based on the Bayes formula, as follows:

$$P(g_1, \dots, g_k | CI^1, \dots, CI^n) = \frac{P(CI^1, \dots, CI^n | g_1, \dots, g_k) \times P(g_1, \dots, g_k)}{P(CI^1, \dots, CI^n)} \quad (13)$$

where $P(CI_1, \dots, CI_n | g_1, \dots, g_k)$: Probability that the condition data is CI_1, \dots, CI_n , if the clustering scheme is g_1 to g_k .

$P(CI_1, \dots, CI_n)$: Probability of having condition vectors of CI_1, \dots, CI_n , regardless of the clustering scheme. In this problem, the clustering scheme is variable but the dataset is fixed. Thus, the parameter $P(CI_1, \dots, CI_n)$ is a constant.

$P(g_1, \dots, g_k)$: Probability that each cluster in the clustering scheme is an M&R project. This probability is determined based on the cluster length and prior knowledge of the PDF of the M&R project length.

Since $P(CI_1, \dots, CI_n)$ is a constant, maximizing Eq. (13) is equivalent to maximizing its nominator. Thus, the problem can be simplified as one of maximizing Eq. (14).

$$P(g_1, \dots, g_k | CI^1, \dots, CI^n) \propto P(CI^1, \dots, CI^n | g_1, \dots, g_k) \times P(g_1, \dots, g_k) \quad (14)$$

The probability of having an M&R project in a particular location is independent of the probability of having another M&R project in another location. In mathematical terms, this fact can be represented as follows:

$$P(g_1, \dots, g_k) = \prod_{i=1}^k P(g_i) \quad (15)$$

Also, by assuming the independence of the CI_i s, the parameter $P(CI_1, \dots, CI_n | g_1, \dots, g_k)$, can be computed as follows:

$$P(CI^1, \dots, CI^n | g_1, \dots, g_k) = \prod_{i=1}^n P(CI^i | g_1, \dots, g_k) \quad (16)$$

Substituting Eqs. (15) and (16), into Eq. (14), the probability that a highway consists of a clustering scheme (g_1 to g_k), given that the condition data for the sections in this highway is CI_i to CI_n , can be computed as follows:

$$P(g_1, \dots, g_k | CI^1, \dots, CI^n) \propto \prod_{i=1}^n P(CI^i | g_1, \dots, g_k) \prod_{j=1}^k P(g_j) \quad (17)$$

where $P(g_i)$ = the probability of having an M&R project with a length equal to g_i .

$P(CI_i | g_1, \dots, g_k)$ = the probability of having a condition data vector of CI_i for section i , given the clustering scheme g_1 to g_k . This probability can be calculated by fitting a multivariate Gaussian PDF onto the cluster of points (i.e., the cluster) to which CI_i belongs. The closer the similarity between the sections within this cluster (in terms of CI patterns and values), the higher this probability will be (due to a tighter PDF).

Finally, among all possible clustering scenarios, the one that gives the highest probability [calculated by Eq. (17)] is selected as the best clustering scenario. There are 2^{n-1} possible clustering scenarios for a roadway corridor with n sections. However, since conducting this outlier analysis on very few data points increases the chance of random effects, it is reasonable to limit the length of a cluster to a practical minimum value (e.g., six sections). The total number of possible clustering scenarios (with a minimum cluster length of n_{\min}) can be calculated as follows:

$$\sum_{k=1}^{n-k \times n_{\min}} \binom{n-k \times n_{\min} - 1}{k-1} \quad (18)$$

For example, a highway with $n=100$ sections can be clustered in $2^{99} \approx 6.4 \times 10^{29}$ different ways. Limiting the minimum number of sections to five, the total number of possible

clustering scenarios decreases to 1.9×10^{22} . Evaluating such a high number of possible clustering scenarios in the absence of a closed form solution would be virtually impossible without the use of efficient optimization techniques. A Genetic Algorithm (GA) was used to evaluate possible clustering scenarios, based on Eq. (17). GA is an optimization method based on natural selection. Basically, this method begins by randomly generating a set of solutions, then tests the solutions for their quality, reproduces new solutions by combining high quality solutions, and finally replaces old solutions with new ones. This process is repeated until a stopping criterion is met (Ashlock 2005).

Step 2: Detection of Outliers within Each Cluster

After clustering the sections within each roadway corridor, the next step in the developed outlier detection method is to detect outliers and differentiate them into gross and pseudo outliers. There are two levels to this method: the section and the point levels. At the point level, the condition data points that differ from other condition values within their cluster in any given cycle are detected (these detected data points are called outlier points). At the section level, the sections that have deterioration histories different from those of the other sections within their cluster are detected.

If an outlier point belongs to a section that has a different condition history from the other sections within its cluster, the section is likely to be affected by different local factors (e.g., drainage condition, subgrade condition, and M&R histories) and is less likely to be erroneous. This point is classified as a pseudo outlier. On the other hand, if an outlier point belongs to a section with a similar condition history to that of other

sections within its cluster that point is likely to be an erroneous observation and is classified as a gross outlier.

An outlierness index is defined for each condition data point within a cluster, based on the IQR of the pavement condition indicator within the cluster in any given cycle. The point outlierness index, or $POI(ci)$, represents the distance between the data point and the median, relative to the IQR itself. The $POI(ci)$ is computed as follows:

$$POI(ci) = \frac{d(ci)}{IQR_{ci}} \quad (19)$$

where: $d(ci)$ is the distance between the condition data point and the median. IQR is computed as the absolute difference between the 75th percentile and the 25th percentile (see Figure 21).

The advantages of this approach for detecting outliers are as follows:

- It allows for measuring the outlierness of all data points on a continuous scale, instead of using a hard classification of data as either an outlier or a non-outlier.
- It uses median and IQR (instead of mean and standard deviation), which are robust and are not affected by the presence of extreme outliers (Spiegelman et al. 2011). This is especially important for preventing the masking effect of the most extreme outliers.
- It is not parametric and does not require the data to follow a particular distribution (e.g., normal distribution). In the case of pavement condition data, since the outlier analysis is performed on multiple cycles of historical condition data, the input data is high-dimensional and very difficult to visualize. In

addition, sufficient knowledge of the distribution of the condition data is rarely present. As a result, non-parametric methods generally are expected to perform better than parametric methods.

- It accounts for the inherent variability within the cluster by considering the IQR.
- Condition data points with a POI(c_i) greater than a user-defined maximum threshold (called α) are identified as outliers.

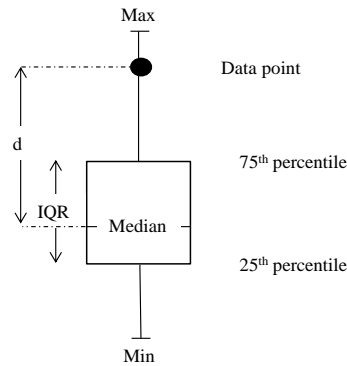


Figure 21 Outlier detection criteria

In contrast to detecting outlier points, the detection of outlying sections (i.e., multiple inspection cycles of condition data), is a multivariate problem. The methods for identifying univariate outliers are based on inarguable orderings of data values. There is no inarguable orderings for multivariate data. However, sub-ordering can be performed based on a distance measure (Jorma et al. 2000).

Euclidian Distance (ED) is the most popular distance measure [Eq. (20)] since it is easy to compute and to interpret (De Maesschalck et al. 2000). However, ED cannot

be used as a measure for comparing the dissimilarity of curves in different clusters since it depends upon the scale and cannot take the correlation in the dataset into account.

$$ED(CI) = \sqrt{(ci_1 - \overline{ci_1})^2 + \dots + (ci_m - \overline{ci_m})^2} \quad (20)$$

where ci_i = the condition index of the section in cycle i and $\overline{ci_i}$ = the average condition index of the cluster to which the section belongs in cycle i . Eq. (20) also can be rewritten in a matrix form, as shown in Eq. (21).

$$ED(CI) = \sqrt{(CI - \overline{CI})^T (CI - \overline{CI})} \quad (21)$$

where T = the matrix transpose symbol. Like DE, Mahalanobis Distance (MD), as defined in Eq. (22), is a measure of distance. By including a covariance matrix (S), MD takes the correlation of data into account. Using MD, it is possible to quantify the distance of an m -dimensional point $CI=(ci_1, \dots, ci_m)$ from a cluster with a mean $M=(\overline{ci_1}, \dots, \overline{ci_m})$ over the cluster to which the section belongs. Research has shown that computing the MD for each point and detecting those points where the MD is the largest (as outliers) works well when a masking effect is not present (Rousseeuw and Leroy 1987). Since in this work the masking effect is controlled by the use of the robust measures of median and IQR, the MD can safely be used for detecting outliers.

$$MD(CI) = \sqrt{(CI - \overline{CI})^T S^{-1} (CI - \overline{CI})} \quad (22)$$

where S = the covariance matrix.

An outlierness index is defined for each pavement section within a cluster based on the MD IQR. As shown in Figure 21, the SOI(CI) represents the distance between the sections of MD and the median, relative to the IQR itself. SOI(CI) is computed as follows:

$$SOI(CI) = \frac{d(MD)}{IQR_{MD}} \quad (23)$$

where $d(MD)$ is the distance between the MD point and the nearest edge of the midspread box.

Finally, sections with an SOI(CI) greater than a user-defined maximum threshold (called β) can be identified as outliers.

Differentiating Gross and Pseudo Outliers

If a condition data point is detected as an outlier, it can either be because that data point is likely to be erroneous (a gross outlier) or because the section has a different behavior from that of the other sections within the cluster (a pseudo outlier). If the POI(ci) of the data point is exceptionally high in one inspection cycle, it should increase suspicion that the conditions data in that specific cycle is erroneous. On the other hand, if both the POI(ci) and SOI(CI) are high, the section itself may be influenced by various local factors (e.g., drainage, subgrade, and construction) that cause the section to behave different from that of other sections in the cluster. In other words, the outlierness of the condition data (in this case) is an indication of unusual section behavior, rather than errors in the data.

Application and Validation of the Developed Method

The developed method was implemented using ArcMap, Python, and MATLAB (code provided in Appendix A) and applied as a case study to the roadway network in TxDOT's Bryan District. The Bryan network consisted of 7,299 0.5-mile pavement sections (3,650 centerline miles) in 2011; it is located in a wet-warm climatic zone with poor subgrade quality.

While it is not possible to validate the developed method using “ground truth” values of condition data since most outliers belong to pavement sections that were in poor condition and have already been repaired, the following efforts were made to provide clues to the validity of the developed method.

- First, the developed method and the classic outlier detection method (currently dominant in the field of pavement management) are applied to the case study and the results are compared.
- Second, independent audit data is used as a reference to which the original data was compared in three ways: audit versus gross outliers, audit versus normal data, and audit versus whole original data (combined normal and gross outliers combined as one dataset).
- Third, possible explanations of the detected pseudo outliers are drawn from pavement engineering knowledge and principles.

Application of Proximity-Based Outlier Detection Method and Classic Method to Case Study

To apply the developed outlier detection technique, a prior knowledge of M&R project length and threshold value for POI(α) (i.e., α) and SOI(β) (i.e., β) needed to be established. M&R project length data was obtained from TxDOT records. Due to the economy of scale and other practical considerations, M&R actions are often applied to a number of neighboring sections and not just the single section. Figure 22 shows the relative frequency distribution of the length of the projects in 2010 in the Bryan district.

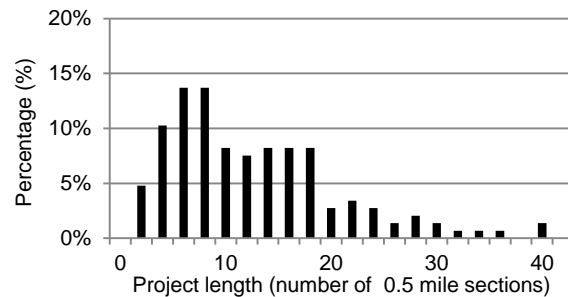


Figure 22 Relative frequency distribution of the length of the M&R projects in 2010 in the Bryan district

The outlier detection method was then run multiple times, using various α and β values as inputs. Finally, $\alpha=2.5$ and $\beta=1.5$ were found to be the values best able to reproduce the detected outliers. These values were then used as input values for analyzing the Bryan network.

The clusters that result from this classic clustering method are not necessarily in close proximity to one another. This is evident in the example cluster shown in Figure 24, used for the Bryan district case study. This cluster consists of pavement sections with a wet-warm climate, poor subgrade, and medium traffic, that fall into pavement family A, with the last treatment type being preventive maintenance [shown as the medium traffic, pavement family A, and last treatment type being preventive maintenance (MAP) cluster in Figure 24].

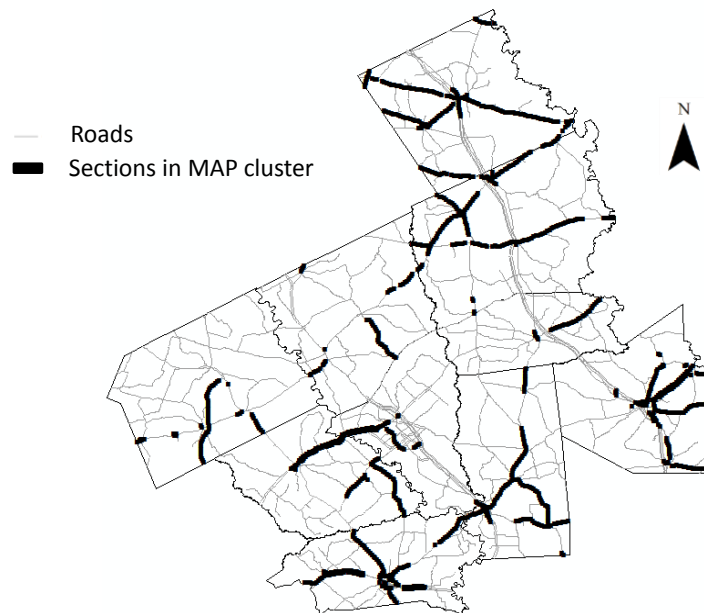


Figure 24 Geographic locations of pavement sections with medium traffic, pavement family A, and last treatment type being preventive maintenance, in the Bryan district in 2011 (classic clustering method)

Figure 26 to Figure 38 show the results of the developed outlier detection method for Highway 36 in Bryan District, as examples. The left side of each Figure shows gross and pseudo outliers in each cluster along Highway 36. Pseudo outliers are shown in red curves and gross outlier points are shown in red. It can be seen that the proposed method has identified clusters that are free of outliers (i.e., clusters # 1, 5, 9 and 10), clusters with gross outliers and no pseudo outliers (i.e., clusters # 2, 3, 4, 6, and 12), and clusters with gross and pseudo outliers (i.e., clusters # 7 and 8). This application suggests that the proposed method is capable of detecting all possible outlier scenarios.

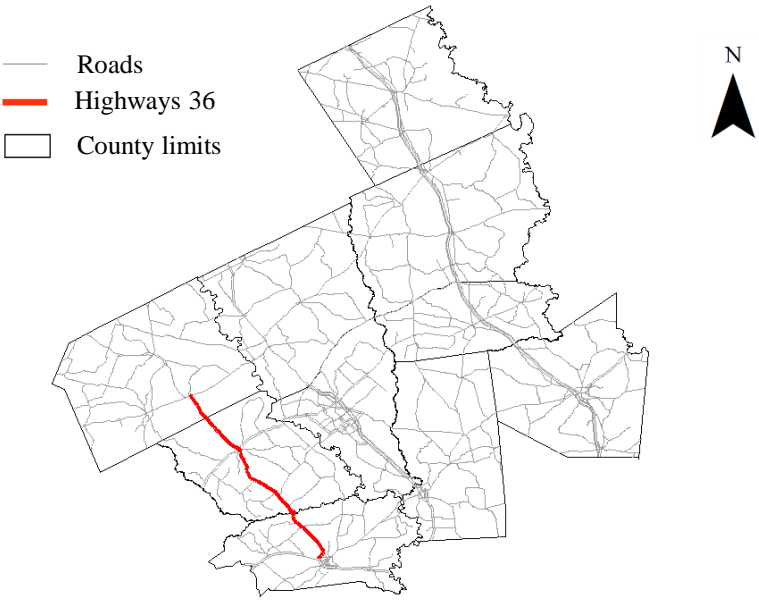


Figure 25 Location of highway 36 on Bryan District map

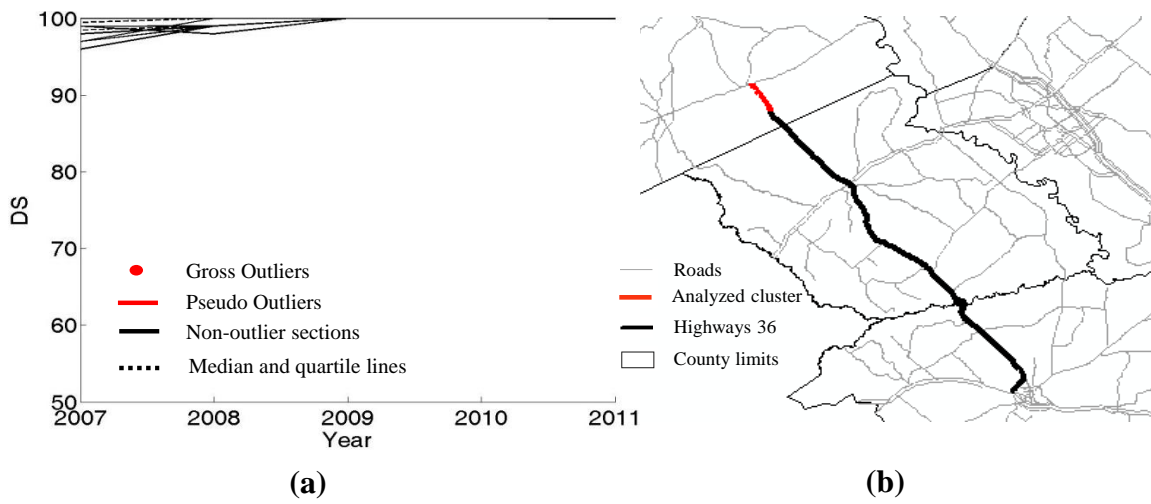


Figure 26 Cluster 1 consisting of eight pavement sections, (a) Historical condition data of sections (b)

Location on map

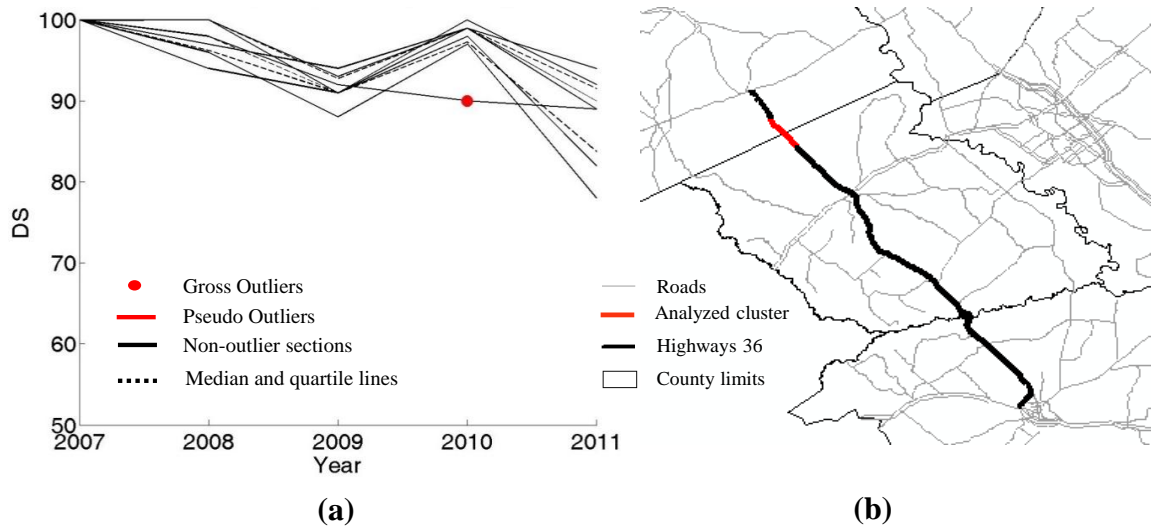


Figure 27 Cluster 2 consisting of seven pavement sections, (a) Historical condition data of sections

(b) Location on map

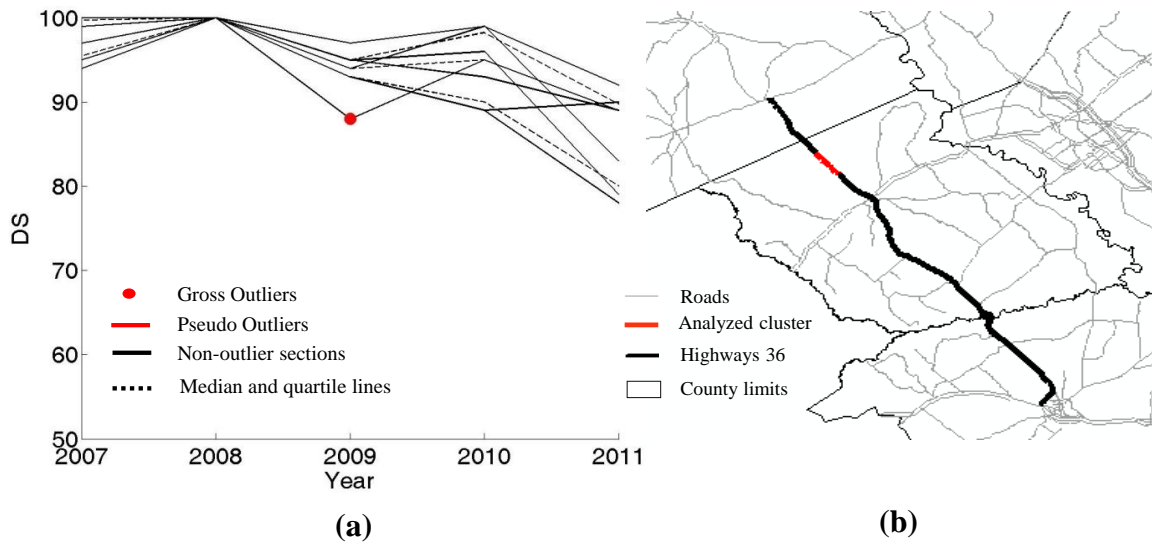


Figure 28 Cluster 3 consisting of seven pavement sections, (a) Historical condition data of sections
(b) Location on map

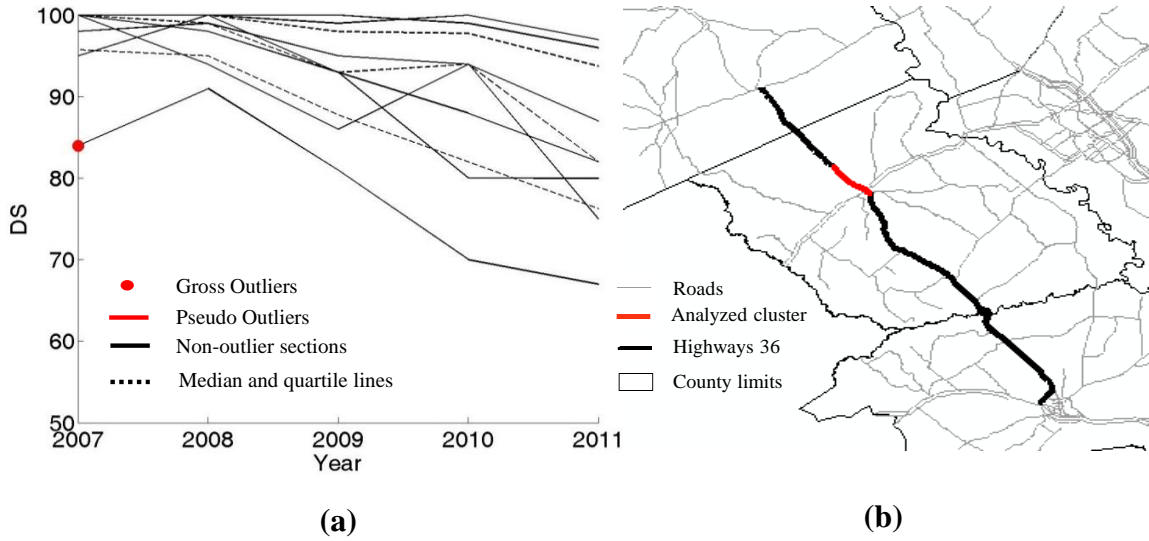


Figure 29 Cluster 4 consisting of seven pavement sections, (a) Historical condition data of sections
(b) Location on map

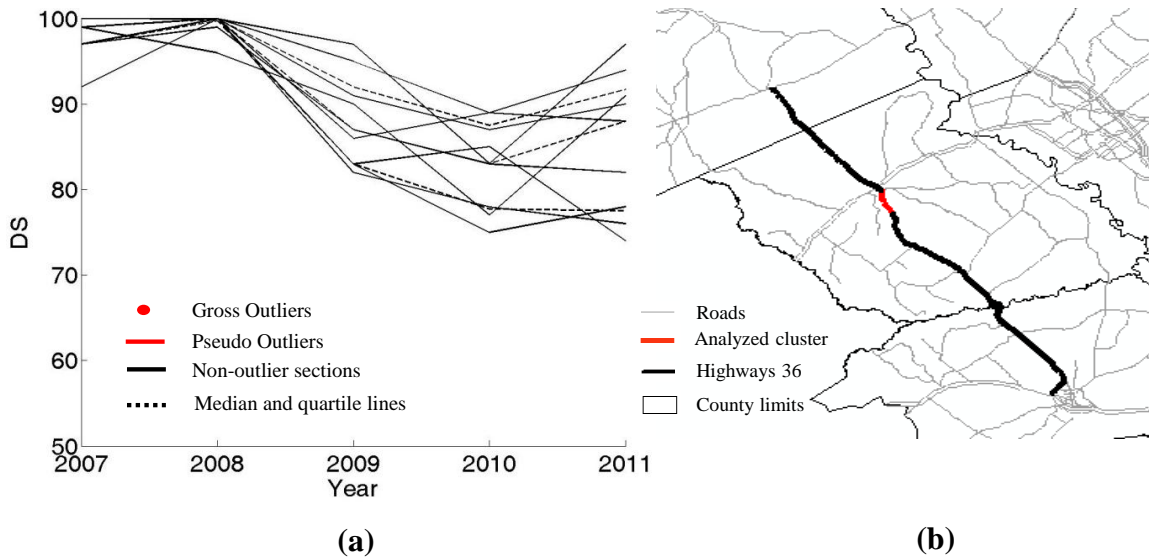


Figure 30 Cluster 5 consisting of nine pavement sections, (a) Historical condition data of sections (b) Location on map

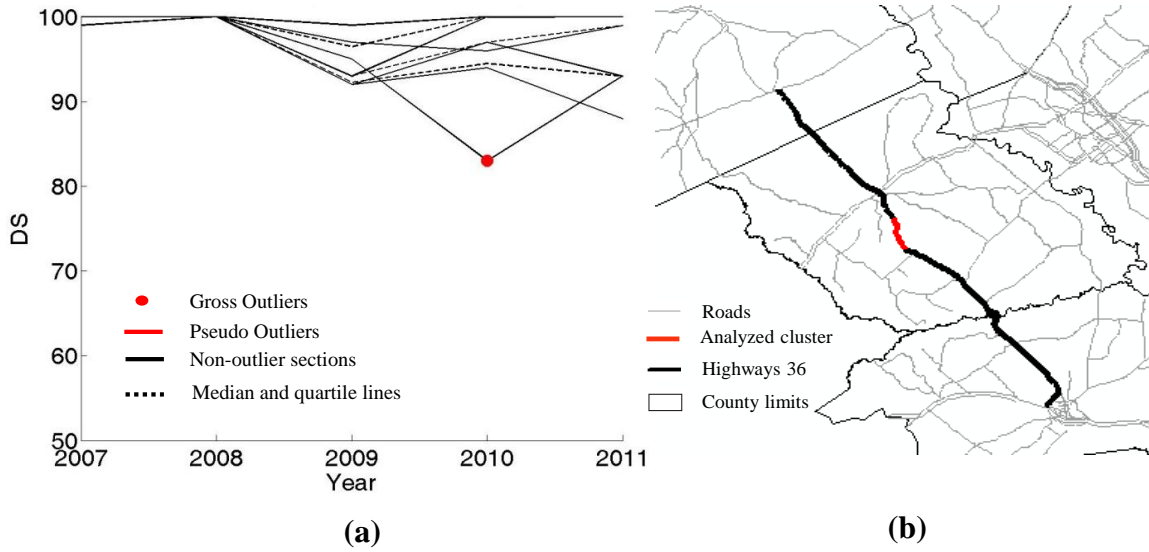


Figure 31 Cluster 6 consisting of seven pavement sections, (a) Historical condition data of sections (b) Location on map

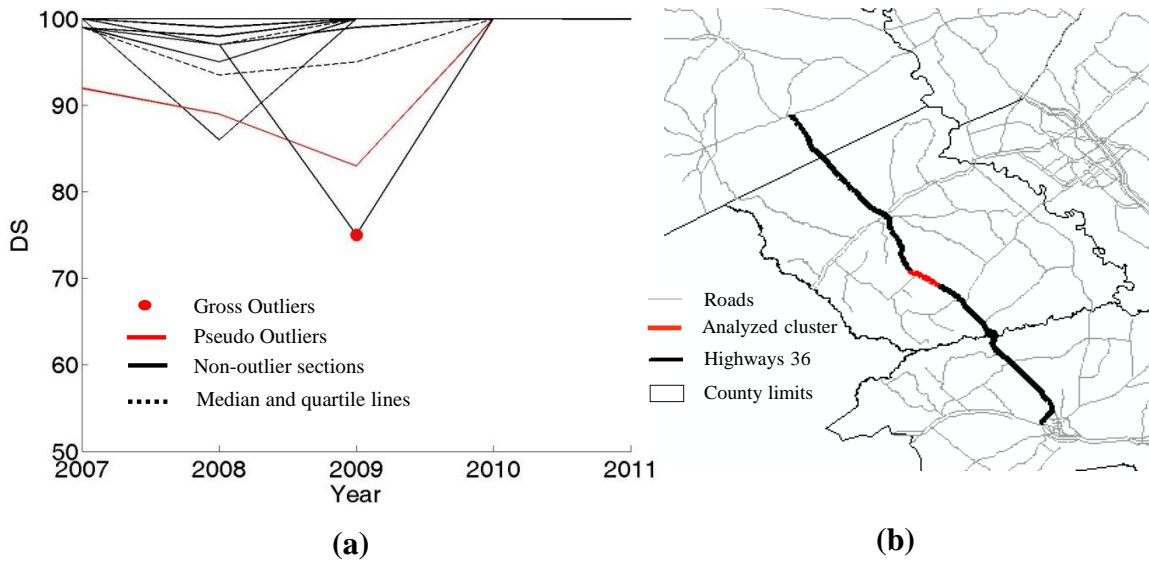


Figure 32 Cluster 7 consisting of nine pavement sections, (a) Historical condition data of sections (b) Location on map

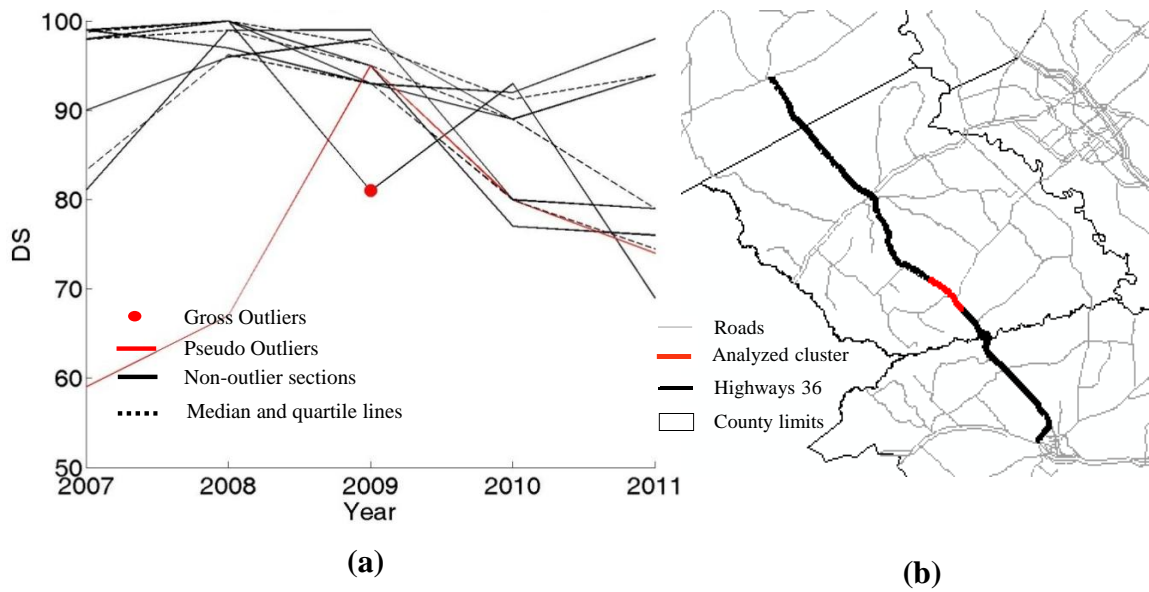


Figure 33 Cluster 8 consisting of seven pavement sections, (a) Historical condition data of sections (b) Location on map

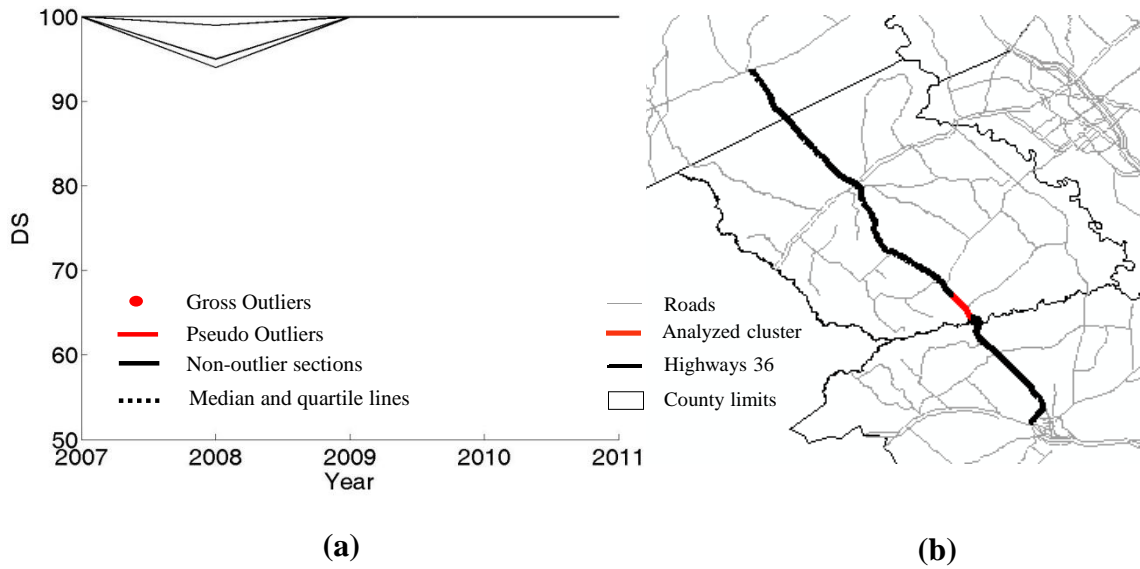


Figure 34 Cluster 9 consisting of nine pavement sections, (a) Historical condition data of sections (b) Location on map

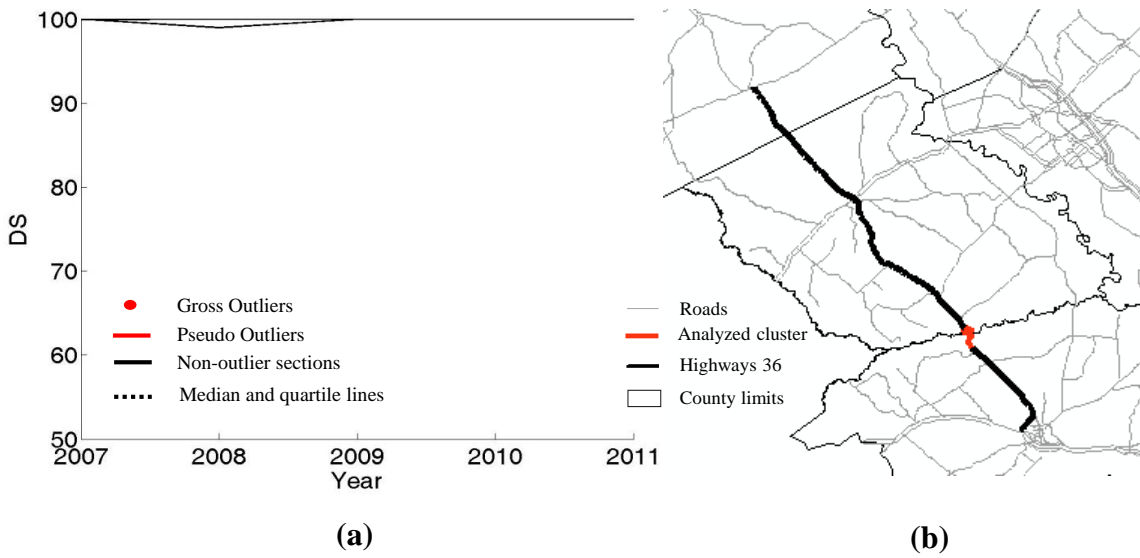


Figure 35 Cluster 10 consisting of seven pavement sections, (a) Historical condition data of sections (b) Location on map

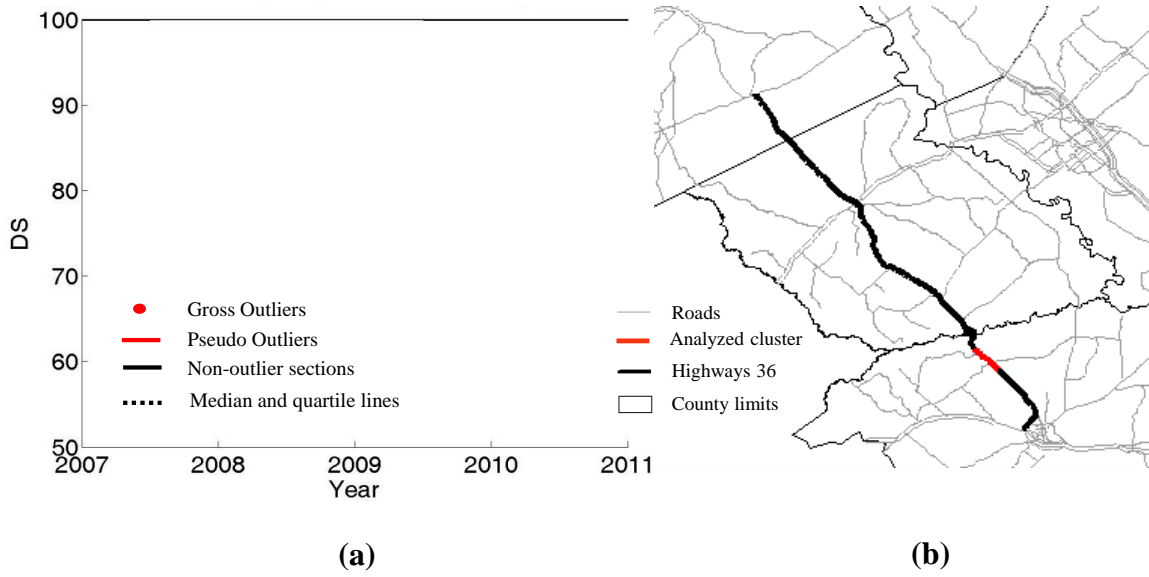


Figure 36 Cluster 11 consisting of eight pavement sections, (a) Historical condition data of sections (b) Location on map

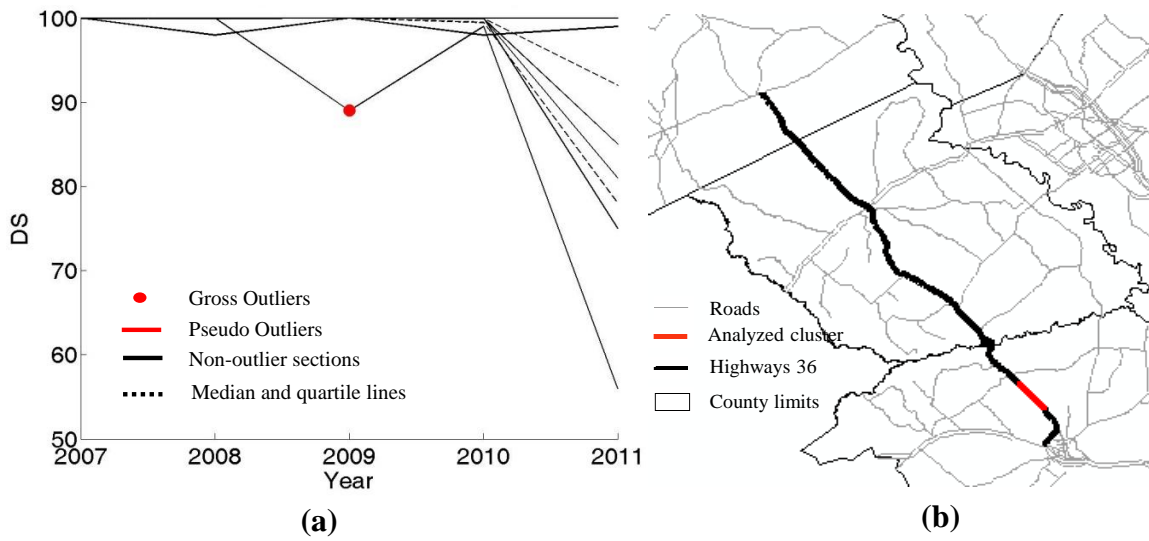


Figure 37 Cluster 12 consisting of eight sections, (a) Historical condition data of sections (b) Location on map

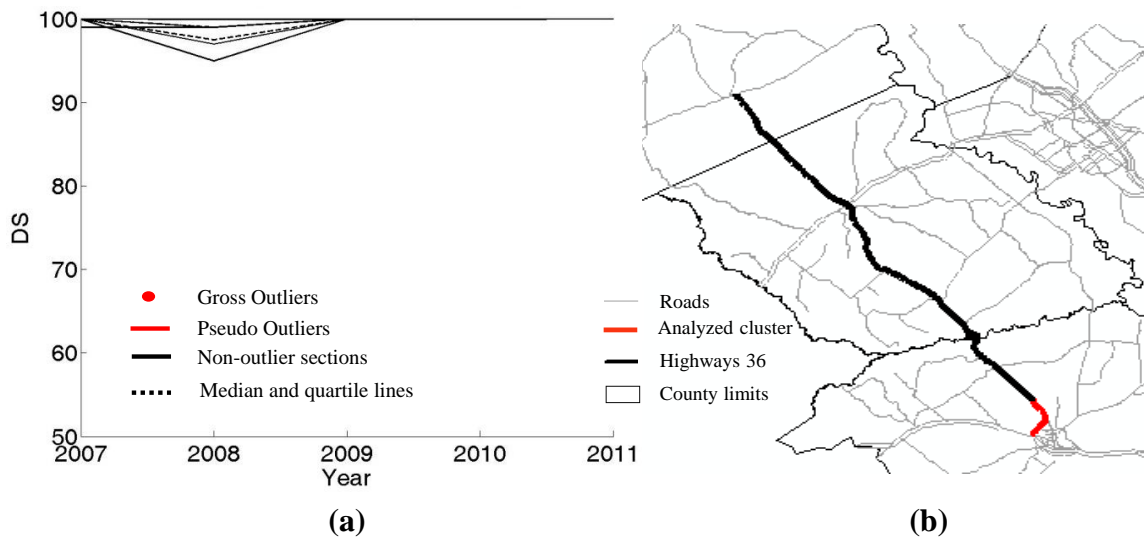


Figure 38 Cluster 13 consisting of seven sections, (a) Historical condition data of sections (b) Location on map

Comparison of Proximity-based to Classic Outlier Detections

The classic form of outlier detection, as has been explained above, was applied to pavement sections with a wet-warm climate, poor subgrade, and medium traffic, in pavement family C, and with the last treatment type being light rehabilitation; the results are presented in Figure 39. The points with a distance of more than $1.5 \times IQR$ from the edges of the midspread box are considered outliers. The crosses represent outlier data. As shown in this Figure, although all of the pavement sections in this cluster share many physical and environmental properties, their overall performance is still highly variable.

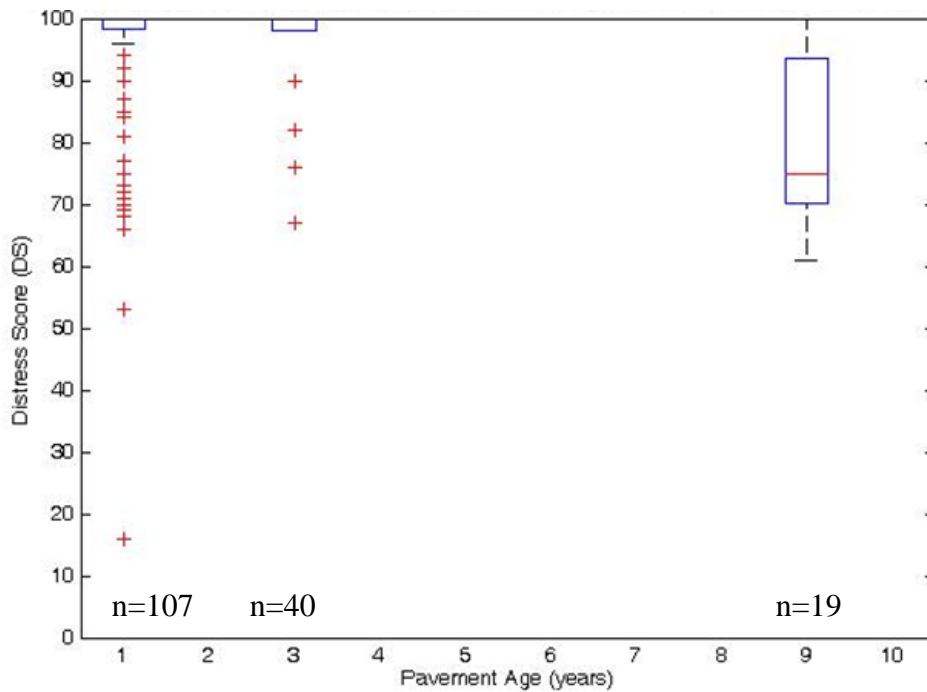


Figure 39 Classic outlier detection method

The implicit assumption behind this method is that the clustered sections behave in a uniform fashion; therefore, outlier sections are suspicious. In practice, however, high levels of variation can still be observed in the behaviors of these so called “uniform” families. As an example, in the case of pavements, AASHTO (2002) considers the following factors to have some influence on pavement performance: subgrade/foundation, traffic loading (axle loads, classification, and forecasting), M&R history, climatic factors (e.g., hourly temperature, precipitation, solar radiation, cloud cover, and wind speed), paving materials properties, structure of the pavement, construction factors, and drainage. However, in practice, only a few of these factors usually are used as the criteria to create uniform clusters because not all of these forms

of data are available for all pavement sections. Even if available, continuum data are recorded as categorical values and, due to practical limitations, the variations within each category contribute to the performance variability within each cluster. For example, temperature is presented in categories of hot and cold instead of as hourly temperature values. Therefore, it is not surprising to see a highly variable performance within a so-called uniform cluster.

The gross outliers detected by the proximity-based method were plotted, as were the outliers detected by the classic method in 2011 (see Figure 40). As Figure 40 indicates, the developed method and the classic method agree in some cases, but disagree in many others. The disagreements between the two methods can be explained as follows.

False positive detection of outliers: The classic method falsely identified some data points as outliers when they were not outliers (see Figure 41), as compared to other data in their proximity (i.e., pavement sections subjected to the same local condition). This result demonstrates the susceptibility of the classic method to the swamping effect. One outlier is considered to swamp a second observation if the latter can be considered an outlier only in the presence of the first outlier. In other words, after the deletion of the first outlier, the second observation becomes a non-outlying observation (Ben-Gal 2005). The impact of this type of misidentification is significant, such that these outliers (shown as yellow ovals) form the majority of outliers detected by the classic method.

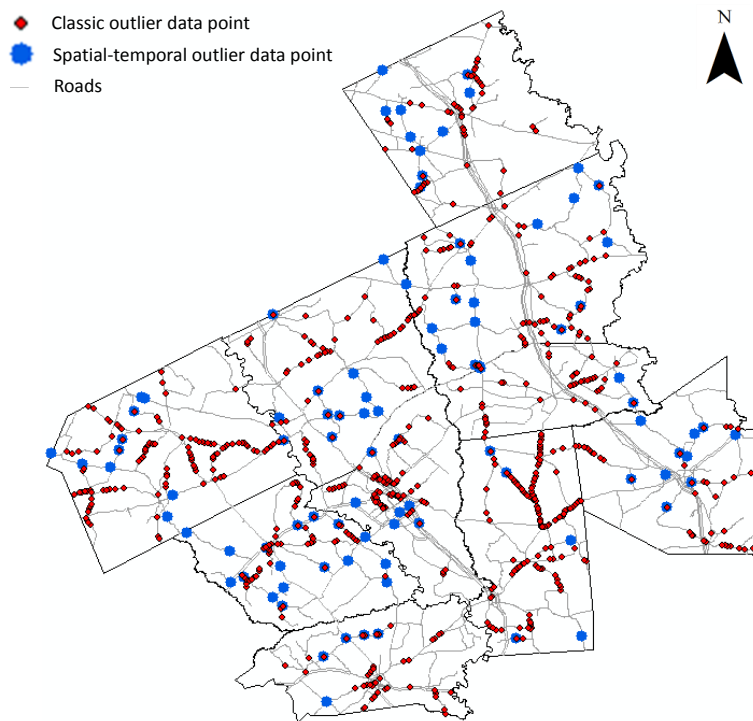


Figure 40 Outliers detected by the classic method versus outliers detected by the proposed spatial-temporal method in Bryan District in 2011

False negative detection of outliers: As shown in Figure 42, the classic method missed true outliers (i.e., it produced false negative detections) due to its susceptibility to the masking effect. Masking effects occur when one outlier masks a second outlier if that second outlier can be considered an outlier by itself, but not in the presence of the first outlier. Thus, after the deletion of the first outlier the second instance emerges as an outlier (Ben-Gal 2005).

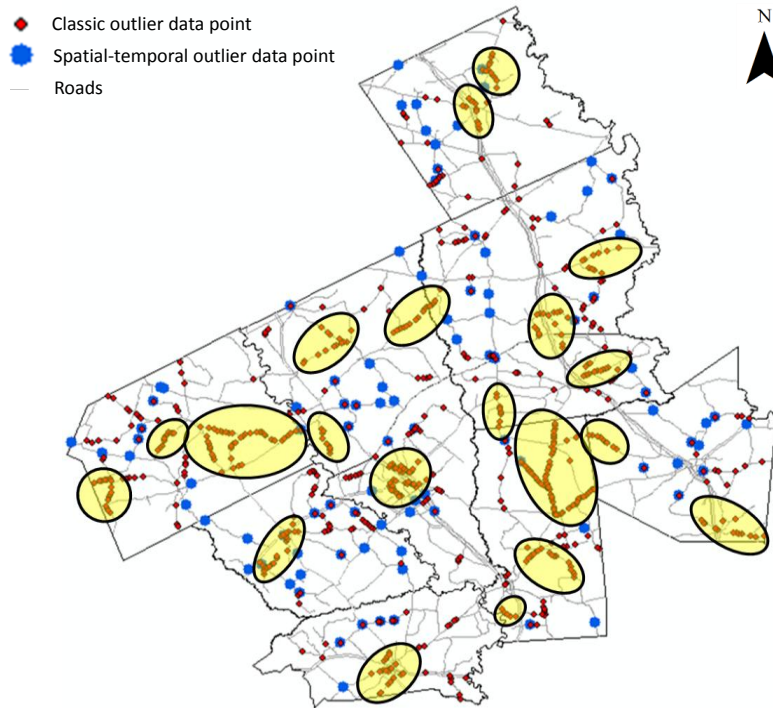


Figure 41 Examples of false positive outliers detected by the classic method

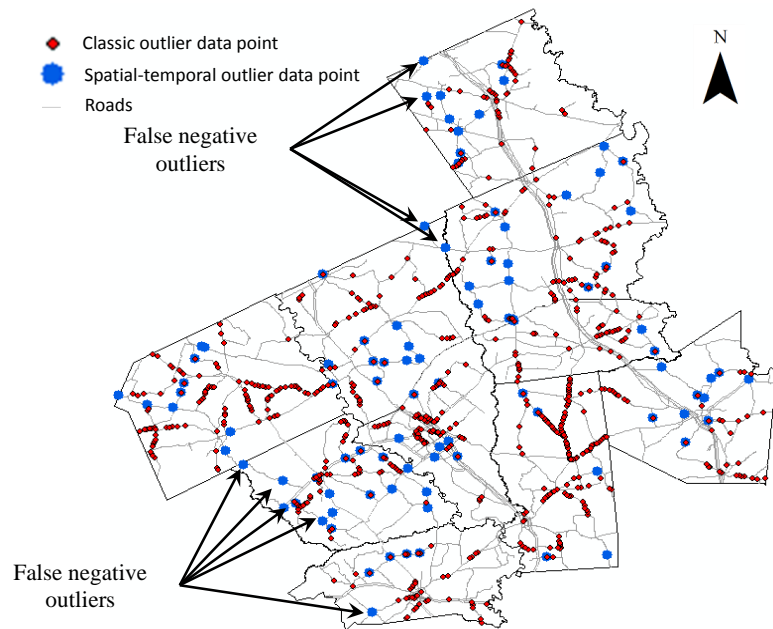


Figure 42 Examples of false negative outliers detected by the classic method

Validation of the Developed Method Using Audit Data

As discussed earlier, audit data is available for approximately five percent of the case study network. The audited pavement sections are selected randomly by the independent auditors, as part of a data quality control process. The vendor's condition data of a county is rejected if more than 15% of the data within that county shows a difference of 10 points or higher in the Distress Score from the auditors data. The descriptive statistics of audited data are provided in Table 2.

Table 2 Descriptive statistics of audited DS data

Data Type	Number of Pavement Sections	No. of Data Points	Mean DS	Standard Deviation of DS
Normal Data	1105	1272	93.2	12.68
Gross Outliers	29	29	73.2	14.2
Whole Original Data (both normal and gross outliers)	1134	1301	92.8	13.1
Audit Data	1134	1301	97.3	8.6

The audit data was used as a reference to which the original data was compared in three ways: audit versus gross outliers, audit versus normal data, and audit versus whole original data. In each case a difference in DS (i.e., ΔDS) was computed as the difference between the audit DS value and the corresponding original DS value. Table 3 compares DS of audit and original ratings of gross outliers. Audit versus normal data are also compared in Appendix B.

Table 3 Comparison between audit and original ratings of gross outliers

Rating #	Highway Name	BRM	BRM Distance	Year	Original DS	Audit DS	Δ DS
1	FM0979 K	0622	0.5	2011	78	100	22
2	FM1179 K	0408	1	2011	67	68	1
3	FM2223 K	0402	0.5	2011	65	100	35
4	US0190 K	0708	1	2011	81	100	19
5	FM0487 K	0570	1	2010	90	98	8
6	FM2223 K	0404	0	2010	69	69	0
7	FM3242 K	0406	0.5	2010	67	100	33
8	SH0014 K	0384	0	2010	60	61	1
9	FM0050 K	0412	1.5	2009	58	59	1
10	FM0060 K	0618	0.5	2009	92	91	1
11	FM0060 K	0632	0.6	2009	51	52	1
12	FM1915 K	0398	0	2009	39	42	3
13	FM2027 K	0388	0	2009	54	54	0
14	US0079 K	0492	0	2009	60	99	39
15	FM0060 K	0606	1	2008	75	100	25
16	FM0158 R	0618	1.6	2008	95	100	5
17	FM0979 K	0608	1	2008	68	69	1
18	SH0036 K	0548	1.5	2008	86	100	14
19	US0290 L	0664	1	2008	82	99	17
20	FM0046 K	0604	1	2007	53	71	18
21	FM0908 K	0598	0.5	2007	71	100	29
22	FM0974 K	0626	1	2007	78	100	22
23	FM0979 K	0628	1	2007	82	99	17
24	FM1687 K	0608	1	2007	68	90	22
25	FM1687 K	0608	1.5	2007	76	61	15
26	FM1696 K	0656	0.5	2007	90	100	10
27	FM3090 K	0430	1	2007	90	100	10
28	SH0090 K	0412	1.5	2007	89	81	8
29	US0190 K	0608	1	2007	88	90	2

Table 3 shows the Δ DS standard deviation and absolute mean values for the three comparisons. It can be seen in Table 4 that Δ DS for the gross outliers has higher

standard deviation and higher mean absolute value than those for the normal data and the combined original data.

Table 4 Mean and standard deviation of Δ DS for the three comparisons (audit versus gross outliers, audit versus normal data, and audit versus whole original data)

Comparison	Mean Absolute ΔDS	Standard Deviation of ΔDS
Audit vs. Normal Data	4.8	9.5
Audit vs. Gross Outliers	13.1	11.5
Audit vs. Whole Original Data (both normal and gross outliers)	5.0	9.6

Unequal variance T-Test is used to determine whether Δ DS of normal data and gross outliers are significantly different. For 95% confidence ($\alpha=0.05$), the probability value (p-value) associated with the hypothesis that “the difference between audit and original data (Δ DS) come from distributions with equal means” is calculated as 0.0082. The hypothesis is rejected since p-value is smaller than α . In other words, it can be inferred with the confidence of 95% that Δ DS of normal data and gross outliers come from populations with significantly different means. Normal and gross outlier populations are both generated in the clustering and outlier detection processes. Therefore, the populations are not randomly chosen and may not be necessarily independent as required by t-test. However, audited normal and gross outlier populations can safely be

assumed independent enough for the t-test since they are randomly-chosen subsets of the normal and gross outlier populations.

Possible Causes of Pseudo Outliers

The developed method classifies the data points into three categories: normal data, gross outliers, and pseudo outliers. In addition, it is possible to classify the pseudo outliers based on their condition data as "up" or "down" outliers. Up outliers indicate higher than expected condition (i.e., pavement sections showing exceptionally high performance) and down outliers indicate lower than expected condition (i.e., pavement sections showing exceptionally poor performance). Poor drainage, expansive soil, over-stabilized base course, under designed structure, and modified base course can cause unexpected pavement behavior (NCHRP 1997; Lytton et al. 2005).

Possible explanations of the detected pseudo outliers are drawn from the literature and are provided in Table 5. Having different construction and M&R history and materials, or higher or lower traffic load in comparison to their neighbors are also possible causes of outlierness for all sections. In Table 5, "High" indicates that the section is in the top 10% of the sections in terms of cracking, rutting, or ride quality performance (i.e., good performance). "Low" indicates that the section is in the bottom 10% of the sections in terms of cracking, rutting, or ride quality performance (i.e., poor performance).

Table 5 Possible explanations for detected pseudo outliers

Highway Name	BRM¹	BRM Dist.²	Up/Down Outlier	Cr. Perf.	Rut. Perf.	Ride Perf.	Possible Causes of Outlierness
SH0036 K	0552	0.5	Down	Low	Low	-	Under Designed Structure, Poor Drainage
FM0111 K	0428	0	Down	-	Low	-	Poor HMA Mix Design
FM0159 K	0434	0	Down	High	Low	Low	Poor Drainage, Under Designed Structure
US0190 R	0672	1.1	Down	Low	Low	High	Under Designed Structure, Poor Drainage
US0190 K	0688	0	Up	-	-	-	
FM0974 K	0634	1.5	Up	-	Low	-	
US0079 K	0484	1.5	Down	High	-	High	
FM2446 K	0616	1.5	Down	High	-	-	
FM0976 K	0598	0	Down	High	High	-	Extensive Patching
SH0090 K	0428	0	Up	High	-	-	Unrecorded Repair
FM1696 K	0650	0.5	Up	-	-	-	
SH0006 K	0560	1	Up	High	High	High	Unrecorded Repair, Modified Base
FM2446 K	0616	1.5	Up	High	High	-	Unrecorded Repair
FM0391 K	0398	1.5	Up	High	-	Low	Unrecorded Repair
FM1373 K	0600	1.5	Up	-	-	-	
US0190 K	0634	1.5	Down	-	High	-	
FM0485 K	0576	1.5	Down	-	Low	Low	Poor HMA Mix Design
FM0975 K	0420	0	Down	Low	-	Low	Over-Stabilized Base
FM0975 K	0424	1	Up	Low	High	Low	Over-Stabilized Base
FM0159 K	0434	0	Up	High	-	-	Unrecorded Repair
FM1600 K	0402	1	Down	Low	-	Low	Expansive Soil, Over-Stabilized Base
FM2346 K	0386	0	Up	High	-	-	Unrecorded Repair
FM3058 K	0568	1	Down	High	High	-	Extensive Patching
SH0036 K	0552	0.5	Up	-	High	High	Modified Base
US0190 K	0622	1	Down	Low	-	-	Over-Stabilized Base
FM0111 K	0424	1	Up	-	High	-	
FM0696 K	0596	1.5	Up	High	-	-	Unrecorded Repair
FM0908 K	0594	1.5	Down	High	-	-	
FM1486 K	0412	1.5	Down	High	High	-	Extensive Patching
FM1600 K	0402	1	Down	-	High	Low	
FM2027 K	0386	0.5	Down	-	-	-	

¹ Beginning reference marker

² The distance from a beginning reference marker to beginning of the pavement section

Table 5 continued

Highway Name	BRM	BRM Dist.	Up/Down Outlier	Cr. Perf.	Rut. Perf.	Ride Perf.	Possible Causes of Outlierness
FM0908 K	0594	1.5	Down	High	-	-	
FM0975 K	0420	0	Down	Low	-	Low	Over-Stabilized Base
FM1600 K	0402	1	Down	Low	-	-	Over-Stabilized Base
FM1696 K	0650	0.5	Down	High	High	-	Extensive Patching
FM2446 K	0620	0.5	Down	High	High	-	Extensive Patching
SH0090 K	0416	1	Down	High	High	-	Extensive Patching
SH0105 K	0666	0.5	Down	-	High	High	
US0190 K	0622	1	Down	Low	High	-	Over-Stabilized Base
US0190 K	0634	1.5	Up	High	High	Low	Unrecorded Repair
US0190 K	0634	1.5	Up	High	Low	-	Unrecorded Repair
FM3061 K	0404	1	Up	High	-	-	Unrecorded Repair
FM3061 K	0404	0.5	Up	High	-	-	Unrecorded Repair
FM0486 K	0500	0.5	Up	High	-	-	Unrecorded Repair
FM2269 K	0394	1	Up	-	-	-	
US0079 K	0484	1.5	Down	-	High	High	
FM0391 K	0398	1.5	Down	High	High	Low	Extensive Patching
FM0437 K	0586	1.5	Up	-	High	-	
SH0021 K	0620	0.5	Down	Low	High	-	Over-Stabilized Base

Impact of Gross Outliers

Gross outliers are extreme either in up or down direction. Therefore, the decision to omitting or keeping them can impact the PMS outputs. To investigate this effect, the test-bench PMS (introduced earlier in Section 3) is run once with gross outliers and once after replacing them with average DS of the cluster they belong to (i.e., their neighboring sections). The results of this investigation are presented in Table 6.

Table 6 PMS outputs before and after omitting outliers

	PMS Output	2011	2012	2013	2014
Without Outliers	Needed Budget	\$ 16,622,455	\$ 36,739,082	\$ 9,965,984	\$ 2,995,837
	DN (%)	93.92%	90.70%	93.98%	97.57%
	Maintenance (%)	3.61%	2.79%	5.21%	2.41%
	Rehabilitation (%)	2.47%	6.51%	0.80%	0.01%
	Good / Very Good (%)	90.18%	90.05%	90.15%	90.42%
	Fair (%)	2.91%	6.94%	5.36%	6.93%
	Poor / Very Poor (%)	6.91%	3.01%	4.48%	2.65%
With Outliers	Needed Budget	\$ 21,867,139	\$ 35,998,720	\$ 9,955,038	\$ 1,983,233
	DN (%)	92.30%	90.76%	93.92%	98.39%
	Maintenance (%)	4.37%	2.79%	5.29%	1.61%
	Rehabilitation (%)	3.34%	6.45%	0.79%	0.00%
	Good / Very Good (%)	90.76%	90.41%	90.75%	90.27%
	Fair (%)	2.79%	6.82%	5.07%	6.60%
	Poor / Very Poor (%)	6.45%	2.77%	4.19%	3.13%

The results suggest that replacing gross outliers with the average DS of their cluster decreases the amount and budget of needed maintenance and rehabilitation in the first year of the planning period. The needed M&R budget decreased by 24% (approximately \$16.6 million vs. \$21.9 million). This suggests that the down outliers overwhelm the up outliers, leading to lower estimated budget when both are eliminated. The needed budget in the next years of planning period, however, is not highly affected.

Section Summary and Concluding Remarks

This section presents the development of a new technique for detecting gross and pseudo outliers in pavement condition data. The developed technique integrates concepts and methods from GIS, Bayesian statistics and Genetic Algorithms to cluster homogenous neighboring pavement sections and then identify gross outliers, as well as pseudo

outliers. Application of the developed technique to a case study (Bryan district) demonstrates how the consideration of geographic proximity and historical patterns can lead to more informative outlier analysis of pavement condition data. The developed technique can potentially be used for:

- Performance modeling: After detecting outliers using the outlier detection method developed for this research, gross outliers, as likely erroneous values, should be further investigated before being included in performance modeling. Pseudo outliers, on the other hand, should be modeled separately because they have different behaviour.
- Data quality control: Identifying the quantity and severity of gross outliers can be a more efficient and less costly data quality assessment framework [see (Buchheit et al. 2005), for example].
- Training condition inspectors: The detected outliers can potentially be used to assess the efficiency of infrastructure condition inspection training programs by checking if the inspectors are able to identify pseudo and gross outliers.
- Improving engineering practices: Comparing pavement section with normal data to pseudo outliers (i.e., pavements that perform exceptionally well or poor) can provide insights regarding the effectiveness of various designs, construction methods, and materials.

6. A SPATIAL-BAYESIAN TECHNIQUE FOR IMPUTING PAVEMENT NETWORK REPAIR DATA

This section presents the development of a new technique for estimating construction and M&R history data from pavement condition data. The developed technique employs Bayesian and spatial statistics in a GIS environment to search for evidence of repair in the condition data (current and past) of group of adjacent pavement sections. While obtaining these data from actual construction records maybe ideal, it may not be practical because construction data and condition data are often collected separately and stored separately in legacy databases (FHWA 2010).

Developed Data Imputation Method

Traditionally, probability is defined for a random event as the relative frequency with which an event occurs in a set of repeated trials. However, in the case of missing M&R history, unknown parameters do not originate from random experiments. Instead, there is uncertainty arising from having insufficient information and not from randomness. As a result, traditional probability concepts are not suitable for addressing this particular problem. To deal with such problems, Bayesian statistics interpret uncertainty as a result of insufficient information and interprets probability as the apparent validity of a hypothesis based on the state of knowledge (Park et al. 2008). The details of Bayesian statistics are well-documented in many textbooks [see for example Bolstad (2007)]

To deal with noisy (i.e., highly variable) pavement condition data, this method allows for collecting evidence of repair by considering the condition data of a group of adjacent pavement sections, rather than a single one.

Technique Parameters

In this thesis, we take advantage of observations from close proximity to estimate the probability of repair. To compute the probability that a repair action has been performed on L adjacent pavement sections, as shown in dark gray color in Figure 43, the following parameters are used in conjunction with Bayes' formula:

- Condition of the pavement sections (CI_1 to CI_L) prior to receiving repair action: Pavement sections with a lower condition rating are more likely to receive a repair action than those in better condition.
- Magnitude of increase or decrease in condition rating ($\Delta 1$ to ΔL): A 20-point increase in a 0-100 condition index is more likely to be a result of repair, and not just data error, in comparison to a rating increase of 2 points, for instance.
- Number of sections with increased and decreased condition rating (from year to year): The probability that a group of sections indeed received a repair action increases as more sections within that group have improved condition rating. In other words, if the condition ratings of many adjacent sections have improved, it is more probable that these sections indeed received an M&R; compared to another group of adjacent sections where the condition ratings of only a few sections have improved.
- Length of group of adjacent sections (L): Due to practical contract letting considerations, pavement projects tend to have a minimum length. Thus, it is more likely to have a project applied to 5 0.5-mile sections, for instance, than to have a pavement project applied to one 0.5-mile section. Note that this length

parameter is expressed here in terms of number of adjacent pavement sections (not miles).

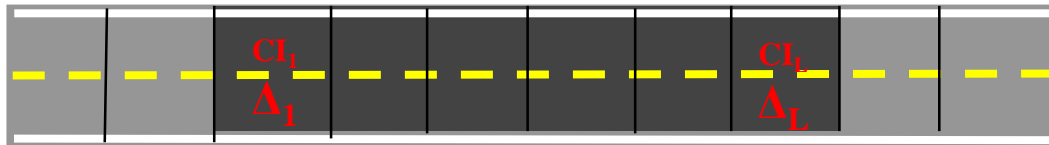


Figure 43 Data used to calculate the probability that a repair action has been performed on a group of L adjacent pavement sections

The total number of possible projects T with length L in the range of $[L_{\min}, L_{\max}]$ in a roadway corridor that consists of N pavement sections is calculated using Eq. (24). As shown in this equation, T increases linearly with N. So, even for large roadway networks, the problem remains tractable.

$$T = \sum_{L=L_{\min}}^{L=L_{\max}} (N - L + 1) = \Delta L \times N - \frac{\Delta L(\Delta L - 1)}{2} \quad (24)$$

where: $\Delta L = L_{\max} - L_{\min}$

For example, for a hypothetical roadway corridor that consists of 10 pavement sections (see Figure 44), the total number of possible projects with L=5 (i.e., project consists of 5 pavement sections) is equal to $N - L + 1 = 6$.

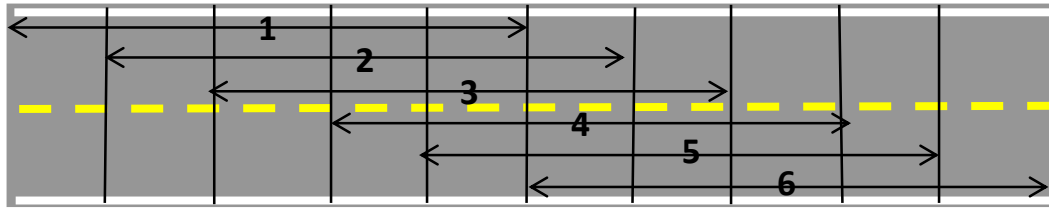


Figure 44 Six possible boundaries for a project spanning over 5 sections of a roadway that consists of 10 sections

To determine the probability that a section was repaired, it is necessary to compute the probabilities of all possible M&R projects which the section is part of. The probability of an M&R project is discussed in the next subsection of this thesis and is expressed as the probability of repair for a group of contiguous sections.

Finally, all sections are ranked in a decreasing order based on their calculated probability of repair. The highest ranked sections are selected so that the total length of selected sections equals to the average annual portion of the highway network that receives repair. It should be noted that the portion of network that receives repair and the project length range can vary from agency to agency, and thus should be determined based on the agency's past experience. For example, an agency might, on average, repair 10% of its pavement network every year and the project length might range from 2 to 5 miles.

Probability of Repair for a Group of Contiguous Sections

This subsection presents a method for computing the probability that any given contiguous group of pavement sections has received repair in any given year. By substituting the above-mentioned parameters, Bayes formula is written as follows:

$$P(R|CI_1, \dots, CI_L, \Delta_1, \dots, \Delta_L, L) = \frac{P(CI_1, \dots, CI_L, \Delta_1, \dots, \Delta_L | R) \times P(R|L)}{P(CI_1, \dots, CI_L, \Delta_1, \dots, \Delta_L)} \quad (25)$$

where:

R: The event of receiving an M&R action.

CI_i: Condition index of the *i*th pavement section in the group

Δ_i: Magnitude of increase or decrease in condition rating for the *i*th section

P(R|CI₁, ..., CI_L, Δ₁, ..., Δ_L, L): Probability that a group of L sections received an M&R action in a given year, given that the CI and Δ values for all sections in the group and the group length (L) are known.

P(CI₁, ..., CI_L, Δ₁, ..., Δ_L|R): Probability that a group of L sections have CI₁ to CI_L and Δ₁ to Δ_L values, if in fact it received an M&R action in that year and it is of length L.

P(R|L): Probability that the group of sections has indeed received an M&R action in a particular year, given that it has a length L. This probability function is constructed based on the total length of the group of pavement sections considered (i.e., without knowing the CI and Δ values of the pavement sections).

P(CI₁, ..., CI_L, Δ₁, ..., Δ_L): Probability that L adjacent pavement sections have condition index values of CI₁ to CI_L and that these values will increase or decrease by Δ₁ to Δ_L in the next year.

To apply Eq. (25) to PMS data, a number of checks and transformations are made, as described in the following paragraphs.

Transformation 1

The probability of the product of a number of statements can be expressed by Chain rule (Koch 2007). Thus, the prior probability of having a repaired group of sections with pre-repair condition index of CI_1 to CI_L and post-repair Δ of Δ_1 to Δ_L can be expanded as follows:

$$P(CI_1, \dots, CI_L, \Delta_1, \dots, \Delta_L | R) = P(\Delta_L | CI_1, \dots, CI_L, \Delta_1, \dots, \Delta_{L-1}, R) \times P(\Delta_{L-1} | CI_1, \dots, CI_L, \Delta_1, \dots, \Delta_{L-2}, R) \\ \times \dots \times P(\Delta_1 | CI_1, \dots, CI_L, R) \times P(CI_L | CI_1, \dots, CI_{L-1}, R) \times \dots \times P(CI_2 | CI_1, R) \times P(CI_1 | R) \quad (26)$$

When a pavement section receives repair action R , the increase or decrease in its CI value (i.e. Δ) depends primarily on the pre-repair CI value and repair type. A dataset from Texas showed that the CI of nearly all repaired sections changes to 100 after an M&R is applied, making $\Delta_i = 100 - CI_i$. For example, if an overlay is applied to a section with an existing CI of 75, Δ will be $100 - 75 = 25$ and if an overlay is applied to another section with CI of 90, $\Delta = 100 - 90 = 10$. Also, CI_i is independent of CI_j (see Transformation 2) and consequently Δ_i is independent of CI_j because $\Delta_i = a - CI_i$, where a is a constant that depends on the repair type. Finally, since CI_i and CI_j are independent, Δ_i and Δ_j would also independent.

It can be shown (Koch 2007) that if statements A and C_1 to C_k are independent, then the probability that “ A is true, given the conditions C_1, \dots, C_k and B are true” is equal to the probability that “ A is true, given that the condition B is true.” Mathematically:

$$P(A | C_1, \dots, C_k, B) = P(A | B) \quad (27)$$

Based on Eq. (27) and the fact that Δ_i is independent of CI_j and Δ_j , it is possible to simplify some terms of Eq. (26), as follows:

For $i=1$ to L :

$$P(\Delta_i | CI_1, \dots, CI_L, \Delta_1, \dots, \Delta_{i-1}, R) = P(\Delta_i | CI_i, R) \quad (28)$$

Transformation 2

To be able to justify this transformation, an investigation of possible spatial correlation between the condition ratings of neighboring sections is warranted. There are different methods for representing spatial dependency in geostatistical data. However, the primary tool to investigate spatial dependency is semivariogram as defined in Eq. (29) (Schabenberger and Gotway 2004).

$$\gamma(d_{ij}) = \frac{1}{2} \text{Var}[CI_i - CI_j] \quad (29)$$

where:

d_{ij} : the distance between the geometric centroid of two pavement sections i and j ; CI_i and CI_j : the CI values of pavement sections i and j , respectively; γ : semiovariogram of pavement sections i and j , which depends only on the distance between sections i and j .

Generally, if there is a spatial dependency in a variable, the semivariogram will have an increasing pattern as the distance increases, ultimately approaching its sill asymptotically. This pattern occurs in a spatially dependent variable because as the distance increases, the variance also increases.

Semivariogram plots were generated for the condition index of randomly-selected roadways of different classifications (Farm-to-Market road, State Highway, and Interstate Highway) using condition data from Texas for two years (2009 and 2010). Figure 45 and Figure 45 show semivariogram plots for State Highway 79 (SH-79) and Interstate Highway 45 (I-45), as examples (remaining plots are not shown here for brevity). Each point in these plots is computed based on the condition index of a pair of sections. The distance extends from 0.5 mile (representing the distance between the centers of two adjacent 0.5-mile sections) to 25 miles (representing the distance between the centers of two 0.5-mile sections that are 25 miles apart). Due to the large number of data points, quantiles are shown as bars instead of points. Hollow points in these figures represent outlier data.

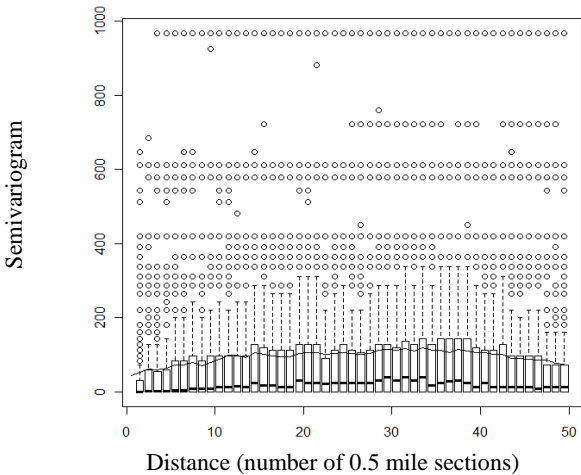


Figure 45 Condition index semivariogram for Highway SH- 79 in 2010

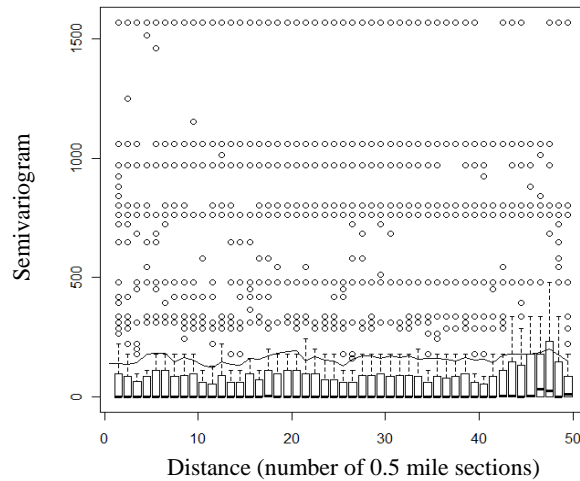


Figure 45 Condition index semivariogram for Highway I- 45 in 2010

The semivariograms shown in Figure 45 and Figure 45 are highly fluctuating and cyclic. This pattern indicates that there is no significant spatial dependency in the condition index data. (Note that the condition index in these plots is TxDOT's 0-100 DS). This finding is not surprising because of the influence of local factors (such as changes in traffic at intersections, variability in subgrade condition, variability in drainage condition, variability in construction quality, etc.) on pavement condition. It should also be mentioned that more than 70% of the sections have DS=100. This leads to similarities in the conditions of sections located far from each other. This is reflected in the cycles and fluctuations of the semivariogram. The CI in this dataset (i.e., DS) is biased toward 100 because it does not account for all distress types, severity levels, and density. For example, even if several types of distress exist in the pavement; as long as they are of low density, DS remains at or near 100. Detailed information about this condition index, and several other pavement condition indexes, is provided in Gharaibeh et al. (2010).

Since no strong spatial correlation exists between the CI of pavement sections in the network (i.e., sections are spatially independent in terms of CI values), it is possible to make the following simplifications:

For $i=1$ to L :

$$P(CI_i | CI_1, \dots, CI_{i-1}, R) = P(CI_i | R) \quad (30)$$

After replacing Eq. (26) with the simplified terms shown in Eq. (29) and (30), the final priori probability can be computed as follows:

$$P(CI_1, \dots, CI_L, \Delta_1, \dots, \Delta_L | R) = \prod_{i=1}^L P(\Delta_i | CI_i, R) \times \prod_{i=1}^L P(CI_i | R) \quad (31)$$

Similarly, the normalizing part of Eq. (25) can be simplified as follows:

$$P(CI_1, \dots, CI_L, \Delta_1, \dots, \Delta_L) = \prod_{i=1}^L P(\Delta_i | CI_i) \times \prod_{i=1}^L P(CI_i) \quad (32)$$

By applying Eq. (31) and (32) to Eq. (25), the equation for computing probability of repair for a group of contiguous sections can be expressed as follows:

$$P(R | CI_1, \dots, CI_L, \Delta_1, \dots, \Delta_L, L) = P(R | L) \times \prod_{i=1}^L \frac{P(CI_i | R)}{P(CI_i)} \times \prod_{i=1}^L \frac{P(\Delta_i | CI_i, R)}{P(\Delta_i | CI_i)} \quad (33)$$

where:

L : Number of 0.5-mile pavement sections in the group.

$P(R|L)$: Probability that a given group of adjacent pavement sections receives a repair action, given that the total length of that group of sections is L (and without knowing the sections' condition).

$P(CI_i|R)$: The percentage of repaired sections that have $CI=CI_i$

$P(CI_i)$: The percentage of all sections that have $CI=CI_i$.

$P(\Delta_i|CI_i,R)$: The percentage of repaired sections with $\Delta=\Delta_i$ that have $CI=CI_i$

Probability of Repair for any Given Section

Suppose that section X can potentially be part of n M&R projects and the probabilities of these projects (computed using Eq. (33)) are P_1 to P_n . Then, the probability that section X has been repaired in any given year ($P(X)$) can be calculated using Eq. (34).

$$P(X) = 1 - \prod_{i=1}^n (1 - P_i) \quad (34)$$

Finally, all pavement sections in the network are ranked in a decreasing order of probability of repair ($P(X)$) and repaired sections are identified such that the total length of these sections equals the annual portion of the network under repair. This input parameter (i.e., annual portion of network under repair) can be readily obtained from the agency's past experience or historical data (see the next part of this section).

Application of the Developed Technique

This section discusses the probability density functions (PDFs) required by Eq. (33) based on actual data from the Beaumont District of TxDOT. To ensure the applicability of these distributions to pavement networks in other TxDOT districts, these PDFs are used to impute M&R history data in the Bryan District. These districts were used in this study due to their similarities in size (i.e., both districts have similar lane-miles of roadway) and location (both districts are located in east-central Texas).

In Texas, pavement condition data is stored directly in the Pavement Management Information System (PMIS) database while maintenance and repair project data is stored

in separate databases. In the PMIS database, pavement sections are identified (referenced) by a unique address which is a combination of district name, county name, highway name, and beginning and end reference mile markers. On the other hand, M&R and construction databases identify projects by a code (called control-section-job or CSJ) and an approximate location on the roadway. Also, in many cases, the repair year and specific work type are missing. Additionally, some routine maintenance projects are done by in-house forces and thus are not recorded in the database at all. Thus, it is very difficult to integrate construction and repair history with condition data from these databases.

Pavement condition data is collected annually by a vendor. However, TxDOT collects “audit” condition data on about 5% of the network for validation purposes. To assess the variability of the collected condition data, standard deviation of error is calculated with the assumption that the true DS value is the average of the audit and vendor values. The DS standard deviation of error for all TxDOT roads was calculated as 5.8%, 7.5%, 8.2%, and 6.5% for years 2007 to 2010, respectively.

A random error with the standard deviation of 5 to 10 is significant in the sense that it can mask the natural year-to-year deterioration of a pavement section and even show a false improvement in condition. This rationale was used by Virginia DOT to assign a maximum acceptable level of variability in distress data: “Data variability for each data element must be smaller than the year-to-year change in that element” (Flintsch and McGhee 2009).

In a perfect dataset, DS of all sections in the network should increase upon receiving a repair action. Otherwise, it should stay constant or decrease with time. In such dataset, temporal patterns in condition data are sufficient to impute missing M&R data without considering the spatial patterns of the data. However such perfect dataset does not exist because, as discussed earlier, pavement condition data is noisy and contains errors. Figure 46 shows changes in DS of pavement sections from 2009 to 2010 in TxDOT's Bryan District.

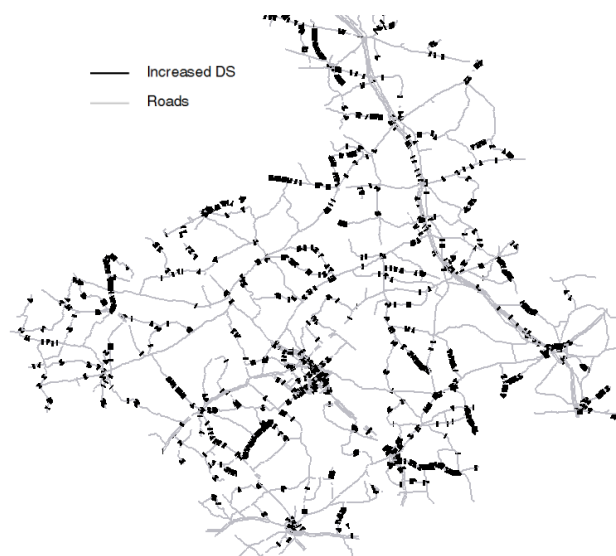


Figure 46 Sections with improved condition in Bryan District from 2009 to 2010

Dark segments in this map represent pavement sections with improved condition from 2009 to 2010. This improvement can either be due to receiving an M&R action or due to an error in the condition data. As discussed earlier, pavement sections receive M&R in

the form of projects. These projects are applied on a number of adjacent pavement sections and it is unlikely to have a project that consists of only one half-mile pavement section. For example, in fiscal year 2009 in the Bryan District, the average project length was 7.2 miles and 95% of projects had a length greater than 2 miles. Thus, one can conclude that the clusters of dark sections represent project boundaries while scattered single dark sections are likely to be erroneous condition data.

For example, Figure 47 shows the historical DS values for a pavement section located on roadway FM0060 (starting at mile marker 0636 +00.5) in Brazos County. This section hasn't received any repair in years 2007 to 2010 and thus the changes in DS values are strictly due to errors in the data. The developed method allows for dealing with this dataset so that accurate inference about M&R history can be made.

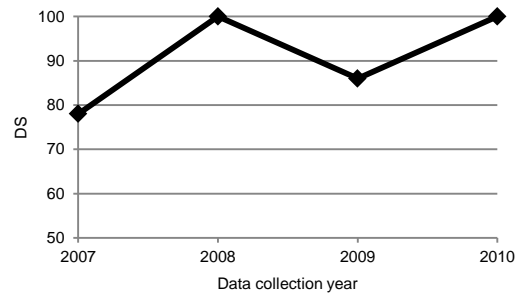


Figure 47 Noisy temporal variation in DS of an example pavement section (Brazos County, Roadway FM0060 starting at 0636 +00.5)

Determining Technique Parameters

The distribution of project lengths completed by Beaumont District of TxDOT in 2009 is shown in Figure 48. This distribution is considered as background information and represents $P(R|L)$ in Eq. (33).

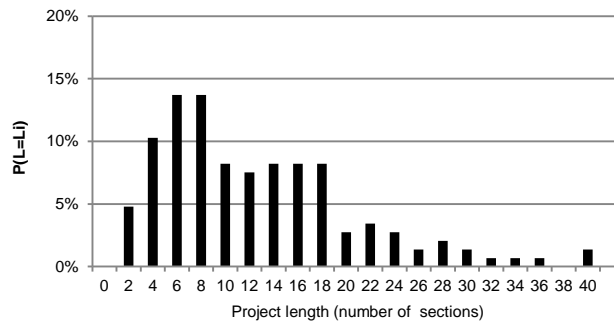


Figure 48 P(R|L): Percentage of projects with various length values

The other PDFs for the Beaumont District dataset are shown in Figure 49 to Figure 52. It should be noted that in Figure 49 and Figure 50, the percentage of sections with DS=100 are 65.4% and 44.2%, respectively. These values are much higher than the other bars and thus are noted separately in these graphs. $P(DS=100|R)=44.2\%$ means that 44.2% of the sections that have been repaired in 2010 had a DS of 100 in 2009. The percentage of sections that have a DS value of 100 in 2009 and were repaired in 2010 is 6.8% ($313/4637=6.8\%$). These percentages raise the question: why so many sections with DS=100 were repaired? Possible explanations include:

1. A pavement section with DS=100 is not necessarily in perfect condition, as DS does not account for all distress types, severity levels, and densities.
2. Due to economy-of-scale and other practical reasons, M&R actions are normally applied to a group of sections and not to isolated individual sections. Thus, sections with DS=100 may be repaired only because the neighboring sections needed to be repaired.
3. Condition data contain inherent error that cannot be eliminated completely.

It should be mentioned that the DS, by definition, is a continuous variable. However, it is normally recorded as a rounded value for practical reasons, as presented in Figure 49 to Figure 50. For example in Figure 49, the bar of DS=90 represents the PDF of a DS range of [89.5, 90.5) and at the same time it represents $P(DS=90)$.

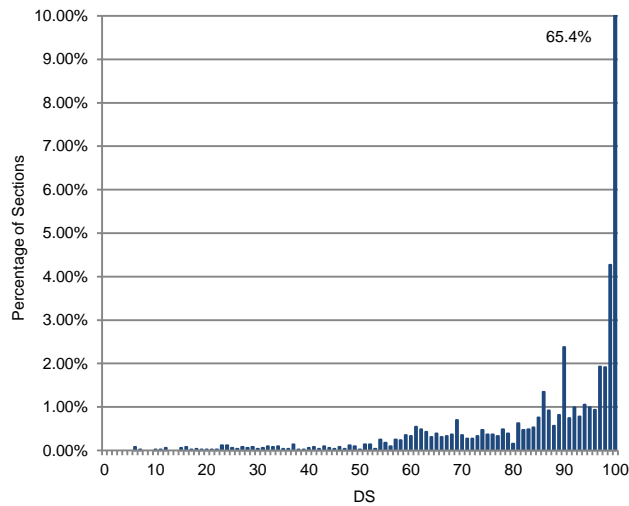


Figure 49 P(DS): Percentage of sections with each DS value

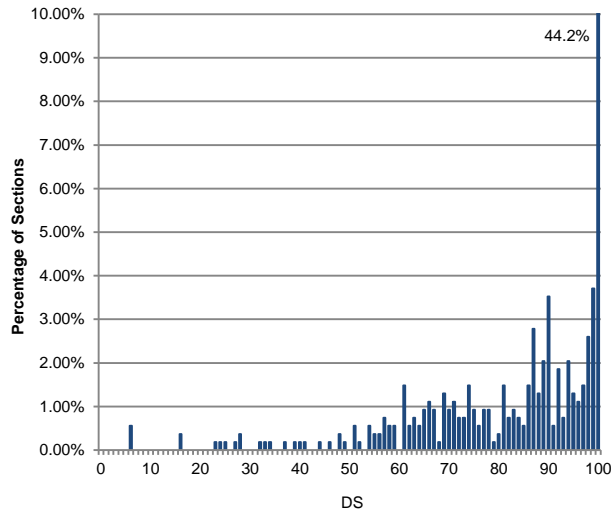


Figure 50 P(DS|R): Percentage of repaired sections with each DS value

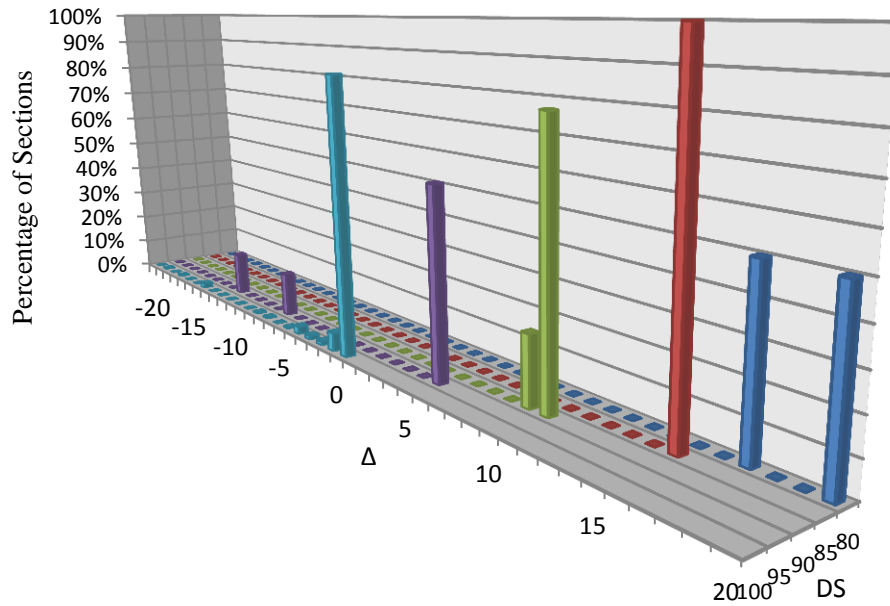


Figure 51 P(Delta|DS,R): Percentage of repaired sections with each Delta value in each DS category

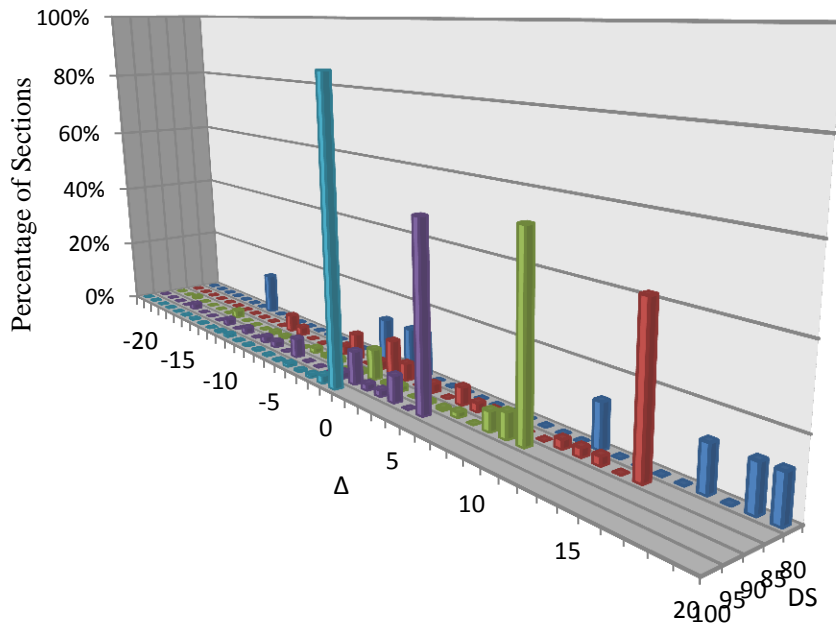


Figure 52 P(Δ |DS): Percentage of sections with each Δ value in each DS category

Table 7 shows a 3.5-mile stretch of pavement (composed of seven half-a-mile sections), as an example calculation. In this example, CI is TxDOT's DS. The DS values are obtained from the pavement management database. Δ is computed as $\Delta=100-DS$ (assuming the repair action increases DS to 100). PDF values for each section are obtained from frequency distribution graphs presented in this section (Figure 49 to Figure 52).

Table 7 Numerical example of key calculations of the developed technique

Section Number (i)	CI	Δ	P(CI)	P(CI R)	P(Δ CI)	P(Δ CI,R)
1	96	4	0.94%	1.11%	65%	100%
2	100	0	65.43%	44.16%	87%	84%
3	90	10	2.38%	3.53%	55%	79%
4	100	0	65.43%	44.16%	87%	84%
5	99	1	4.27%	3.71%	65%	70%
6	94	6	1.05%	2.04%	57%	64%
7	98	2	1.91%	2.60%	64%	86%

The length of this group is 3.5 miles and according to the distribution shown in Figure 48, $P(R|L)=14.38\%$. By applying the parameters shown in Table 7 to Eq. (33), the probability that this group of pavement sections has received an M&R action between the two consecutive data collection years is calculated to be 88%.

Results and Discussion

Using Eq. (33) and (34) in conjunction with the PDFs plotted in the previous section, the probability that an M&R action has been applied is calculated for each section in the network. Then the sections are sorted based on this probability. These calculations are implemented in ArcGIS using the Python coding language (code of computer program is presented in Appendix C). Finally, the sections with the highest probability of being repaired are picked so that the sum of their lengths is equal to a predefined portion of the total length of the highway network. In this analysis, a 10% portion is used because, on

average, TxDOT repairs 10% of the network annually. This input parameter (i.e., network repaired portion, hereafter abbreviated as NRP) should be determined based on the agency's past experience or historical data.

Each section classified by the developed method falls in one of four classes: True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). Thus, the precision and accuracy of this binary classification method can be defined as follows:

$$\text{Precision} = \frac{TP}{TP + FP} \quad (35)$$

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (36)$$

where:

True Positive (TP): The number of sections that have been correctly classified as repaired.

True Negative (TN): The number of sections that have been correctly classified as not repaired.

False Positive (FP): The number of sections that have been incorrectly classified as repaired.

False Negative (FN): The number of sections that have been incorrectly classified as not repaired.

Precision represents the effectiveness of the method in correctly detecting repaired sections. Accuracy, on the other hand, is the proportion of true results (both true positive and true negative), in the whole population of pavement sections in network. It represents the veracity of the developed method.

To measure effectiveness of the developed method, the Beaumont District PDFs are used to estimate repair projects in a different district (the Bryan District). The Bryan District highway network consists of 7090 sections (3600 miles). 74% of the sections that were classified by the developed spatial-Bayesian technique as repaired were actually repaired in 2010; indicating that precision of this technique has 74% precision. The accuracy in this case was 95%, indicating that 95% of all sections in the network were classified correctly. Predicted versus actual repaired sections are shown in Figure 53 (thick gray segments represent actual M&R projects and dark segments represent M&R projects estimated by the developed technique).

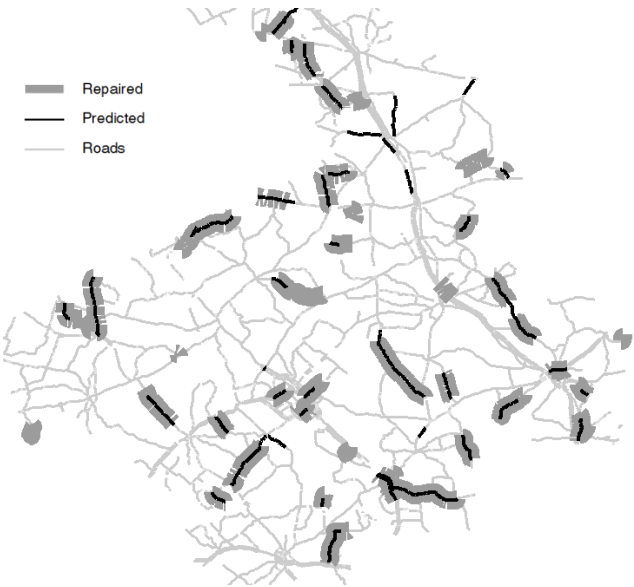


Figure 53 Estimated M&R projects (dark segments) versus actual M&R projects (thick gray segments) in Bryan District

The sensitivity of the developed method to the NRP input parameter is shown in Figure 54. Note that NRP was varied between 1% and 20% since it is unlikely that a highway agency will repair more than 20% or less than one percent of its network in any given year. This figure shows that accuracy ranges between 90% and 95%, with a maximum accuracy achieved at an optimum NRP of 10%. This optimum value of NRP occurs because as NRP increases, the number of TPs increases, but at the same time, the number of FPs also increases (as more TNs change to FPs). However, the number of FPs increases at a faster rate compared to the number of TPs; resulting in a decreased precision with increased NRP.

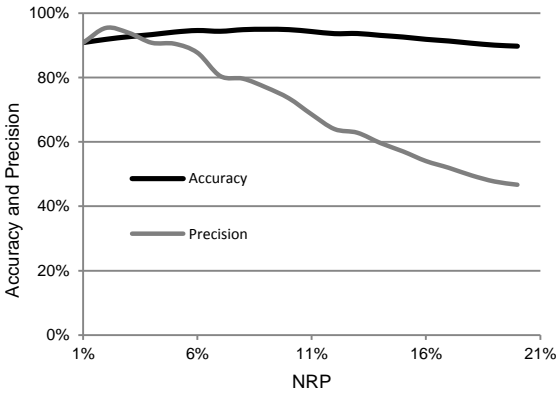


Figure 54 Sensitivity of the accuracy and precision of the developed Method to the NRP input parameter

Limitations of the developed method and possible causes of error in the results include:

1. Presence of systematic error in the CI data: More than 60% of falsely classified (FP and FN) sections are those with deteriorated condition (i.e. lower CI value) after they have been repaired, or those with increased CI values without receiving M&R. This indicates that the CI data for these sections is likely to be erroneous.
2. Inability of the CI to reflect the true condition of the pavement: For example, TxDOT's DS does not account for pavement skid resistance and some individual distress types. In such situations, the agency may make repair decisions based on condition indicators that are not accounted for in the CI.
3. Presence of small projects: The calculated probability of repair for short projects can be highly sensitive to error in the CI data.
4. Non-representative or inaccurate prior information: Low quality prior information can lead to inaccurate prior probability distribution functions and consequently reduce the accuracy and precision of the developed method.
5. The derivation of Eq. (34) implicitly assumes that the various project probabilities are independent of each other. This simplifying assumption can be violated since M&R projects can share common pavement sections. To account for this dependency, further research would be needed in modeling the covariance structure of M&R projects.

The accuracy and precision of the developed method can be improved by addressing the above limitations and possible sources of error. However, reducing systematic error in the CI data has the greatest impact on the effectiveness of the developed method.

Figure 55 shows estimated projects using the section-by-section method. In this method, the estimation of repair history is done solely based on each section's condition history from 2009 to 2010 (i.e., ignoring condition of neighboring sections). In this way, 10% of the sections with the highest DS increase are estimated to have received an M&R action. This results in a prediction precision of 47%; which is much lower than the 74% precision of the developed spatial-Bayesian technique.

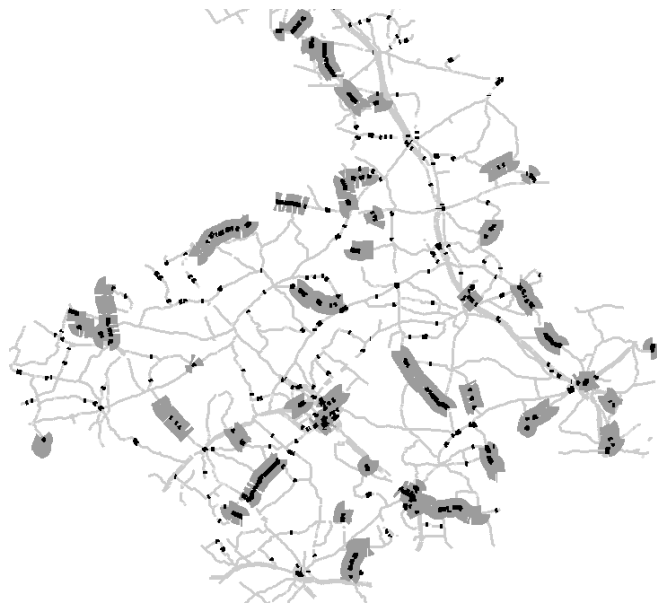


Figure 55 M&R projects estimated solely based on improvement in DS on a section-by-section basis (dark segments) versus actual M&R projects (thick gray segments) in Bryan District

To quantify the effect of considering spatial patterns on the robustness of this data imputation method, an artificially error-free pavement condition dataset was manufactured. Then, a normally-distributed random noise in the DS with average of zero

and standard deviation of σ was introduced to the dataset. The precision of the developed spatial-Bayesian technique and the section-by-section method (which ignores the spatial patterns in condition and repair data) for estimating past M&R projects were computed and plotted in Figure 56.

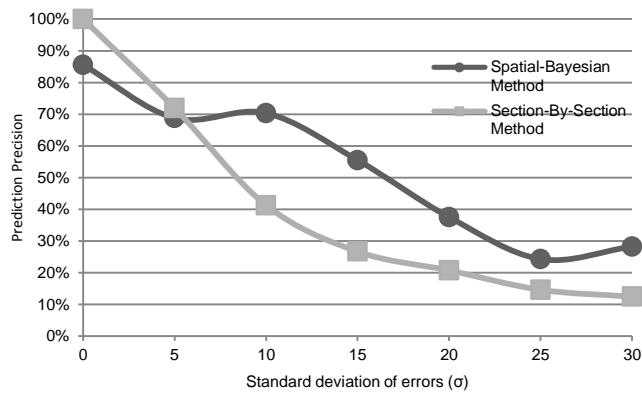


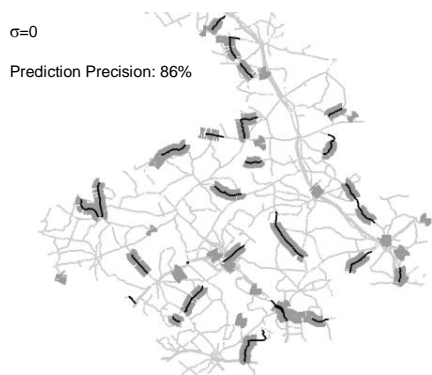
Figure 56 Estimating the precision of spatial-Bayesian and section-by-section methods versus standard deviation of error in pavement condition

Figure 56 shows how the precision of both methods decreases as the standard deviation of error in the pavement condition dataset increases. The non-monotonic trend of the spatial-Bayesian method in this plot, as opposed to the monotonic trend of the section-by-section method, can be attributed to the PDFs and the interactions among this prior information. Also, the spatial-Bayesian method is not 100% precise even if no error is present in the condition data. This is because of the effect of prior information about project length (i.e., project length PDF). For example, an isolated single 0.5-mile section might have been repaired and its DS increased (and thus should be classified as an M&R

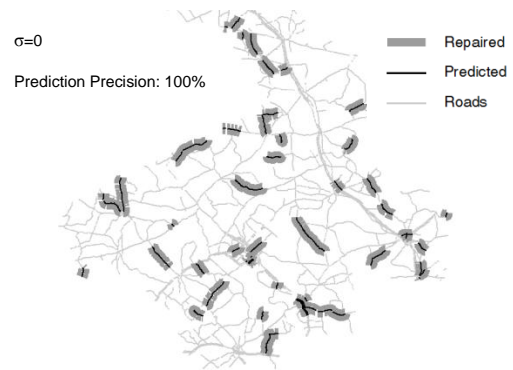
project in itself), but the spatial-Bayesian method considers this information unlikely because the project length prior PDF shows that 0.5-mile M&R projects are very unlikely. Similar situation can occur for very long projects (also considered to be unlikely based on project length PDF).

The precision of the spatial-Bayesian method remains stable (ranging between 70% and 85%) up to a DS standard deviation of error of 10 DS points. This behavior suggests that a 10% error in the condition data is a critical point where the method is no longer stable. As discussed earlier, σ usually ranges between 5 and 10 DS points. On the other hand, the precision of the section-by-section method decreases steadily as the error level in DS increases.

Predicted versus actual repaired sections for $\sigma=0, 5, 10,$ and 15 using the spatial-Bayesian method and the section-by-section method are shown in Figure 57 through Figure 60. As shown in these figures, for $\sigma>10$, the section-by-section method approaches a random estimation of M&R projects (as can be seen in the dotted map). However, for spatial-Bayesian method the results remain fairly consistent. These comparisons indicate that the developed technique is fairly robust.



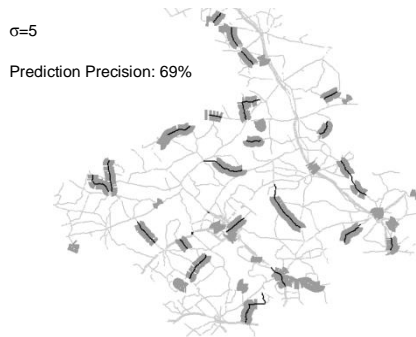
(a)



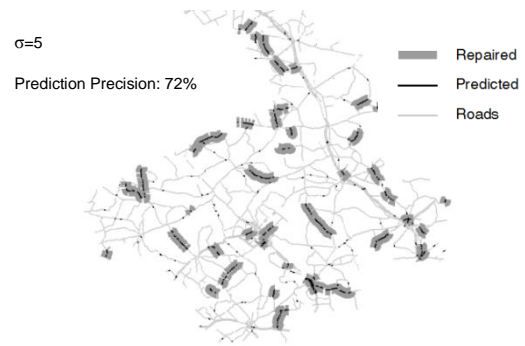
(b)

Figure 57 Predicted versus actual repaired sections for $\sigma=0$: (a) using the spatial-Bayesian method

(b) based on DS increases



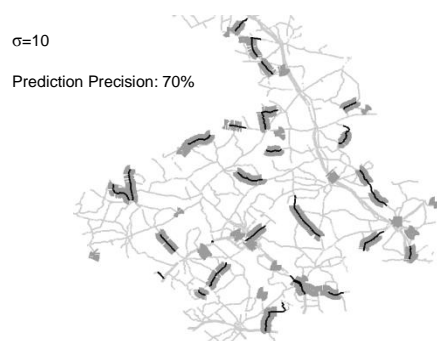
(a)



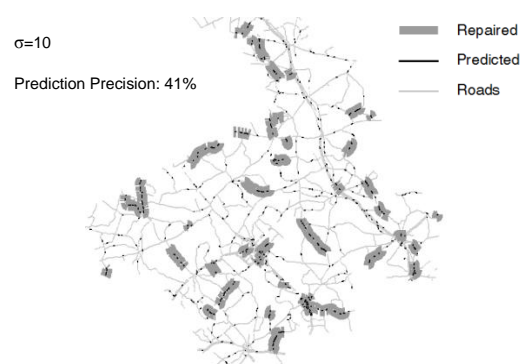
(b)

Figure 58 Predicted versus actual repaired sections for $\sigma=5$: (a) using the spatial-Bayesian method

(b) based on DS increases



(a)



(b)

Figure 59 Predicted versus actual repaired sections for $\sigma=10$: (a) using the spatial-Bayesian method

(b) based on DS increases

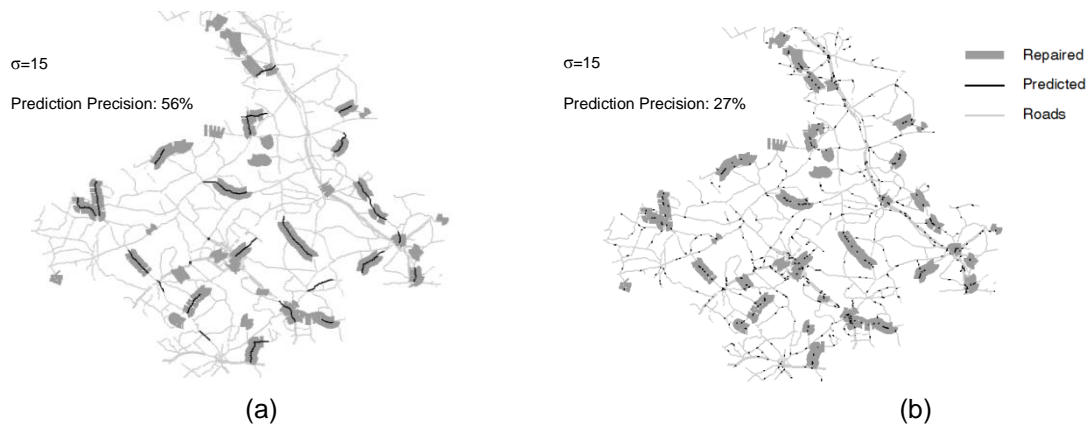


Figure 60 Predicted versus actual repaired sections for $\sigma=15$: (a) using the spatial-Bayesian method (b) based on DS increases

Section Summary and Concluding Remarks

Pavement M&R data and condition data are often stored in disparate heterogeneous databases that are difficult to integrate (especially legacy databases). A GIS-based Bayesian method has been developed for imputing construction and M&R history of a pavement network by recognizing spatial and temporal patterns in pavement condition data. This thesis shows that combined spatial and temporal patterns in condition data can be used for estimating the construction and M&R history of pavement networks more accurately.

Analysis of error in pavement condition rating data from the Bryan district case study showed a standard deviation of error ranging from 5 to 10 points (for a 0-100 condition index). For this range of error in condition data, the developed technique has 74% precision in estimating repair data while the estimation of repair history solely based on each section's condition history (i.e., ignoring spatial data) has only 47% precision.

7. SUMMARY, CONTRIBUTION, CONCLUSIONS, AND RECOMMENDATIONS

Infrastructure management is a data-driven process. It relies on large databases that contain information on the system's inventory, condition, M&R treatment history, utilization, and cost. This data feeds into analytical models that assess infrastructure current conditions, predict future conditions, and develop optimal M&R strategies. Thus, complete and accurate data is essential to a reliable infrastructure management system. In the infrastructure management arena, much of past research has focused on data collection processes and technologies, as well as data analysis (performance modeling, optimization, etc.); little work has been done to help understand the impact of data quality on the outputs of infrastructure management systems and to develop efficient methods for assessing and improving data quality. This research fills this gap in the literature, with focus on pavement management systems.

The aim of this research is to assess and enhance the quality of infrastructure data through the development of novel computational techniques. First, this research provided a quantitative assessment of the effects of error magnitude and type (both systematic and random) in pavement condition data on the accuracy of PMS outputs (i.e., forecasted needed budget and M&R activities in a multi-year planning period). Second, a new technique for detecting gross and pseudo outliers in pavement condition data was developed and tested. The developed technique integrates concepts and methods from GIS, Bayesian statistics and Genetic Algorithms to cluster homogenous neighboring pavement sections and then identify gross outliers (data values that are likely to be erroneous), as well as pseudo outliers (pavement sections affected by

isolated local conditions). Third, a new technique for estimating construction and M&R history from pavement condition data was developed and tested. The developed technique employs Bayesian and spatial statistics in a GIS environment to search for evidence of repair in the condition data (current and past) of group of adjacent pavement sections. This technique is especially beneficial when M&R data and condition data are stored in disparate heterogeneous databases that are difficult to integrate (i.e., legacy databases).

Contribution

This study contributes to advancing the infrastructure management paradigm in two primary ways: (a) it provides in-depth understanding of the impact of errors in condition data on the outputs of infrastructure management systems, and (b) it provides efficient computational methods for improving infrastructure data quality.

The main merit of the developed techniques is their ability to integrate methods and principles from Bayesian and spatial statistics, GIS, and operations research in an efficient manner. The application of these techniques to a real-world cases study (pavement network in Bryan district) demonstrated the potential benefits of these techniques to infrastructure managers and engineers.

Conclusions

Key conclusions of this research are grouped into three categories and summarized as follows:

- Conclusions related to assessing the effect of error in pavement condition data on PMS outputs:

- Forecasted average annual budget needed for M&R is highly sensitive to both systematic and random errors in pavement condition data. This output of PMS can be distorted significantly by random errors in pavement condition data usually considered “acceptable” or inevitable.
- Predicted portions of the network in need of rehabilitation, maintenance, and “do nothing” are less sensitive to errors in pavement condition data, compared to forecasted needed budgets. For example, a ± 10 standard error in DS can result in an overestimation of the rehabilitation portion by as much as 2%, an overestimation of the maintenance portion by as much as 3.8%, and an underestimation of the "do nothing" portion by 5.8%, with a 95% level of confidence.
- The sensitivity of a PMS needs analysis at the section level (i.e., M&R type assigned to individual sections) to random errors in pavement condition data is higher than that at the aggregated level (i.e., portions of the network assigned different M&R types).
- PMS outputs are more sensitive to negative errors in condition data (i.e., an overestimation of distress and an underestimation of the condition index) than to positive errors (an underestimation of the distress and an overestimation of the condition index).
- Systematic errors in pavement condition data have a more significant impact on PMS outputs than random errors because they affect the entire

network in the same manner (i.e., either increase or decrease the condition of the whole network).

- The effects of pavement condition data errors on PMS needs analysis outputs persist throughout the planning period.
- Conclusions related to detecting pseudo and gross outliers in pavement condition data:
 - Classic clustering methods (i.e., ones that do not account for local spatial and temporal patterns in condition data) are dominant in the infrastructure management literature. Pavement groups identified by these methods can be too broad and distributed over a wide geographic area; leading to nonhomogeneous pavement sections within such clusters.
 - Gross outliers (condition data values that are likely to be erroneous), as well as pseudo outliers (pavement sections affected by isolated local conditions that caused unusual pavement performance) can be identified and delineated by analyzing spatial and temporal patterns in condition data.
 - The developed proximity-based technique is less susceptible to the masking and swamping effects, compared to classic data filtering methods prevalent in pavement engineering.
- Conclusions related to estimation of construction and M&R history from condition data:

- Construction and M&R history of pavement networks can be estimated fairly reliably by analyzing spatial and temporal patterns in condition data.
- The devised technique was found to be robust in the presence of random error in pavement condition data. For condition data with a standard deviation of error ranging from 5 to 10 points (for a 0-100 condition index), the developed technique has 74% precision in estimating repair data.
- Possible sources of classification error in the developed method include presence of systematic error in condition data, inability of the condition index to reflect the true condition of the pavement, presence of small projects, and use of non-representative or inaccurate prior information. The accuracy and precision of the developed method can be improved by addressing these possible sources of error.

Recommendations

Key recommendations based on the results of this research are to:

- Optimize pavement condition data collection processes by focusing on error types and levels that have the greatest impact on PMS outputs.
- Base pavement condition data quality acceptance criteria on analyses of the impact of data quality on PMS outputs, rather than strictly precision analyses and practical considerations.

- Employ the developed outlier detection technique to improve pavement performance modeling. Gross outliers as likely erroneous values should be further investigated before being included in performance modeling. Pseudo outliers should be modeled separately because they behave differently. Markov models could be investigated as a possible performance prediction approach that allows for considering the differences between normal pavement sections and pseudo outlier sections.
- Apply the developed outlier detection technique to identify candidate sections for field auditing
- Detected outliers can potentially be used to assess the efficiency of infrastructure condition inspection training programs by checking if the inspectors are able to identify pseudo and gross outliers.
- Compare pavement sections with normal data to those identified as pseudo outliers (i.e., pavements that perform exceptionally well or poor) to gain insights regarding the effectiveness of various designs, construction methods, and materials
- Investigate improving the run time of the computer code of the developed clustering technique.

Finally, it is recommended to extend the developed techniques to other linear infrastructure systems (e.g. roadways, pipelines, power transmission lines).

REFERENCES

- AASHTO (1993). *Guide for Design of Pavement Structures*, American Association of State Highway and Transportation Officials, Washington, D. C.
- AASHTO (2001). *Pavement Management Guide*, American Association of State Highway and Transportation Officials, Washington, D.C.
- AASHTO (2002). *Guide for Mechanistic-Empirical Design of New and Rehabilitated Pavement Structures*, American Association of State Highway and Transportation Officials, Washington, D.C.
- Adachi, T., and Ellingwood, B. R. (2009). "Serviceability assessment of a municipal water system under spatially correlated seismic intensities." *Computer-Aided Civil and Infrastructure Engineering*, 24(4), 237-248.
- Adeli, H. (2001). "Neural networks in civil engineering: 1989–2000." *Computer-Aided Civil and Infrastructure Engineering*, 16(2), 126.
- Al Qudrah, A. A., and Alsugair (1998). "Artificial neural network approach for pavement maintenance." *Journal of Computing in Civil Engineering*, 12(4), 249.
- Ashlock, D. (2005). *Evolutionary Computation for Modeling and Optimization*, Springer, New York.
- Ballou, D. P., Pazer, H. L., Belardo, S., and Klein, B. (1987). "Implication of data quality for spreadsheet analysis." *Data Base*, 18(3), 13-19.
- Bansal, A., Kauffman, R. J., and Weitz, R. R. (1993). "Comparing the modeling performance of regression and neural networks as data quality varies: a business value approach." *Journal of Management Information Systems*, 10(1), 11-32.

- Bartholomew, D. J. (1975). "Errors of prediction for Markov chain models." *Journal of the Royal Statistical Society*. 3(1), 444.
- Ben-Gal, I. (2005). *Data Mining and Knowledge Discovery Handbook: A Complete Guide for Practitioners and Researchers*, Kluwer Academic Publishers, New York, 1-12.
- Boardman, A. E., Greenberg, D. H., Vining, A. R., and Weimer, D. L. (2005). *Cost Benefit Analysis: Concepts and Practice*, Upper Saddle River, New Jersey.
- Bogus, S. M., Migliaccio, G. C., and Cordova, A. A. (2010a). "Assessment of data quality for evaluations of manual pavement distress." *Transportation Research Record: Journal of the Transportation Research Board*, 21(70), 1-8.
- Bogus, S. M., Song, J., Waggerman, R., and Lenke, L. R. (2010b). "Rank correlation method for evaluating manual pavement distress data variability." *Journal of Infrastructure Systems*, 16(1), 66-72.
- Bolstad, W. M. (2007). *Introduction to Bayesian Statistics*, John Wiley & Sons, New York. 21-23.
- Brassel, K., Bucher, F., Stephan, E. M., and Vickovski, A. (1995). *Elements of Spatial Data Quality*, Elsevier, New York , 34-36.
- Buchheit, R. B., Garrett Jr, J. H., McNeil, S., and Chen, P. (2005). "Automated procedure to assess civil infrastructure data quality: Method and validation." *Journal of Infrastructure Systems*, 11(3), 180-189.
- Button, J. W., Freeman, T. J., Fry, G. T., Jones, S., Smith, R. E., and Estakhri, C. K. (2001). "Paired measurements strategy to analyze the effects of construction

- specification changes on quality of HMA surface courses." Texas Transportation Institute, College Station, Texas.
- Byrne, M., Ricardo, I., and Tony, P. (2011). "A generalised approach to outlier identification in pavement condition data." *International Journal of Pavement Engineering* 12(6), 1-11.
- Cao, X. R. (2000). "A unified approach to Markov decision problems and performance sensitivity analysis." *Automatica*, 36(5), 771.
- Cao, X. R., and Xi-Ren, C. (1998). "Algorithms for sensitivity analysis of Markov systems through potentials and perturbation realization." *IEEE Transactions on Control Systems Technology*, 6(4), 482.
- Castillo, E. (2009). "Diagnostics for non-linear regression." *Journal of statistical computation and simulation*, 79(9), 1109.
- Chan, W. T., and Fwa (1993). "Priority rating of highway maintenance needs by neural networks." *Journal of Transportation Engineering*, 119(3), 419.
- Chen, W. (2005). "Analytical variance-based global sensitivity analysis in simulation-based design under uncertainty." *Journal of mechanical design*, 127(5), 875.
- Cheneviere, P. (2006). "Cost benefit analysis aspects related to long-life pavements." *The International journal of pavement engineering*, 7(2), 145.
- Cheng, Z., Amin, H., Zayed, T. M., and Wainer, G. (2005). "Representation and analysis of spatial resources in construction simulation." *Proc., Proceedings of the 2005 Winter Simulation Conference, 4-7 Dec. 2005*, IEEE, 8.

- Cho, H., and Olivera, F. (2009). "Effect of the spatial variability of land use, soil type, and precipitation on streamflows in small watersheds." *Journal of the American Water Resources Association*, 45(3), 673-686.
- Coleman, M. D., and Gardiner, T. D. (2009). "Sensitivity of model-based quantitative FTIR to instrumental and spectroscopic database error sources." *Vibrational Spectroscopy*, 51(2), 177-183.
- Dai, L. (1997). "Rate of convergence for derivative estimation of discrete-time Markov chains via finite-difference approximation with common random numbers." *SIAM journal on applied mathematics*, 57(2), 731-751.
- Daleiden, J. F., and Simpson, A. L. (1998). "Off-the-wall pavement distress variability study." *Transportation Research Record: Journal of the Transportation Research Board*, 16(43), 62-70.
- De la Garza, J. M., Drew, D. R., and Chasey, A. D. (1998). "Simulating highway infrastructure management policies." *Journal of Management in Engineering*, 14(5), 64-72.
- De Maesschalck, R., Jouan-Rimbaud, D., and Massart, D. L. (2000). "The Mahalanobis distance." *Chemometrics and Intelligent Laboratory Systems*, 50(1), 1-18.
- De Oliveira, D. P., Neill, D. B., Garrett, J. H., Jr., and Soibelman, L. (2011). "Detection of patterns in water distribution pipe breakage using spatial scan statistics for point events in a physical network." *Journal of Computing in Civil Engineering*, 25(1), 21-30.

- Deshpande, V. P., Damnjanovic, I. D., and Gardoni, P. (2010). "Reliability-based optimization models for scheduling pavement rehabilitation." *Computer-Aided Civil and Infrastructure Engineering*, 25(4), 227-237.
- Dewan, S. A., and Smith, R. E. (2003). "Modifying local agency pavement management system to support governmental accounting standards board 34 requirements." *Proc., Eight International Conference on Low-Volume Roads 2003, June 22, 2003 - June 25, 2003*, National Research Council, 15-23.
- Edwards, D., and Hamson, M. (1990). *Guide to Mathematical Modelling*, CRC Press, Boca Raton, Fl.
- Everitt, B. S., Landau, S., Morven, L., and Stahl, D. (2011). *Cluster Analysis*, John Wiley and Sons, London.
- FHWA (2005). *Highway Economic Requirements System-State Version*, US Department of Transportation, Federal Highway Administration, Washington, D.C.
- FHWA (2010). *Data Integration Primer*, Federal Highway Administration, Washington D.C.
- FHWA (2012). "Funding Tables", Federal Highway Administration, <http://www.fhwa.dot.gov/safetealu/fundtables.htm>, Accessed 2/21/2012.
- Flintsch, G., and McGhee, K. K. (2009). *Quality Management of Pavement Condition Data Collection.*, National Cooperative Highway Research Program, Washington, D. C.

- Flintsch, G. W. (1996). "Artificial neural network for selecting pavement rehabilitation projects." *Transportation Research Record: Journal of the Transportation Research Board*, 1524(1), 185.
- Flintsch, G. W. (2004a). "Application of engineering economic analysis tools for pavement management." *Transportation research. Part A, Policy and practice*, 79(1), 1.
- Flintsch, G. W. (2004b). "Soft computing applications in infrastructure management." *Journal of Infrastructure Systems*, 10(4), 157.
- Fu, M. C. (1992). "Extensions and generalizations of smoothed perturbation analysis in a generalized semi-Markov process framework." *IEEE Transactions on Automatic Control*, 37(10), 1483.
- Fuller, W. A. (2006). *Measurement Error Models*, Wiley, New York.
- Gharaibeh, N. G., Freeman, T., Saliminejad, S., Chang-Albitres, C., Weissman, J., Weissman, A., Wimsatt, A., and Gurganus, C. (2012). "Evaluation and development of pavement scores, performance models and needs Estimates for the TxDOT Pavement Management Information System." Texas Department of Transportation, Austin, Tx.
- Gharaibeh, N. G., Zou, Y., and Saliminejad, S. (2010). "Assessing the agreement among pavement condition indexes." *Journal of Transportation Engineering*, 136(8), 765-772.

- Golroo, A., and Tighe, S. (2009). "Use of soft computing applications to model pervious concrete pavement condition in cold climates." *Journal of Transportation Engineering*, 135(11), 791-800.
- Grabe, M. (2010). *Generalized Gaussian Error Calculus*, Springer, New York, NY.
- Griffith, C., and Chamman, D. (2000). *Pavement Surface Distress Data Collection for Pavement Management Information System*, Texas Department of Transportation, Austin, Tx.
- Grivas, D. A., and Shen, Y.-C. (1995). "Use of fuzzy relations to manage decisions in preserving civil infrastructure." *Transportation Research Record: Journal of the Transportation Research Board*, 1497(1), 10-18.
- Gurganus, C., and Gharaibeh, N. G. (2012). "Pavement preservation project selection and prioritization: A competitive approach." *91st Annual Meeting of the Transportation Research Board*, Transportation Research Board, Washington, D. C.
- Hadley, D. (2011). *Sensitivity analysis within cost benefit analysis*, Science Policy Integration for Coastal Systems Assessment, Norwich, England.
- Haider, S. W., Chatti, K., Buch, N., Lyles, R. W., Pulipaka, A. S., and Gilliland, D. (2007). "Statistical analysis of in-service pavement performance data for LTPP SPS-1 and SPS-2 experiments." *Journal of Transportation Engineering*, 133(6), 378-388.

- Hajek, J. J., and Hurdal, B. (1993). "Comparison of rule-based and neural network solutions for a structured selection problem.", *Transportation Research Board* , 13(39), 1-7.
- Hawkins, D. M. (1980). *Identification of Outliers*, Springer, New York, 4-6.
- Heintz, J., Pollin, R., and Garrett-Peltier, H. (2009). *How Infrastructure Investments Support the U.S. Economy: Employment, Productivity and Growth*, Institute of Political Economy Research, Amherst, MA.
- Hodge, V., and Austin, J. (2004). "A survey of outlier detection methodologies." *Artificial Intelligence Review*, 22(2), 85-126.
- Hoem, J. M. (1973). *Levels of Error in Population Forecasts.*, Statistisk Sentralbya, Oslo, Norway.
- Hu, X., Nonomura, Y., and Kohno, M. (2011). *Monte Carlo Simulation*, Springer Handbook of Metrology and Testing, Springer Berlin Heidelberg, Berlin, Germany.
- Hudson, R. W., uddin, W., and Haas, R. (1997). *Infrastructure Management: Integrating Design, Construction, Maintenance, Rehabilitation and Renovation*, McGraw-Hill, New York.
- Humplick, F. (1992). "Highway pavement distress evaluation. Modeling measurement error." *Transportation Research Record: Journal of the Transportation Research Board*, 26(2), 135-135.

- Hutcheson, R. S. (2009). "Function-based design tools for analyzing the behavior and sensitivity of complex systems during conceptual design." PhD, Texas A&M University, College Station, Tx.
- Jia, J.-H., and Wang, C. (2010). "Spatial analysis based on GIS in the application of construction of urban underground rail." *Proc., 2nd International Conference on Information Science and Engineering, ICISE2010, December 4, 2010 - December 6, 2010*, IEEE Computer Society, 4030-4032.
- Jie, G., and Caldas, C. H. (2008). "Data processing for real-time construction site spatial modeling." *Automation in Construction*, 17(5), 526-535.
- Jorma, L., Martti, J., and Erna, K. (2000). "Informal identification of outliers in medical data." *14th European conference on Artificial Intelligence*, Berlin, Germany.
- Kim, S., Damnjanovic, I., and Gunby, M. (2010). "Effects of pavement spatial variability on contractor's management strategies." *Journal of Infrastructure Systems*, 16(4), 231-240.
- Kingston, G. (2001). "Cost benefit analysis in theory and practice." *Australian Economic Review*, 34(4), 478.
- Klein, B. D., and Rossin, D. F. (1999). "Data quality in neural network models: effect of error rate and magnitude of error on predictive accuracy." *Omega*, 27(5), 569-582.
- Knorr, E. M., and Ng, R. T. (1998). "Algorithms for mining distance-based outliers in large datasets." *Proceedings of the 24rd International Conference on Very Large Data Bases*, Morgan Kaufmann Publishers Inc., 392-403.

- Koch, K., R., (2007). *Introduction to Bayesian Statistics*, Springer, Bonn, Germany.
- Lajnef, N., Rhimi, M., Chatti, K., Mhamdi, L., and Faridazar, F. (2011). "Toward an integrated smart sensing system and data interpretation techniques for pavement fatigue monitoring." *Computer-Aided Civil and Infrastructure Engineering*, 26(7), 513-523.
- Landers, S., Robson, M., and Cowe Falls, L. (2002). "Development of quality assurance and control procedures for network level contract pavement surface condition surveys." *Proc., 5th International Conference on Managing Pavements*, Seattle, Washington.
- Larson, C. D., Sami, N., and Luhr, D. R. (2000). "Structured approach to managing quality of pavement distress data: Virginia department of transportation experience." *Transportation Research Record: Journal of the Transportation Research Board*, 16(99), 72-80.
- Laudon, K. C. (1986). "Data quality and due process in large interorganizational record systems." *Communications of the ACM*, 29(1), 4.
- Lee, S., and Adams, T. M. (2004). "Spatial model for path planning of multiple mobile construction robots." *Computer-Aided Civil and Infrastructure Engineering*, 19(4), 231-245.
- Lee, Y.-H., and Darter, M. I. (1994). *Development of Pavement Prediction Models*, University of Illinois at Urbana-Champaign, Urbana-Champaign, Il.
- Lee, Y.-H., Mohseni, A., and Darter, M. I. (1993). "Simplified pavement performance models." *Transportation Research Record, No. 1397*, 7-14.

- Li, L., Zhu, L., and Sui, D. Z. (2007). "A GIS-based Bayesian approach for analyzing spatial-temporal patterns of intra-city motor vehicle crashes." *Journal of Transport Geography*, 15(4), 274-285.
- Li, N. (1996). "Reliability-based processing of Markov chains for modeling pavement network deterioration." *Transportation Research Record: Journal of the Transportation Research Board*, 1524(1), 203.
- Liang, J., F, (2011). "Application of fuzzy AHP to the comparison and selection of cement concrete pavement overlaying schemes." *Advanced Materials Research*, 243-249.
- Livneh, M. " (2004). Repeatability and reproducibility of manual pavement distress survey methods " *Proc., 3rd International Conference on Managing Pavements*. Seattle, Washington.
- Lou, Z. (2001). "Application of neural network model to forecast short-term pavement crack condition: Florida case study." *Journal of Infrastructure Systems*, 7(4), 166.
- Lytton, R. L. (1987). "Concepts of pavement performance prediction and modeling." *Proc., 2nd North American Conference on Managing Pavements*, Ministry of Communication and Transportation, Ontario, Canada.
- Lytton, R. L., Aubeny, C., and Bulut, R. (2005). *Design Procedure for Pavements on Expansive Solis*, Texas Department of Transportation, Austin, Tx.
- Madanat, S. (1993). "Incorporating inspection decisions in pavement management." *Transportation Research, Part B (Methodological)*, 27B, 425-438.

- Manzella, A., and McNeil, S. (2006). *Complex Systems Models as Tools for Asset Management*, Department of Civil and Environmental Engineering of University of Delaware, Delaware.
- Mao, M., and Yu, X. (2009). "A method of expressway pavement performance prediction." *Proc., 2nd International Conference on Transportation Engineering, ICTE 2009, July 25, 2009 - July 27, 2009*, American Society of Civil Engineers, Washington, D.C., 433-438.
- Matejicek, L. (2011). "Spatio-temporal analysis of environmental pollution in urban areas: A case study of the environment in the city of Prague." *Proc., 19th International Congress on Modelling and Simulation - Sustaining Our Future: Understanding and Living with Uncertainty*, Modelling and Simulation Society of, 1909-1915.
- McNeil, S., and Humplick, F. (1991). "Evaluation of errors in automated pavement-distress data acquisition." *Journal of Transportation Engineering*, 117(2), 224-241.
- McNinch, T., J., Dong, J. K., and Colling, T. (2008). "Validation of the pavement performance models used In Michigan's RoadSoft infrastructure management system & development of parent deterioration curves for local agencies." *Mid-Continent Transportation Research Forum Program* Midwest Regional University Transportation Center, University of Wisconsin – Madison.

- McQueen, J. M., and Timm, D. H. (2005). "Statistical analysis of automated versus manual pavement condition surveys." *Transportation Research Record: Journal of the Transportation Research Board*, 1940(1), 55-62.
- Migliaccio, G. C., Bogus, S. M., and Cordova, A. A. (2011). "Continuous quality improvement techniques for data collection in asset management systems." *ASCE Journal of Construction Engineering and Management*, 138(11), 127-141.
- Ming-Syan, C., Jiawei, H., and Yu, P. S. (1996). "Data mining: an overview from a database perspective." *Knowledge and Data Engineering, IEEE Transactions on*, 8(6), 866-883.
- Mishalani, R. G., and Gong, L. (2008). "Evaluating impact of pavement condition sampling advances on life-cycle management." *Transportation Research Record: Journal of the Transportation Research Board*, 20(68), 3-9.
- Mishalani, R. G., and Koutsopoulos, H. N. (2002). "Modeling the spatial behavior of infrastructure condition." *Transportation Research, Part B (Methodological)*, 36B(2), 171-194.
- Mooney, M. A., Khanna, V., Yuan, F., Parsons, T., and Miller, G. A. (2005). "Web-based pavement infrastructure management system." *Journal of Infrastructure Systems*, 11(4), 241-249.
- Morian, D., Stoeffels, S., and Firth, D. J. (2002). "Quality management of pavement performance data." *Pavement Evaluation 2002*, Blacksburg, Va.
- NCHRP (1997). *Pavement Subsurface Drainage Systems*, National Cooperative Highway Research Program, Washington, D. C.

- NIST (2003). "NIST/SEMATECH e-Handbook of Statistical Methods", <http://www.itl.nist.gov/div898/handbook/>, Accessed 4/1/2012.
- NmDOT (2007). "The New Mexico department of transportation's pavement inspection program, a successful partnership." Office of Highway and Transportation Operations, New Mexico Department of Transportation,, Santa, Fe, 68.
- Nunez, M. M., and Shahin, M. Y. (1986). "Pavement condition data analysis and modeling." *Transportation Research Record*, 10(70), 125-132.
- Olivera, F. (2001). "Extracting hydrologic information from spatial data for HMS modeling." *Journal of Hydrologic Engineering*, 6(6), 524-531.
- Olivera, F., Koka, S., and Nelson, J. (2006). "A GIS application for the analysis of hydrologic networks using vector spatial sata." *Transactions in GIS 10*, 12(7), 355-375.
- Olivera, F., and Maidment, D. (1999). "Geographic information systems (GIS)-based spatially distributed model for runoff routing." *Water Resources Research*, 35(4), 1155-1164.
- Ong, G. P., Noureldin, A. S., and Sinha, K. C. (2010). "An information management approach for pavement condition data integrity quality assurance." *Proc., Presented in the Transportation Research Board 89th Annual Meeting*, Transportation Research Board, Washington, D.C.
- Ong, G. P., Noureldin, S., and Sinha, K. C. (2009). "Methodology to evaluate quality of pavement surface distress data collected by automated techniques."

Transportation Research Record: Journal of the Transportation Research Board, 20(93), 3-11.

Ong, G. P., and Sinha, K. C. (2008). "Evaluating international roughness index data quality at project and network levels." *Proc., 2008 Airfield and Highway Pavements Conference: Efficient Pavements Supporting Transportation's Future*, American Society of Civil Engineers, 385-396.

OwusuAbabio, S., and Owusu, A. (1998). "Effect of neural network topology on flexible pavement cracking prediction." *Computer-Aided Civil and Infrastructure Engineering*, 13(5), 349.

Papadimitriou, S., Kitagawa, H., Gibbons, P. B., and Faloutsos, C. (2003). "LOCI: fast outlier detection using the local correlation integral." *Proc., Data Engineering, 2003. Proceedings. 19th International Conference on*, 315-326.

Park, E. S., Smith, R. E., Freeman, T. J., Spiegelman, C. H. (2008), "A Bayesian approach for improved pavement performance prediction" *Journal of Applied Statistics*, 35(11), 1219-1238.

Prechaverakul, S. (1995). "Using a knowledge-based expert system and fuzzy logic for minor rehabilitation projects in Ohio." *Traffic safety*, 1497(19).

Rada, G. R. (1997). "Assessment of long-term pavement performance program manual distress data variability: bias and precision." *Transportation Research Record: Journal of the Transportation Research Board*, 1592(1), 151.

Rada, G. R., Wu, C. L., Elkins, G. E., Bhandari, R. K., and Bellinger, W. Y. (1998). "Update of long-term pavement performance manual distress data variability:

- Bias and precision." *Transportation Research Record: Journal of the Transportation Research Board*, 1643(1), 71-79.
- Redman, T. C., and Godfrey, A. B. (1996). *Data Quality for the Information Age*, Artech House Publishers, Boston, MA.
- Rizk, K., Wagen, J. F., and Gardiol, F. (2000). "Influence of database accuracy on two-dimensional ray-tracing-based predictions in urban microcells." *Vehicular Technology, IEEE Transactions on*, 49(2), 631-642.
- Rosenberg, B. (1974). "Error rates in CRSP and Compustat data bases and their implications." *The journal of finance*, 29(4), 1303.
- Rousseeuw, P. J., and Leroy, A. M. (1987). *Robust Regression and Outlier Detection*, John Wiley & Sons., New York.
- Saliminejad, S., and Gharaibeh, N. G. (2012). "A spatial-bayesian technique for imputing pavement network repair data." *Computer-Aided Civil and Infrastructure Engineering*, 27(8), 594-607.
- Saltelli, A., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M., and Tarantola, S. (2008). *Global Sensitivity Analysis - The Primer*, John Wiley & Sons, West Sussex, England.
- Saltelli, A., Tarantola, S., Campolongo, F., and Ratto, M. (2004). *Sensitivity Analysis in Practice: A Guide to Assessing Scientific Models*, John Wiley & Sons, West Sussex, England.
- Schabenberger, O., and Gotway, C. (2004). *Statistical Methods for Spatial Data Analysis*, Chapman & Hall/CRC, New York.

- Shahin, M. Y. (2005). *Pavement Management for Airports, Roads, and Parking Lots*, Springer, New York.
- Shahin, M. Y., Kohn, S. D., Lytton, R. L., and McFaraland, W. F. (1985). "Pavement M&R budget optimization using the incremental benefit-cost technique." *1st North American Pavement Management Conference, Washington, D.C.*
- Shahin, M. Y., Nunez, M. M., Broten, M. R., Carpenter, S. H., and Sameh, A. (1987). "New techniques for modeling pavement deterioration." *Transportation Research Record*, 1123(1), 40-46.
- Shekharan, A. R. (1998). "Effect of noisy data on pavement performance prediction by artificial neural networks." *Transportation Research Record: Journal of the Transportation Research Board*, 1643(1), 7.
- Shekharan, R., Frith, D., Chowdhury, T., Larson, C., and Morian, D. (2007). "Effects of comprehensive quality assurance/quality control plan on pavement management." *Transportation Research Record: Journal of the Transportation Research Board*, 1990(1), 65-71.
- Smith, R. E. (2002). "Integrating pavement preservation into a local agency pavement management system." *Transportation Research Record*, 1795(1), 27-32.
- Spencer, B. D. (1994). "Sensitivity of benefit-cost analysis of data programs to monotone misspecification." *Journal of statistical planning and inference*, 39(1), 19.
- Spiegelman, C. H., Park, E. S., and Rilett, L. R. (2011). *Transportation Statistics and Microsimulation*, Chapman & Hall/CRC, New York.

- Stampley, B., Miller, B., Smith, R. E., and Scullion, T. (1993). *Pavement Management Information System Concepts, Equations, and Analysis Models*, Texas Department of Transportation, Austin, Tx.
- Sulieman, H., McLellan, P. J., and Bacon, D. W. (2001). "A profile-based approach to parametric sensitivity analysis of nonlinear regression models." *Technometrics*, 43(4), 425-433.
- Taha, M. A., and Hanna, A. S. (1995). "Evolutionary neural network model for the selection of pavement maintenance strategy." *Transportation Research Record: Journal of the Transportation Research Board*, 1497(1), 70-76.
- Takeyama, Y., Shimada, Y., and Fukuda, T. (1990). "Evaluation system for managing asphalt pavements based on Markov chain model." *Doboku Gakkai-shi/Journal of the Japan Society of Civil Engineers*, 420(V-13), 135-135.
- TRB (2009). *Glossary of Highway Quality Assurance Terms*, Transportation Research Board, Washington, D. C.
- TxDOT (2005). *Status of Statewide Pavement Condition Goal*, Texas Department of Transportation, Austin, Tx.
- Valsami, G. (2000). "Non-linear regression analysis with errors in both variables: estimation of co-operative binding parameters." *Biopharmaceutics & drug disposition*, 21(1), 7.
- Vandevalk-Ostander, A., Guerre, J., and Harrison, F. (2003). *Review of Data Integration Practices and Their Applications to Transportation Asset Management*, Federal Highway Administration, Washington, D. C.

- Vlahogianni, E. I., Karlaftis, M. G., and Golias, J. C. (2007). "Spatio-temporal short-term urban traffic volume forecasting using genetically optimized modular networks." *Computer-Aided Civil and Infrastructure Engineering*, 22(5), 317-325.
- Waldo, D. R. (2005). "Accuracy and bias of race/ethnicity codes in the Medicare enrollment database." *Health care financing review*, 26(2), 61.
- Wand, Y., and Wang, R. Y. (1996). "Anchoring data quality dimensions in ontological foundations." *Communications of the ACM*, 39(11), 86-95.
- Wang, K. C. P., and Li, Q. (2011). "Pavement smoothness prediction based on Fuzzy and gray theories." *Computer-Aided Civil and Infrastructure Engineering*, 26(1), 69-76.
- Wang, K. C. P., and Zaniewski, J. (1995). "Sensitivity of pavement network optimization system to its prediction models." *Transportation Research Record: Journal of the Transportation Research Board*, 1308(1508), 22.
- Wang, X., and Kockelman, K. M. (2009). "Forecasting network data spatial interpolation of traffic counts from texas data." *Transportation Research Record*, 2105(1), 100-108.
- Yang, C., Tsai, Y., and Wang, Z. (2009). "Algorithm for spatial clustering of pavement segments." *Computer-Aided Civil and Infrastructure Engineering*, 24(2), 93-108.
- Yang, J. (2003). "Forecasting overall pavement condition with neural networks: application on Florida highway network." *Transportation Research Record: Journal of the Transportation Research Board*, 1853(1), 3.

- Ying, L., and Salari, E. (2010). "Beamlet transform-based technique for pavement crack detection and classification." *Computer-Aided Civil and Infrastructure Engineering*, 25(8), 572-580.
- Zhang, B., Wang, Q., Sun, Q., and Li, S. (2009). "Spatio-temporal simulation of water quality based on SD-GIS accidental water pollution." *Wuhan Daxue Xuebao (Xinxi Kexue Ban), Geomatics and Information Science of Wuhan University*, 34(3), 348-351.
- Zhang, W., Vidal, R., Heyden, A., and Ma, Y. (2007). "Nonparametric estimation of multiple structures with outliers." *Dynamic Vision*, Pringer Verlag, VA, 60-74.
- Zhang, Z., Smith, S. G., and Hudson, W. R. (2001). "Geographic information system implementation plan for pavement management information system: Texas department of transportation." *Transportation research record*, 1769(1), 46-50.

APPENDIX A: CODE OF COMPUTER PROGRAM DEVELOPED FOR
DETECTING PROXIMITY-BASED OUTLIERS IN PAVEMENT CONDITION

DATA

```
ShapefileAddress='C:\Siamak\Studies\Papers\9-Automated Outlier  
Detection Method for Pavement Surface Distress  
Data\Data\Shapefiles\Bryan\PointSections.shp';  
    RankedSHPAddress = 'C:\Siamak\Studies\Papers\9-Automated Outlier  
Detection Method for Pavement Surface Distress  
Data\Data\Shapefiles\Bryan\PointSectionsW1.shp';  
    ClusteredSHPAddress = 'C:\Siamak\Studies\Papers\9-Automated Outlier  
Detection Method for Pavement Surface Distress  
Data\Data\Shapefiles\Bryan\PointSectionsW2.shp';  
    FiguresAddress='C:\Siamak\Studies\Papers\9-Automated Outlier  
Detection Method for Pavement Surface Distress Data\Outputs';  
    OI DetectedAddress = 'C:\Siamak\Studies\Papers\9-Automated Outlier  
Detection Method for Pavement Surface Distress  
Data\Data\Shapefiles\Bryan\PointSectionsW6.shp';  
    %    FiguresAddress='C:\Outputs';
```

```
Alpha=2.5; % Limit on Point  
Beta=1; %Limit on section  
Gamma=3; %Limit on the error Gamm>Beta/Alpha  
Theta=7; %Mahalanobis Distance Limit for ourlier sections  
PracticalLimit=5;  
DistLimit=0.017;
```

```
HighwayNames={'BS0006R' 'BS0006S' 'BS0021H' 'BS0036J' 'BU0084R'  
'BU0290F' 'FM0002' 'FM0003' 'FM0027' 'FM0039' 'FM0046' 'FM0050'  
'FM0060' 'FM0080' 'FM0109' 'FM0111' 'FM0149' ...  
'FM0158' 'FM0159' 'FM0166' 'FM0230' 'FM0244' 'FM0246' 'FM0247'  
'FM0332' 'FM0362' 'FM0379' 'FM0389' 'FM0390' 'FM0391' 'FM0405' 'FM0416'  
'FM0437' 'FM0485' 'FM0486' 'FM0487' ...  
'FM0488' 'FM0489' 'FM0542' 'FM0553' 'FM0577' 'FM0579' 'FM0594'  
'FM0696' 'FM0745' 'FM0811' 'FM0831' 'FM0832' 'FM0833' 'FM0845' 'FM0908'  
'FM0912' 'FM0937' 'FM0974' 'FM0975' ...  
'FM0976' 'FM0977' 'FM0978' 'FM0979' 'FM0980' 'FM1073' 'FM1097'  
'FM1119' 'FM1124' 'FM1146' 'FM1147' 'FM1155' 'FM1179' 'FM1227' 'FM1331'  
'FM1361' 'FM1362' 'FM1363' 'FM1364' ...  
'FM1365' 'FM1366' 'FM1367' 'FM1370' 'FM1371' 'FM1372' 'FM1373'  
'FM1374' 'FM1375' 'FM1394' 'FM1428' 'FM1444' 'FM1445' 'FM1449' 'FM1451'  
'FM1452' 'FM1469' 'FM1486' 'FM1511' ...  
'FM1512' 'FM1580' 'FM1600' 'FM1618' 'FM1644' 'FM1687' 'FM1688'  
'FM1696' 'FM1697' 'FM1712' 'FM1748' 'FM1774' 'FM1786' 'FM1791' 'FM1848'  
'FM1915' 'FM1935' 'FM1940' 'FM1948' ...
```

```

'FM1963' 'FM2000' 'FM2027' 'FM2038' 'FM2039' 'FM2095' 'FM2096'
'FM2116' 'FM2154' 'FM2155' 'FM2158' 'FM2159' 'FM2193' 'FM2223' 'FM2268'
'FM2269' 'FM2289' 'FM2293' 'FM2296' ...
'FM2346' 'FM2347' 'FM2413' 'FM2445' 'FM2446' 'FM2447' 'FM2485'
'FM2502' 'FM2539' 'FM2547' 'FM2548' 'FM2549' 'FM2550' 'FM2562' 'FM2570'
'FM2620' 'FM2621' 'FM2628' 'FM2679' ...
'FM2693' 'FM2726' 'FM2774' 'FM2776' 'FM2777' 'FM2780' 'FM2793'
'FM2818' 'FM2819' 'FM2821' 'FM2865' 'FM2929' 'FM2935' 'FM2954' 'FM2988'
'FM2989' 'FM3058' 'FM3059' 'FM3060' ...
'FM3061' 'FM3090' 'FM3091' 'FM3178' 'FM3179' 'FM3242' 'FM3285'
'FM3411' 'FM3454' 'FM3455' 'FM3456' 'FM3478' 'FM3501' 'FS0003' 'FS0390'
'FS1366' 'FS2155' 'FS2296' 'IH0045' ...
'PR0012' 'PR0040' 'PR0040A' 'PR0057' 'PR0064' 'RE0004' 'SH0006'
'SH0007' 'SH0014' 'SH0019' 'SH0021' 'SH0030' 'SH0036' 'SH0040' 'SH0047'
'SH0075' 'SH0090' 'SH0105' 'SH0150' ...
'SH0164' 'SH0179' 'SH0237' 'SH0308' 'SHOSR' 'SL0083' 'SL0160'
'SL0208' 'SL0262' 'SL0361' 'SL0429' 'SS0059' 'SS0067' 'SS0104' 'SS0113'
'SS0114' 'SS0125' 'SS0174' 'SS0231' ...
'SS0234' 'SS0515' 'US0077' 'US0079' 'US0084' 'US0190' 'US0287'
'US0290'};

```

```

cd(fullfile(FolderAddress)) %for working with the maps more easily
OIDetectedAddress = 'C:\Siamak\Studies\Papers\9-Automated Outlier
Detection Method for Pavement Surface Distress
Data\Data\Shapefiles\Bryan\PointSectionsW3.shp';

```

```

RankSections(ShapefileAddress, HighwayNames, RankedSHPAddress, DistLimit);
ClusterFunc(RankedSHPAddress, ClusteredSHPAddress, HighwayNames);

```

```

OIDetectionFunc(ClusteredSHPAddress, FiguresAddress, OIDetectedAddress, HighwayNames, Alpha, Beta, Gamma, Theta, PracticalLimit);

```

```

function RoundedClusterLengths=ArrayRounder (ClusterLengths)
ClusterNum=size (ClusterLengths, 2);
SectionNum=round (sum ( (ClusterLengths) ));
RoundedClusterLengths=round (ClusterLengths);
RoundingError=SectionNum-sum (RoundedClusterLengths);
    if (RoundingError== -1)

RoundedClusterLengths (ClusterNum)=RoundedClusterLengths (ClusterNum) -1;
        elseif (RoundingError== -2)

RoundedClusterLengths (ClusterNum)=RoundedClusterLengths (ClusterNum) -1;
            RoundedClusterLengths (ClusterNum-
1)=RoundedClusterLengths (ClusterNum-1) -1;
                elseif (RoundingError== -3)

RoundedClusterLengths (ClusterNum)=RoundedClusterLengths (ClusterNum) -1;
                    RoundedClusterLengths (ClusterNum-
1)=RoundedClusterLengths (ClusterNum-1) -1;
                        RoundedClusterLengths (ClusterNum-
2)=RoundedClusterLengths (ClusterNum-2) -1;
                            else

RoundedClusterLengths (ClusterNum)=RoundedClusterLengths (ClusterNum) +Rou
ndingError;
                                end
RoundedClusterLengths;

```

```

function [ C1Num, OI, OI7,
OI7N,OI8,OI8N,OI9,OI9N,OI10,OI10N,OI11,OI11N,OI7F,OI8F,OI9F,OI10F,OI11F
] = C1NumByObjID( ObjectID, HighwayMatrix )
%UNTITLED Summary of this function goes here
% Detailed explanation goes here
%This functions searches the Highways matrix for Sections with a known
%Object ID and finds their associated Outlierness Index
[RowsNum,ColumnNum]=size(HighwayMatrix);
C1Num=0;
for i=1:RowsNum
    if (HighwayMatrix(i,1)==ObjectID)
        C1Num=HighwayMatrix(i,8);
        OI=HighwayMatrix(i,9);
        OI7=HighwayMatrix(i,10);
        OI7N=HighwayMatrix(i,11);
        OI8=HighwayMatrix(i,12);
        OI8N=HighwayMatrix(i,13);
        OI9=HighwayMatrix(i,14);
        OI9N=HighwayMatrix(i,15);
        OI10=HighwayMatrix(i,16);
        OI10N=HighwayMatrix(i,17);
        OI11=HighwayMatrix(i,18);
        OI11N=HighwayMatrix(i,19);
        OI7F=HighwayMatrix(i,20);
        OI8F=HighwayMatrix(i,21);
        OI9F=HighwayMatrix(i,22);
        OI10F=HighwayMatrix(i,23);
        OI11F=HighwayMatrix(i,24);
    end
end
end

```

```

function
[a]=ClusterFunc(RankedSHPAddress,ClusteredSHPAddress,HighwayNames)

%Setting the Environment

%Computer Addresses

NetworkData=shaperead(RankedSHPAddress,'Selector',...
{@(v1,v2,v3,v4,v5) (v1>=1) && (v2>=1) && (v3>=1) && (v4>=1) &&
(v5>=1),'DS2011','DISTRESS_S','DS2009','DS2008','DS2007'});

%data Extraction
HWNArray=size(HighwayNames);
HighwayNum=HWNArray(1,2);

[TotalSecNum,aa]=size(NetworkData);

for HN=1:HighwayNum
    if (HN~=188)
    % for HN=1:10
    % for ThisHighwayName=HighwayNames
        ThisHighwayName=HighwayNames(HN);
        HN
    %     strcmp('HN=',num2str(HN))
    %     strcmp('HighwayName= ',ThisHighwayName)
        ThisHwSecNum=0;
        clear HighwayMatrix
        for ii=1:TotalSecNum
            if (strcmp(NetworkData(ii).SIGNED_HIG,ThisHighwayName))
                ThisHwSecNum=ThisHwSecNum+1;
                HighwayMatrix(ThisHwSecNum,1)=NetworkData(ii).OBJECTID;
                HighwayMatrix(ThisHwSecNum,2)=NetworkData(ii).Rank;
                HighwayMatrix(ThisHwSecNum,3)=NetworkData(ii).DS2007;
                HighwayMatrix(ThisHwSecNum,4)=NetworkData(ii).DS2008;
                HighwayMatrix(ThisHwSecNum,5)=NetworkData(ii).DS2009;
                HighwayMatrix(ThisHwSecNum,6)=NetworkData(ii).DISTRESS_S;
                HighwayMatrix(ThisHwSecNum,7)=NetworkData(ii).DS2011;
                %HighwayMatrix(ThisHwSecNum,8)=NetworkData(ii).ClNum;
                %HighwayMatrix(ThisHwSecNum,9)=NetworkData(ii).OI;

            end
        end
        strcmp('Number of Section in this highway is:
',num2str(ThisHwSecNum));

        n=0;
        if (ThisHwSecNum>5)

```



```

        options2 =
gaoptimset('MutationFcn',@mutationadaptfeasible,'CrossoverFcn',@crossover
scattered,'CrossoverFraction',0.3,'PopInitRange',Bound,'Display','off
','StallGenL',20,'Generations',100, ...

'PopulationSize',100,'StallTimeLimit',1000,'TolFun',1e-
6,'EliteCount',2);
    try
        [ClusterLengths,fval] =
ga(fitnessFcn,numberOfVariables,[],[],MyAeq,MyBeq,LB,UB,[],options2);
    catch ME
        msg=ME.message
    end
    if (fval<fmin)
        fmin=fval;
        clear ClusterLengthsMax;
        ClusterLengthsMax=ArrayRounder(ClusterLengths);
        NCmax=NC;
    end
end

%Save Cluster Number
clear SortedThisHighwayMatrix;
CumNum=1;
SortedThisHighwayMatrix=sortrows(HighwayMatrix,2);
[a,NC]=size(ClusterLengthsMax);
for j=1:NC
    OutliernessIndex=0;
    ThisClusterSecNums=ClusterLengthsMax(j);
    if (ThisClusterSecNums>0)
        clear OutliernessIndex;
        %Calculate OI
        ThisClusterNumber=j;
        if (ThisClusterSecNums>1)
            clear ThisClusterArray;
            for k=1:ThisClusterSecNums

SortedThisHighwayMatrix(CumNum,8)=ThisClusterNumber;

ThisClusterArray(k,:)=SortedThisHighwayMatrix(CumNum,3:7);
                CumNum=CumNum+1;
            end

%                labels = {'2007','2008','2009','2010','2011'};
%                figure;
parallelcoords(ThisClusterArray(:,:),'labels',labels);
%                title( strcat('cluster num
',num2str(ThisClusterNumber),' in highway ',ThisHighwayName, ' ,
including ',num2str(ThisClusterSecNums), ' sections' ));
%                ylim([50,100]);

end

```

```

        end
    end

    %All in One Figure
    %     ClNumForChart=SortedThisHighwayMatrix(:,8);
    %     labels = {'2007','2008','2009','2010','2011'};
    %     figure;
parallelcoords(SortedThisHighwayMatrix(:,3:7),'group',ClNumForChart,'la
bels',labels)
%     title('All Highways');

else
    for j=1:ThisHwSecNum
        HighwayMatrix(j,8)=0;
    end
end

        %Save Cluster Number in the database .IO field
    for k=1:TotalSecNum
        if (strcmp(NetworkData(k).SIGNED_HIG,ThisHighwayName))
            %NetworkData(k).OI=OIByObjID(NetworkData(k).OBJECTID,
SortedThisHighwayMatrix );
            NetworkData(k).ClNum=ClNumByObjID(NetworkData(k).OBJECTID,
SortedThisHighwayMatrix );
        end
    end

end
%     100*HN/HighwayNum;
end
shapewrite(NetworkData,ClusteredSHPAddress);
shapeinfo(ClusteredSHPAddress)
a='Done!'

```

```
function BRMNum=ConvertBRMTextToNum(BRMText)
BRMText1=BRMText(1:4);
BRMSignT=BRMText(6);
BRMText2=BRMText(7:10);

BRMNum1=str2num(BRMText1);
BRMNum2=str2num(BRMText2);
if (BRMSignT=='+')
    BRMNum= BRMNum1+BRMNum2;
else
    BRMNum= (-1) * (BRMNum1+BRMNum2);
end
```

```
function [ OI ] = OIByObjID( ObjectID, HighwayMatrix )
%UNTITLED Summary of this function goes here
% Detailed explanation goes here
%This functions searches the Highways matrix for Sections with a known
%Object ID and finds their associated Outlierness Index
[RowsNum,ColumnNum]=size(HighwayMatrix);
for i=1:RowsNum
    if (HighwayMatrix(i,1)==ObjectID)
        OI=HighwayMatrix(i,9);
    end
end
end
```

```

function
OIDetectionFunc(ClusteredSHPAddress, FiguresAddress, OI DetectedAddress, Hi
ghwayNames, Alpha, Beta, Gamma, Theta, PracticalLimit)
NetworkData=shaperead(ClusteredSHPAddress, 'Selector', ...
{@(v1,v2,v3,v4,v5,v6) (v1>=1) && (v2>=1) && (v3>=1) && (v4>=1) &&
(v5>=1) &&
(v6>=1), 'DS2011', 'DISTRESS_S', 'DS2009', 'DS2008', 'DS2007', 'ClNum'});
[TotalSecNum, aa]=size(NetworkData);

MyLineWidth=1;
SecIsOutlier='false';

for k=1:TotalSecNum
    NetworkData(k).OI=0;
end

%data Extraction
HWNArray=size(HighwayNames);
HighwayNum=HWNArray(1,2);

[TotalSecNum, aa]=size(NetworkData);

for HN=1:HighwayNum
    if (HN~=188)

        % for HN=1:1
        ThisHighwayName=HighwayNames(HN);
        % if (strcmp(ThisHighwayName, 'FM0416'))
        %     nn=1;
        % end
        % ThisHighwayName='SH0036';
        %ThisHighwayName='US0079'; %Biggest
        %ThisHighwayName='';
        ThisHwSecNum=0;
        clear HighwayMatrix
        for ii=1:TotalSecNum
            if (strcmp(NetworkData(ii).SIGNED_HIG, ThisHighwayName))
                ThisHwSecNum=ThisHwSecNum+1;
                HighwayMatrix(ThisHwSecNum,1)=NetworkData(ii).OBJECTID;
                HighwayMatrix(ThisHwSecNum,2)=NetworkData(ii).Rank;
                HighwayMatrix(ThisHwSecNum,3)=NetworkData(ii).DS2007;
                HighwayMatrix(ThisHwSecNum,4)=NetworkData(ii).DS2008;
                HighwayMatrix(ThisHwSecNum,5)=NetworkData(ii).DS2009;
                HighwayMatrix(ThisHwSecNum,6)=NetworkData(ii).DISTRESS_S;
                HighwayMatrix(ThisHwSecNum,7)=NetworkData(ii).DS2011;
                HighwayMatrix(ThisHwSecNum,8)=NetworkData(ii).ClNum;
                %HighwayMatrix(ThisHwSecNum,9)=NetworkData(ii).OI;
                %HighwayMatrix(ThisHwSecNum,10)=NetworkData(ii).OI7;
                %HighwayMatrix(ThisHwSecNum,11)=NetworkData(ii).OI7N;
                %HighwayMatrix(ThisHwSecNum,12)=NetworkData(ii).OI8;
                %HighwayMatrix(ThisHwSecNum,13)=NetworkData(ii).OI8N;
            end
        end
    end
end

```

```

        %HighwayMatrix(ThisHwSecNum,14)=NetworkData(ii).OI9;
        %HighwayMatrix(ThisHwSecNum,15)=NetworkData(ii).OI9N;
        %HighwayMatrix(ThisHwSecNum,16)=NetworkData(ii).OI10;
        %HighwayMatrix(ThisHwSecNum,17)=NetworkData(ii).OI10N;
        %HighwayMatrix(ThisHwSecNum,18)=NetworkData(ii).OI11;
        %HighwayMatrix(ThisHwSecNum,19)=NetworkData(ii).OI11N;
        %HighwayMatrix(ThisHwSecNum,20)=NetworkData(ii).OI7F;
        %HighwayMatrix(ThisHwSecNum,21)=NetworkData(ii).OI8F;
        %HighwayMatrix(ThisHwSecNum,22)=NetworkData(ii).OI9F;
        %HighwayMatrix(ThisHwSecNum,23)=NetworkData(ii).OI10F;
        %HighwayMatrix(ThisHwSecNum,24)=NetworkData(ii).OI11F;
    end
    end
    strcmp('Number of Section in this highway is:
',num2str(ThisHwSecNum));
    %ClusteringArray=HighwayMatrix(:,3:6);
    if (ThisHwSecNum>0)
        MaxClNum= max(HighwayMatrix(:,8));
        ThisHwSecNum=size(HighwayMatrix(),1);

        for ThisClNum=1:MaxClNum
            FoundSecNum=0;
            clear ClusterMatrix;
            for ThisHwSeci=1:ThisHwSecNum
                if (HighwayMatrix(ThisHwSeci,8)==ThisClNum)
                    FoundSecNum=FoundSecNum+1;
                    ClusterMatrix(FoundSecNum,1)=HighwayMatrix(ThisHwSeci,1);
%OBJECTID;
                    ClusterMatrix(FoundSecNum,2)=HighwayMatrix(ThisHwSeci,2);
%Rank;

ClusterMatrix(FoundSecNum,3)=HighwayMatrix(ThisHwSeci,3)+normrnd(0.01,0
.01);    %DS2007;

ClusterMatrix(FoundSecNum,4)=HighwayMatrix(ThisHwSeci,4)+normrnd(0.01,0
.01);    %DS2008;

ClusterMatrix(FoundSecNum,5)=HighwayMatrix(ThisHwSeci,5)+normrnd(0.01,0
.01);    %DS2009;

ClusterMatrix(FoundSecNum,6)=HighwayMatrix(ThisHwSeci,6)+normrnd(0.01,0
.01);    %DISTRESS_S;

ClusterMatrix(FoundSecNum,7)=HighwayMatrix(ThisHwSeci,7)+normrnd(0.01,0
.01);    %DS2011
                    ClusterMatrix(FoundSecNum,8)=HighwayMatrix(ThisHwSeci,8);
%ClNum;
                    %ClusterMatrix(FoundSecNum,9)=HighwayMatrix(ThisHwSeci,9);
%OI;

%ClusterMatrix(FoundSecNum,10)=HighwayMatrix(ThisHwSeci,10);    %OI7;

%ClusterMatrix(FoundSecNum,11)=HighwayMatrix(ThisHwSeci,11); %OI7N;

```

```

        %ClusterMatrix(FoundSecNum,12)=HighwayMatrix(ThisHwSeci,12); %OI8;
%ClusterMatrix(FoundSecNum,13)=HighwayMatrix(ThisHwSeci,13); %OI8N;
%ClusterMatrix(FoundSecNum,14)=HighwayMatrix(ThisHwSeci,14); %OI9;
%ClusterMatrix(FoundSecNum,15)=HighwayMatrix(ThisHwSeci,15); %OI9N;
%ClusterMatrix(FoundSecNum,16)=HighwayMatrix(ThisHwSeci,16); %OI10;
%ClusterMatrix(FoundSecNum,17)=HighwayMatrix(ThisHwSeci,17); %OI10N;
%ClusterMatrix(FoundSecNum,18)=HighwayMatrix(ThisHwSeci,18); %OI11;
%ClusterMatrix(FoundSecNum,19)=HighwayMatrix(ThisHwSeci,19); %OI11N;
%ClusterMatrix(FoundSecNum,20)=HighwayMatrix(ThisHwSeci,20); %OI7F;
%ClusterMatrix(FoundSecNum,21)=HighwayMatrix(ThisHwSeci,21); %OI8F;
%ClusterMatrix(FoundSecNum,22)=HighwayMatrix(ThisHwSeci,22); %OI9F;
%ClusterMatrix(FoundSecNum,23)=HighwayMatrix(ThisHwSeci,23); %OI10F;
%ClusterMatrix(FoundSecNum,24)=HighwayMatrix(ThisHwSeci,24); %OI11F;
        end
        end
        ThisClusterSize=size(ClusterMatrix,1);
        %Calc OIs

clear CIMatrix;
clear MDArray;
clear MedianArray;
clear P25Array;
clear P75Array;
clear IQR0;
clear Median0;
clear Percentile075;
clear Percentile025;
clear OIArray;

if (ThisClusterSize>5)

for i=1:ThisClusterSize
clear TempMatrix;
k=1;
for j=1:ThisClusterSize
if (j~=i)
TempMatrix(k,:)=ClusterMatrix(j,3:7);
CIMatrix(k,:)=ClusterMatrix(j,3:7);
k=k+1;
end
end

```



```

        end

TempMatrix(ThisClusterSize,:)=ClusterMatrix(ThisClusterSize,3:7);
CIMatrix(ThisClusterSize,:)=ClusterMatrix(ThisClusterSize,3:7);

%         MDArray(i)=mahal(ClusterMatrix(i,3:7),TempMatrix(:,,:));

    end

%         Multivariate Normal Test
%         Mulnortest(CIMatrix(:,,:),0.05)

for i=1:ThisClusterSize
    MDArray(i)=mahal(ClusterMatrix(i,3:7),CIMatrix(:,,:));
end

%Arrays
for i=1:5
    MedianArray(i)=median(CIMatrix(:,i));
    P25Array(i)=prctile(CIMatrix(:,i),25);
    P75Array(i)=prctile(CIMatrix(:,i),75);
end

%Calc OI0 using IQR
IQR0=iqr(MDArray(:));
Median0=median(MDArray(:));
Percentile075=prctile(MDArray(:),75);
Percentile025=prctile(MDArray(:),25);
for i=1:ThisClusterSize
    if (MDArray(i)>Median0)
        OIArray(i)=(MDArray(i)-Median0)/IQR0;
    else
        OIArray(i)=0.01;
    end
end

%Calc OI7N

clear MDArray7N;
clear MedianArray7N;
clear P25Array7N;
clear P75Array7N;
clear IQR7N;
clear Median7N;
clear Percentile0757N;
clear Percentile0257N;
clear OINot7Array;
OrderArray=[4,5,6,7];

```

```

        CIOrderArray=[2,3,4,5];
        for i=1:ThisClusterSize

MDArray7N(i)=mahal(ClusterMatrix(i,OrderArray),CIMatrix(:,CIOrderArray)
);
        end
        for i=1:4
            MedianArray7N(i)=median(CIMatrix(:,CIOrderArray(i)));
            P25Array7N(i)=prctile(CIMatrix(:,CIOrderArray(i)),25);
            P75Array7N(i)=prctile(CIMatrix(:,CIOrderArray(i)),75);
        end
        IQR7N=iqr(MDArray7N(:));
        Median7N=median(MDArray7N(:));
        Percentile0757N=prctile(MDArray7N(:),75);
        Percentile0257N=prctile(MDArray7N(:),25);
        for i=1:ThisClusterSize
            if (MDArray7N(i)>Median7N)
                OINot7Array(i)=(MDArray7N(i)-Median7N)/IQR7N;
            else
                OINot7Array(i)=0.01;
            end
        end
        end

        %Calc OI8N
        clear MDArray8N;
        clear MedianArray8N;
        clear P25Array8N;
        clear P75Array8N;
        clear IQR8N;
        clear Median8N;
        clear Percentile0758N;
        clear Percentile0258N;
        clear OINot8Array;
        OrderArray=[3,5,6,7];
        CIOrderArray=[1,3,4,5];
        for i=1:ThisClusterSize

MDArray8N(i)=mahal(ClusterMatrix(i,OrderArray),CIMatrix(:,CIOrderArray)
);
        end
        for i=1:4
            MedianArray8N(i)=median(CIMatrix(:,CIOrderArray(i)));
            P25Array8N(i)=prctile(CIMatrix(:,CIOrderArray(i)),25);
            P75Array8N(i)=prctile(CIMatrix(:,CIOrderArray(i)),75);
        end
        IQR8N=iqr(MDArray8N(:));
        Median8N=median(MDArray8N(:));
        Percentile0758N=prctile(MDArray8N(:),75);
        Percentile0258N=prctile(MDArray8N(:),25);
        for i=1:ThisClusterSize
            if (MDArray8N(i)>Median8N)
                OINot8Array(i)=(MDArray8N(i)-Median8N)/IQR8N;
            else

```

```

        OINot8Array(i)=0.01;
    end
end

%Calc OI9N
clear MDArr9N;
clear MedianArr9N;
clear P25Arr9N;
clear P75Arr9N;
clear IQR9N;
clear Median9N;
clear Percentile0759N;
clear Percentile0259N;
clear OINot9Array;
OrderArray=[3,4,6,7];
CIOrderArray=[1,2,4,5];
for i=1:ThisClusterSize

MDArr9N(i)=mahal(ClusterMatrix(i,OrderArray),CIMatrix(:,CIOrderArray)
);
    end
    for i=1:4
        MedianArr9N(i)=median(CIMatrix(:,CIOrderArray(i)));
        P25Arr9N(i)=prctile(CIMatrix(:,CIOrderArray(i)),25);
        P75Arr9N(i)=prctile(CIMatrix(:,CIOrderArray(i)),75);
    end
    IQR9N=iqr(MDArr9N(:));
    Median9N=median(MDArr9N(:));
    Percentile0759N=prctile(MDArr9N(:),75);
    Percentile0259N=prctile(MDArr9N(:),25);
    for i=1:ThisClusterSize
        if (MDArr9N(i)>Median9N)
            OINot9Array(i)=(MDArr9N(i)-Median9N)/IQR9N;
        else
            OINot9Array(i)=0.01;
        end
    end
end

%Calc OI10N
clear MDArr10N;
clear MedianArr10N;
clear P25Arr10N;
clear P75Arr10N;
clear IQR10N;
clear Median10N;
clear Percentile07510N;
clear Percentile02510N;
clear OINot10Array;
OrderArray=[3,4,5,7];
CIOrderArray=[1,2,3,5];
for i=1:ThisClusterSize
MDArr10N(i)=mahal(ClusterMatrix(i,OrderArray),CIMatrix(:,CIOrderArray)
);

```

```

end
for i=1:4
    MedianArray10N(i)=median(CIMatrix(:,CIOrderArray(i)));
    P25Array10N(i)=prctile(CIMatrix(:,CIOrderArray(i)),25);
    P75Array10N(i)=prctile(CIMatrix(:,CIOrderArray(i)),75);
end
IQR10N=iqr(MDArray10N(:));
Median10N=median(MDArray10N(:));
Percentile07510N=prctile(MDArray10N(:),75);
Percentile02510N=prctile(MDArray10N(:),25);
for i=1:ThisClusterSize
    if (MDArray10N(i)>Median10N)
        OINot10Array(i)=(MDArray10N(i)-Median10N)/IQR10N;
    else
        OINot10Array(i)=0.01;
    end
end
end

%Calc OI11N
clear MDArray11N;
clear MedianArray11N;
clear P25Array11N;
clear P75Array11N;
clear IQR11N;
clear Median11N;
clear Percentile07511N;
clear Percentile02511N;
clear OINot11Array;
OrderArray=[3,4,5,6];
CIOrderArray=[1,2,3,4];
for i=1:ThisClusterSize

MDArray11N(i)=mahal(ClusterMatrix(i,OrderArray),CIMatrix(:,CIOrderArray
));
end
for i=1:4
    MedianArray11N(i)=median(CIMatrix(:,CIOrderArray(i)));
    P25Array11N(i)=prctile(CIMatrix(:,CIOrderArray(i)),25);
    P75Array11N(i)=prctile(CIMatrix(:,CIOrderArray(i)),75);
end
IQR11N=iqr(MDArray11N(:));
Median11N=median(MDArray11N(:));
Percentile07511N=prctile(MDArray11N(:),75);
Percentile02511N=prctile(MDArray11N(:),25);
for i=1:ThisClusterSize
    if (MDArray11N(i)>Median11N)
        OINot11Array(i)=(MDArray11N(i)-Median11N)/IQR11N;
    else
        OINot11Array(i)=0.01;
    end
end
end

```

```

%Calc OI7
clear mean1;
mean1=mean(ClusterMatrix(:,3));

clear mu1;
clear sigma1;
clear Median1;
clear IQR1;
clear Percentile175;
clear Percentile125;
clear OI7Array;
[mu1,sigma1]=normfit(CIMatrix(:,1));
%Calc OI7 using IQR
IQR1=iqr(CIMatrix(:,1));
Median1=median(CIMatrix(:,1));
Percentile175=prctile(CIMatrix(:,1),75);
Percentile125=prctile(CIMatrix(:,1),25);
for i=1:ThisClusterSize
    if ((CIMatrix(i,1)>Percentile175) ||
(CIMatrix(i,1)<Percentile125)) && (abs(CIMatrix(i,1)-
Median1)>PracticalLimit))
        OI7Array(i)=min(abs(CIMatrix(i,1)-
Median1),abs(CIMatrix(i,1)-Median1))/IQR1;
    else
        OI7Array(i)=0;
    end
end

%Calc OI7N

for i=1:ThisClusterSize
    clear TempMatrix;
    k=1;
    for j=1:ThisClusterSize
        if (j~=i)
            TempMatrix(k,:)=ClusterMatrix(j,4:7);
            k=k+1;
        end
    end
    dArray(i)=mahal(ClusterMatrix(i,4:7),TempMatrix(:,,:));
TempMatrix(ThisClusterSize,:)=ClusterMatrix(ThisClusterSize,4:7);
end
[mu,sigma]=normfit(dArray(:));
for i=1:ThisClusterSize
    if (normcdf(dArray(i),mu,sigma)>0.5)
        ClusterMatrix(i,11)=normcdf(dArray(i),mu,sigma);
    else
        ClusterMatrix(i,11)=normcdf(2*mu-dArray(i),mu,sigma);
    end
end

```

```

        end
    end

    %Calc OI8
    clear mu2;
    clear sigma2;
    clear Median2;
    clear IQR2;
    clear Percentile275;
    clear Percentile225;
    clear OI8Array;
    mean2=mean(ClusterMatrix(:,4));

%
%
%
    %Calc OI8 using IQR
    [mu2,sigma2]=normfit(dArray2(:));
    [mu2,sigma2]=normfit(CIMatrix(:,2));
    %Calc OI7 using IQR
    IQR2=iqr(CIMatrix(:,2));
    Median2=median(CIMatrix(:,2));
    Percentile275=prctile(CIMatrix(:,2),75);
    Percentile225=prctile(CIMatrix(:,2),25);
    for i=1:ThisClusterSize
        if ((CIMatrix(i,2)>Percentile275) ||
(CIMatrix(i,2)<Percentile225)) && (abs(CIMatrix(i,2)-
Median2)>PracticalLimit))
            OI8Array(i)=min(abs(CIMatrix(i,2)-
Median2),abs(CIMatrix(i,2)-Median2))/IQR2;
        else
            OI8Array(i)=0;
        end
    end
end

%Calc OI8N
for i=1:ThisClusterSize
    clear TempMatrix;
    k=1;
    for j=1:ThisClusterSize
        if (j~=i)
            TempMatrix(k,:)=ClusterMatrix(j,[3,5,6,7]);
            k=k+1;
        end
    end
    dArray(i)=mahal(ClusterMatrix(i,[3,5,6,7]),TempMatrix(:,,:));
TempMatrix(ThisClusterSize,:)=ClusterMatrix(ThisClusterSize,[3,5,6,7]);
end
    [mu,sigma]=normfit(dArray(:));
    for i=1:ThisClusterSize

```

```

        if (normcdf(dArray(i),mu,sigma)>0.5)
            ClusterMatrix(i,13)=normcdf(dArray(i),mu,sigma);
        else
            ClusterMatrix(i,13)=normcdf(2*mu-dArray(i),mu,sigma);
        end
    end
end

%OI9
clear mu3;
clear sigma3;
clear Median3;
clear IQR3;
clear Percentile375;
clear Percentile325;
clear OI3Array;
mean3=mean(ClusterMatrix(:,5));

%      %Using Mahalanobis Distance for Each year

[mu3,sigma3]=normfit(CIMatrix(:,3));
%Calc OI7 using IQR
IQR3=iqr(CIMatrix(:,3));
Median3=median(CIMatrix(:,3));
Percentile375=prctile(CIMatrix(:,3),75);
Percentile325=prctile(CIMatrix(:,3),25);
for i=1:ThisClusterSize
    if ((CIMatrix(i,3)>Percentile375) ||
(CIMatrix(i,3)<Percentile325)) && (abs(CIMatrix(i,3)-
Median3)>PracticalLimit))
        OI9Array(i)=min(abs(CIMatrix(i,3)-
Median3),abs(CIMatrix(i,3)-Median3))/IQR3;
    else
        OI9Array(i)=0;
    end
end
end

%Calc OI9N
for i=1:ThisClusterSize
    clear TempMatrix;
    k=1;
    for j=1:ThisClusterSize
        if (j~=i)
            TempMatrix(k,:)=ClusterMatrix(j,[3,4,6,7]);
            k=k+1;
        end
    end
    dArray(i)=mahal(ClusterMatrix(i,[3,4,6,7]),TempMatrix(:,:));
    TempMatrix(ThisClusterSize,:)=ClusterMatrix(ThisClusterSize,[3,4,6,7]);
end
end

```

```

[mu, sigma]=normfit(dArray(:));
for i=1:ThisClusterSize
    if (normcdf(dArray(i), mu, sigma)>0.5)
        ClusterMatrix(i,15)=normcdf(dArray(i), mu, sigma);
    else
        ClusterMatrix(i,15)=normcdf(2*mu-dArray(i), mu, sigma);
    end
end

%Calc OI10
clear mu4;
clear sigma4;
clear Median4;
clear IQR4;
clear Percentile475;
clear Percentile425;
clear OI4Array;
mean4=mean(ClusterMatrix(:,6));

%      %Using Mahalanobis Distance for Each year

%Calc OI10 using IQR
[mu4, sigma4]=normfit(CIMatrix(:,4));
%Calc OI7 using IQR
IQR4=iqr(CIMatrix(:,4));
Median4=median(CIMatrix(:,4));
Percentile475=prctile(CIMatrix(:,4),75);
Percentile425=prctile(CIMatrix(:,4),25);
for i=1:ThisClusterSize
    if (((CIMatrix(i,4)>Percentile475) ||
(CIMatrix(i,4)<Percentile425)) && (abs(CIMatrix(i,4)-
Median4)>PracticalLimit))
        OI10Array(i)=min(abs(CIMatrix(i,4)-
Median4), abs(CIMatrix(i,4)-Median4))/IQR4;
    else
        OI10Array(i)=0;
    end
end

%Calc OI10N
for i=1:ThisClusterSize
    clear TempMatrix;
    k=1;
    for j=1:ThisClusterSize
        if (j~=i)
            TempMatrix(k,:)=ClusterMatrix(j,[3,4,5,7]);
            k=k+1;
        end
    end
end

```



```

        dArray(i)=mahal(ClusterMatrix(i,[3,4,5,7]),TempMatrix(:,:));
TempMatrix(ThisClusterSize,:)=ClusterMatrix(ThisClusterSize,[3,4,5,7]);
end
[mu,sigma]=normfit(dArray(:));
for i=1:ThisClusterSize
    if (normcdf(dArray(i),mu,sigma)>0.5)
        ClusterMatrix(i,17)=normcdf(dArray(i),mu,sigma);
    else
        ClusterMatrix(i,17)=normcdf(2*mu-dArray(i),mu,sigma);
    end
end

end

%Calc OI11
clear mu5;
clear sigma5;
clear Median5;
clear IQR5;
clear Percentile575;
clear Percentile525;
clear O5Array;
mean5=mean(ClusterMatrix(:,7));

[mu5,sigma5]=normfit(CIMatrix(:,5));
%Calc OI7 using IQR
IQR5=iqr(CIMatrix(:,5));
Median5=median(CIMatrix(:,5));
Percentile575=prctile(CIMatrix(:,5),75);
Percentile525=prctile(CIMatrix(:,5),25);
for i=1:ThisClusterSize
    if ((CIMatrix(i,5)>Percentile575) ||
(CIMatrix(i,5)<Percentile525)) && (abs(CIMatrix(i,5)-
Median5)>PracticalLimit))
        OI11Array(i)=min(abs(CIMatrix(i,5)-
Median5),abs(CIMatrix(i,5)-Median5))/IQR5;
    else
        OI11Array(i)=0;
    end
end

end

%Calc OI11N
for i=1:ThisClusterSize
    clear TempMatrix;
    k=1;
    for j=1:ThisClusterSize
        if (j~=i)

```

```

        TempMatrix(k,:) = ClusterMatrix(j, 3:6);
        k=k+1;
    end
end
dArray(i) = mahal(ClusterMatrix(i, 3:6), TempMatrix(:, :));

TempMatrix(ThisClusterSize, :) = ClusterMatrix(ThisClusterSize, 3:6);
end
[mu, sigma] = normfit(dArray(:));
for i=1:ThisClusterSize
    if (normcdf(dArray(i), mu, sigma) > 0.5)
        ClusterMatrix(i, 19) = normcdf(dArray(i), mu, sigma);
    else
        ClusterMatrix(i, 19) = normcdf(2*mu - dArray(i), mu, sigma);
    end
end

%
    ClusterMatrix(i, 9) = OIArray(i);    %OI
ClusterMatrix(i, 9) = 0;    %OI
ClusterMatrix(i, 10) = OI7Array(i);    %OI7
ClusterMatrix(i, 11) = OINot7Array(i); %OINot7
ClusterMatrix(i, 12) = OI8Array(i);    %OI8
ClusterMatrix(i, 13) = OINot8Array(i); %OINot8
ClusterMatrix(i, 14) = OI9Array(i);    %OI9
ClusterMatrix(i, 15) = OINot9Array(i); %OINot9
ClusterMatrix(i, 16) = OI10Array(i);   %OI10
ClusterMatrix(i, 17) = OINot10Array(i); %OINot10
ClusterMatrix(i, 18) = OI11Array(i);   %OI11
ClusterMatrix(i, 19) = OINot11Array(i); %OINot11

ClusterMatrix(i, 20) = ClusterMatrix(i, 10) /
ClusterMatrix(i, 11); %Calc OI7F
ClusterMatrix(i, 21) = ClusterMatrix(i, 12) /
ClusterMatrix(i, 13); %Calc OI8F
ClusterMatrix(i, 22) = ClusterMatrix(i, 14) /
ClusterMatrix(i, 15); %Calc OI9F
ClusterMatrix(i, 23) = ClusterMatrix(i, 16) /
ClusterMatrix(i, 17); %Calc OI10F
ClusterMatrix(i, 24) = ClusterMatrix(i, 18) /
ClusterMatrix(i, 19); %Calc OI11F

SecIsOutlier = 'false';
PointIsOutlier = 'false';
OutYearNum = 0;
if (ClusterMatrix(i, 10) < Alpha)

else
    if (OINot7Array(i) < Beta)
        OutYearNum = OutYearNum + 1;
    else

```

```

                SecIsOutlier='true';           %The section is truly
dissimilar        ClusterMatrix(i,9)=1;
                end
            end
            if (ClusterMatrix(i,12)<Alpha)
        else
            if (OINot8Array(i)<Beta)
                OutYearNum=OutYearNum+1;
            else
                SecIsOutlier='true';           %The section is truly
dissimilar        ClusterMatrix(i,9)=1;
                end
            end
            if (ClusterMatrix(i,14)<Alpha)
        else
            if (OINot9Array(i)<Beta)
                OutYearNum=OutYearNum+1;
            else
                SecIsOutlier='true';           %The section is truly
dissimilar        ClusterMatrix(i,9)=1;
                end
            end
            if (ClusterMatrix(i,16)<Alpha)
        else
            if (OINot10Array(i)<Beta)
                OutYearNum=OutYearNum+1;
            else
                SecIsOutlier='true';           %The section is truly
dissimilar        ClusterMatrix(i,9)=1;
                end
            end
            if (ClusterMatrix(i,18)<Alpha)
        else
            if (OINot11Array(i)<Beta)
                OutYearNum=OutYearNum+1;
            else
                SecIsOutlier='true';           %The section is truly
dissimilar        ClusterMatrix(i,9)=1;
                end
            end
            if (OutYearNum>1)
                ClusterMatrix(i,9)=1;
            end
        end
    end
end

```

```

ThisHighwayMatrixSize = size(HighwayMatrix,1);
%Record OIs to HighwayMatrix
for ThisSec=1:ThisClusterSize
    for ThisSec2=1:ThisHighwayMatrixSize
        if (HighwayMatrix(ThisSec2,1)==ClusterMatrix(ThisSec,1))
            HighwayMatrix(ThisSec2,9)=ClusterMatrix(ThisSec,9);
%OI;
            HighwayMatrix(ThisSec2,10)=ClusterMatrix(ThisSec,10);
%OI7;
            HighwayMatrix(ThisSec2,11)=ClusterMatrix(ThisSec,11);
%OI7N;
            HighwayMatrix(ThisSec2,12)=ClusterMatrix(ThisSec,12);
%OI8;
            HighwayMatrix(ThisSec2,13)=ClusterMatrix(ThisSec,13);
%OI8N;
            HighwayMatrix(ThisSec2,14)=ClusterMatrix(ThisSec,14);
%OI9;
            HighwayMatrix(ThisSec2,15)=ClusterMatrix(ThisSec,15);
%OI9N;
            HighwayMatrix(ThisSec2,16)=ClusterMatrix(ThisSec,16);
%OI10;
            HighwayMatrix(ThisSec2,17)=ClusterMatrix(ThisSec,17);
%OI10N;
            HighwayMatrix(ThisSec2,18)=ClusterMatrix(ThisSec,18);
%OI11;
            HighwayMatrix(ThisSec2,19)=ClusterMatrix(ThisSec,19);
%OI11N;
            HighwayMatrix(ThisSec2,20)=ClusterMatrix(ThisSec,20);
%OI7F;
            HighwayMatrix(ThisSec2,21)=ClusterMatrix(ThisSec,21);
%OI8F;
            HighwayMatrix(ThisSec2,22)=ClusterMatrix(ThisSec,22);
%OI9F;
            HighwayMatrix(ThisSec2,23)=ClusterMatrix(ThisSec,23);
%OI10F;
            HighwayMatrix(ThisSec2,24)=ClusterMatrix(ThisSec,24);
%OI11F;
        end
    end

end

figure1 = figure('Visible','off');
axes1 =
axes('Parent',figure1,'YTickLabel',{'50','60','70','80','90','100'},'YTick',
[50,60,70,80,90,100],'XTickLabel',{'2007','2008','2009','2010','2011'},...
'XTick',[1 2 3 4 5],'fontsize',20);

XData1=1:5;
ylim([50,100.1]);
ylabel('Distress Score');
for i=1:ThisClusterSize

```

```

ClusterMeanDS2007=mean(ClusterMatrix(1:ThisClusterSize,3));
ClusterMeanDS2008=mean(ClusterMatrix(1:ThisClusterSize,4));
ClusterMeanDS2009=mean(ClusterMatrix(1:ThisClusterSize,5));
ClusterMeanDS2010=mean(ClusterMatrix(1:ThisClusterSize,6));
ClusterMeanDS2011=mean(ClusterMatrix(1:ThisClusterSize,7));

MClusterMeanArray=[ClusterMeanDS2007,ClusterMeanDS2008,ClusterMeanDS2009,ClusterMeanDS2010,ClusterMeanDS2011];
ThisSecMeanDist=1/5*(abs(ClusterMeanDS2007-ClusterMatrix(i,3))+abs(ClusterMeanDS2008-ClusterMatrix(i,4))+abs(ClusterMeanDS2009-ClusterMatrix(i,5))+abs(ClusterMeanDS2010-ClusterMatrix(i,6))+abs(ClusterMeanDS2011-ClusterMatrix(i,7)));

line(1:5,MedianArray(:),'Parent',axes1,'Tag','coords','Color',[0 0 0],'LineWidth',MyLineWidth,'LineStyle','--'); %Median line in blue

line(1:5,P25Array(:),'Parent',axes1,'Tag','coords','Color',[0 0 0],'LineWidth',MyLineWidth,'LineStyle','--'); %Median line in blue

line(1:5,P75Array(:),'Parent',axes1,'Tag','coords','Color',[0 0 0],'LineWidth',MyLineWidth,'LineStyle','--'); %Median line in blue

OINMax=max([ClusterMatrix(i,10),ClusterMatrix(i,12),ClusterMatrix(i,14),ClusterMatrix(i,16),ClusterMatrix(i,18)]);
if (ClusterMatrix(i,10)==OINMax)
    MaxYear=1;
elseif (ClusterMatrix(i,12)==OINMax)
    MaxYear=2;
elseif (ClusterMatrix(i,14)==OINMax)
    MaxYear=3;
elseif (ClusterMatrix(i,16)==OINMax)
    MaxYear=4;
elseif (ClusterMatrix(i,18)==OINMax)
    MaxYear=5;
end

MoreThan2Years=false;
SortedOIArray=sort([ClusterMatrix(i,10),ClusterMatrix(i,12),ClusterMatrix(i,14),ClusterMatrix(i,16),ClusterMatrix(i,18)]);
OImax1=SortedOIArray(5);
OImax2=SortedOIArray(4);

if (OImax1>Beta && OImax2>Beta)
    MoreThan2Years=true;
end

% if ((ClusterMatrix(i,9)>Beta && ThisSecMeanDist>PracticalLimit)|| MDArray(i)>Theta || MoreThan2Years==true) %the section is outlier

```

```

%           if ((MaxYear==1) && (ClusterMatrix(i,10)>Alpha) &&
(OINot7Array(i)<Beta) && MoreThan2Years==false)
%
line(XData1,ClusterMatrix(i,3:7),'Parent',axes1,'Tag','coords','Color',
[0 0 0],'LineWidth',MyLineWidth); %Regular Sections in Blue
%           if (abs(ClusterMeanDS2007-
ClusterMatrix(i,3))>PracticalLimit)
%
line(1:1,ClusterMatrix(i,3:3),'Parent',axes1,'Tag','coords','Color',[1
0 0],'MarkerFaceColor',[1 0 0],...
%           'MarkerSize',10,'Marker','o','Color',[1 0 0]);
%Red: Erronous
%           end
%           elseif ((MaxYear==2) && (ClusterMatrix(i,12)>Alpha) &&
(OINot8Array(i)<Beta) && MoreThan2Years==false)
%
line(XData1,ClusterMatrix(i,3:7),'Parent',axes1,'Tag','coords','Color',
[0 0 0],'LineWidth',MyLineWidth); %Regular Sections in Blue
%           if (abs(ClusterMeanDS2007-
ClusterMatrix(i,4))>PracticalLimit)
%
line(2:2,ClusterMatrix(i,4:4),'Parent',axes1,'Tag','coords','Color',[1
0 0],'MarkerFaceColor',[1 0 0],...
%           'MarkerSize',10,'Marker','o','Color',[1 0 0]);
%Red: Erronous
%           end
%           elseif ((MaxYear==3) && (ClusterMatrix(i,14)>Alpha) &&
(OINot9Array(i)<Beta) && MoreThan2Years==false)
%
line(XData1,ClusterMatrix(i,3:7),'Parent',axes1,'Tag','coords','Color',
[0 0 0],'LineWidth',MyLineWidth); %Regular Sections in Blue
%           if (abs(ClusterMeanDS2007-
ClusterMatrix(i,5))>PracticalLimit)
%
line(3:3,ClusterMatrix(i,5:5),'Parent',axes1,'Tag','coords','Color',[1
0 0],'MarkerFaceColor',[1 0 0],...
%           'MarkerSize',10,'Marker','o','Color',[1 0 0]);
%Red: Erronous
%           end
%           elseif ((MaxYear==4) && (ClusterMatrix(i,16)>Alpha) &&
(OINot10Array(i)<Beta) && MoreThan2Years==false)
%
line(XData1,ClusterMatrix(i,3:7),'Parent',axes1,'Tag','coords','Color',
[0 0 0],'LineWidth',MyLineWidth); %Regular Sections in Blue
%           if (abs(ClusterMeanDS2007-
ClusterMatrix(i,6))>PracticalLimit)
%
line(4:4,ClusterMatrix(i,6:6),'Parent',axes1,'Tag','coords','Color',[1
0 0],'MarkerFaceColor',[1 0 0],...
%           'MarkerSize',10,'Marker','o','Color',[1 0 0]);
%Red: Erronous
%           end
%

```

```

%           elseif ((MaxYear==5) && (ClusterMatrix(i,18)>Alpha) &&
(OINot11Array(i)<Beta) && MoreThan2Years==false)
%
line(XData1,ClusterMatrix(i,3:7),'Parent',axes1,'Tag','coords','Color',
[0 0 0],'LineWidth',MyLineWidth); %Regular Sections in Blue
%           if (abs(ClusterMeanDS2007-
ClusterMatrix(i,7))>PracticalLimit)
%
line(5:5,ClusterMatrix(i,7:7),'Parent',axes1,'Tag','coords','Color',[1
0 0],'MarkerFaceColor',[1 0 0],...
%           'MarkerSize',10,'Marker','o','Color',[1 0 0]);
%Red: Erronous
%           end
%           else
%
line(1:5,ClusterMatrix(i,3:7),'Parent',axes1,'Tag','coords','Color',[1
0 0],'LineWidth',MyLineWidth); %Outlier Sections in Red
%           end
%           else
%
line(XData1,ClusterMatrix(i,3:7),'Parent',axes1,'Tag','coords','Color',
[0 0 0],'LineWidth',MyLineWidth); %Regular Sections in Blue
%
%           %Outlier Point
%           if (ClusterMatrix(i,10)>Alpha &&
abs(ClusterMeanDS2007-ClusterMatrix(i,3))>PracticalLimit)
%
line(1:1,ClusterMatrix(i,3:3),'Parent',axes1,'Tag','coords','Color',[1
0 0],'MarkerFaceColor',[1 0 0],...
%           'MarkerSize',10,'Marker','o','Color',[1 0 0]);
%Red: Erronous
%           end
%           if (ClusterMatrix(i,12)>Alpha &&
abs(ClusterMeanDS2008-ClusterMatrix(i,4))>PracticalLimit)
%
line(2:2,ClusterMatrix(i,4:4),'Parent',axes1,'Tag','coords','Color',[1
0 0],'MarkerFaceColor',[1 0 0],...
%           'MarkerSize',10,'Marker','o','Color',[1 0 0]);
%Red: Erronous
%           end
%           if (ClusterMatrix(i,14)>Alpha &&
abs(ClusterMeanDS2009-ClusterMatrix(i,5))>PracticalLimit)
%
line(3:3,ClusterMatrix(i,5:5),'Parent',axes1,'Tag','coords','Color',[1
0 0],'MarkerFaceColor',[1 0 0],...
%           'MarkerSize',10,'Marker','o','Color',[1 0 0]);
%Red: Erronous
%           end
%           if (ClusterMatrix(i,16)>Alpha &&
abs(ClusterMeanDS2010-ClusterMatrix(i,6))>PracticalLimit)
%
line(4:4,ClusterMatrix(i,6:6),'Parent',axes1,'Tag','coords','Color',[1
0 0],'MarkerFaceColor',[1 0 0],...

```

```

%                               'MarkerSize',10,'Marker','o','Color',[1 0 0]);
%Red: Erronous
%                               end
%                               if (ClusterMatrix(i,18)>Alpha &&
abs(ClusterMeanDS2011-ClusterMatrix(i,7)>PracticalLimit)
%
line(5:5,ClusterMatrix(i,7:7),'Parent',axes1,'Tag','coords','Color',[1
0 0],'MarkerFaceColor',[1 0 0],...
%                               'MarkerSize',10,'Marker','o','Color',[1 0 0]);
%Red: Erronous
%                               end
%                               end

                                if (ClusterMatrix(i,9)==1)

line(1:5,ClusterMatrix(i,3:7),'Parent',axes1,'Tag','coords','Color',[1
0 0],'LineWidth',MyLineWidth); %Outlier section
                                else

line(XData1,ClusterMatrix(i,3:7),'Parent',axes1,'Tag','coords','Color',
[0 0 0],'LineWidth',MyLineWidth); %Normal Section
                                if (ClusterMatrix(i,10)>Alpha)

line(1:1,ClusterMatrix(i,3:3),'Parent',axes1,'Tag','coords','Color',[1
0 0],'MarkerFaceColor',[1 0 0],...
                                'MarkerSize',10,'Marker','o','Color',[1 0 0]);
%Red: Erronous
                                elseif (ClusterMatrix(i,12)>Alpha)

line(2:2,ClusterMatrix(i,4:4),'Parent',axes1,'Tag','coords','Color',[1
0 0],'MarkerFaceColor',[1 0 0],...
                                'MarkerSize',10,'Marker','o','Color',[1 0 0]);
%Red: Erronous
                                elseif (ClusterMatrix(i,14)>Alpha)

line(3:3,ClusterMatrix(i,5:5),'Parent',axes1,'Tag','coords','Color',[1
0 0],'MarkerFaceColor',[1 0 0],...
                                'MarkerSize',10,'Marker','o','Color',[1 0 0]);
%Red: Erronouss
                                elseif (ClusterMatrix(i,16)>Alpha)

line(4:4,ClusterMatrix(i,6:6),'Parent',axes1,'Tag','coords','Color',[1
0 0],'MarkerFaceColor',[1 0 0],...
                                'MarkerSize',10,'Marker','o','Color',[1 0 0]);
%Red: Erronou
                                elseif (ClusterMatrix(i,18)>Alpha)

line(5:5,ClusterMatrix(i,7:7),'Parent',axes1,'Tag','coords','Color',[1
0 0],'MarkerFaceColor',[1 0 0],...

```



```

                                'MarkerSize',10,'Marker','o','Color',[1 0 0]);
%Red: Erronous
                                end
                                end

%
%
%           if (ClusterMatrix(i,17)>Beta &&
ThisSecMeanDist>PracticalLimit) %the section is outlier
%           if ((MaxYear==1) && (ClusterMatrix(i,10)>Alpha) &&
(OINot7Array(i)<Beta) && MoreThan2Years==false)
%
% line(XData1,ClusterMatrix(i,3:7),'Parent',axes1,'Tag','coords','Color',
[0 0 0],'LineWidth',MyLineWidth); %Regular Sections in Blue
%           if (abs(ClusterMeanDS2007-
ClusterMatrix(i,3))>PracticalLimit)
%
% line(1:1,ClusterMatrix(i,3:3),'Parent',axes1,'Tag','coords','Color',[1
0 0],'MarkerFaceColor',[1 0 0],...
%           'MarkerSize',10,'Marker','o','Color',[1 0 0]);
%Red: Erronous
%           end
%           elseif ((MaxYear==2) && (ClusterMatrix(i,12)>Alpha) &&
(OINot8Array(i)<Beta) && MoreThan2Years==false)
%
% line(XData1,ClusterMatrix(i,3:7),'Parent',axes1,'Tag','coords','Color',
[0 0 0],'LineWidth',MyLineWidth); %Regular Sections in Blue
%           if (abs(ClusterMeanDS2007-
ClusterMatrix(i,4))>PracticalLimit)
%
% line(2:2,ClusterMatrix(i,4:4),'Parent',axes1,'Tag','coords','Color',[1
0 0],'MarkerFaceColor',[1 0 0],...
%           'MarkerSize',10,'Marker','o','Color',[1 0 0]);
%Red: Erronous
%           end
%           elseif ((MaxYear==3) && (ClusterMatrix(i,14)>Alpha) &&
(OINot9Array(i)<Beta) && MoreThan2Years==false)
%
% line(XData1,ClusterMatrix(i,3:7),'Parent',axes1,'Tag','coords','Color',
[0 0 0],'LineWidth',MyLineWidth); %Regular Sections in Blue
%           if (abs(ClusterMeanDS2007-
ClusterMatrix(i,5))>PracticalLimit)
%
% line(3:3,ClusterMatrix(i,5:5),'Parent',axes1,'Tag','coords','Color',[1
0 0],'MarkerFaceColor',[1 0 0],...
%           'MarkerSize',10,'Marker','o','Color',[1 0 0]);
%Red: Erronous
%           end
%

```

```

%           elseif ((MaxYear==4) && (ClusterMatrix(i,16)>Alpha) &&
(OINot10Array(i)<Beta) && MoreThan2Years==false)
%
line(XData1,ClusterMatrix(i,3:7),'Parent',axes1,'Tag','coords','Color',
[0 0 0],'LineWidth',MyLineWidth); %Regular Sections in Blue
%           if (abs(ClusterMeanDS2007-
ClusterMatrix(i,6))>PracticalLimit)
%
line(4:4,ClusterMatrix(i,6:6),'Parent',axes1,'Tag','coords','Color',[1
0 0],'MarkerFaceColor',[1 0 0],...
%           'MarkerSize',10,'Marker','o','Color',[1 0 0]);
%Red: Erronous
%           end
%           elseif ((MaxYear==5) && (ClusterMatrix(i,18)>Alpha) &&
(OINot11Array(i)<Beta) && MoreThan2Years==false)
%
line(XData1,ClusterMatrix(i,3:7),'Parent',axes1,'Tag','coords','Color',
[0 0 0],'LineWidth',MyLineWidth); %Regular Sections in Blue
%           if (abs(ClusterMeanDS2007-
ClusterMatrix(i,7))>PracticalLimit)
%
line(5:5,ClusterMatrix(i,7:7),'Parent',axes1,'Tag','coords','Color',[1
0 0],'MarkerFaceColor',[1 0 0],...
%           'MarkerSize',10,'Marker','o','Color',[1 0 0]);
%Red: Erronous
%           end
%           else
%
line(1:5,ClusterMatrix(i,3:7),'Parent',axes1,'Tag','coords','Color',[1
0 0],'LineWidth',MyLineWidth); %Outlier Sections in Red
%           end
%           else
%
line(XData1,ClusterMatrix(i,3:7),'Parent',axes1,'Tag','coords','Color',
[0 0 0],'LineWidth',MyLineWidth); %Regular Sections in Blue
%
%           %Outlier Point
%           if (ClusterMatrix(i,10)>Alpha &&
abs(ClusterMeanDS2007-ClusterMatrix(i,3))>PracticalLimit)
%
line(1:1,ClusterMatrix(i,3:3),'Parent',axes1,'Tag','coords','Color',[1
0 0],'MarkerFaceColor',[1 0 0],...
%           'MarkerSize',10,'Marker','o','Color',[1 0 0]);
%Red: Erronous
%           end
%           if (ClusterMatrix(i,12)>Alpha &&
abs(ClusterMeanDS2008-ClusterMatrix(i,4))>PracticalLimit)
%
line(2:2,ClusterMatrix(i,4:4),'Parent',axes1,'Tag','coords','Color',[1
0 0],'MarkerFaceColor',[1 0 0],...
%           'MarkerSize',10,'Marker','o','Color',[1 0 0]);
%Red: Erronous
%           end

```

```

%           if (ClusterMatrix(i,14)>Alpha &&
abs(ClusterMeanDS2009-ClusterMatrix(i,5))>PracticalLimit)
%
line(3:3,ClusterMatrix(i,5:5),'Parent',axes1,'Tag','coords','Color',[1
0 0],'MarkerFaceColor',[1 0 0],...
%           'MarkerSize',10,'Marker','o','Color',[1 0 0]);
%Red: Erronous
%           end
%           if (ClusterMatrix(i,16)>Alpha &&
abs(ClusterMeanDS2010-ClusterMatrix(i,6))>PracticalLimit)
%
line(4:4,ClusterMatrix(i,6:6),'Parent',axes1,'Tag','coords','Color',[1
0 0],'MarkerFaceColor',[1 0 0],...
%           'MarkerSize',10,'Marker','o','Color',[1 0 0]);
%Red: Erronous
%           end
%           if (ClusterMatrix(i,18)>Alpha &&
abs(ClusterMeanDS2011-ClusterMatrix(i,7))>PracticalLimit)
%
line(5:5,ClusterMatrix(i,7:7),'Parent',axes1,'Tag','coords','Color',[1
0 0],'MarkerFaceColor',[1 0 0],...
%           'MarkerSize',10,'Marker','o','Color',[1 0 0]);
%Red: Erronous
%           end
%           end

end
title( strcat('cluster num ',num2str(ThisClNum),' in highway
',ThisHighwayName, ' , including ',num2str(ThisClusterSize), '
sections' ));
xlabel('Year');
ylabel('DS');
FileString= char(strcat(FiguresAddress,'\ ',ThisHighwayName,'-
',num2str(ThisClNum),'.jpg'));
saveas(figure1,FileString);
close(figure1);
end
end

%Save Cluster Number
j=1;
SortedClusterHighwayMatrix=sortrows(HighwayMatrix,8);
NCj=1;
SecNum=size(HighwayMatrix,1);
for i=1:SecNum-1
    if
(SortedClusterHighwayMatrix(i,8)==SortedClusterHighwayMatrix(i+1,8))
        NCj=NCj+1;
    else
        ClusterLenghtsMax(j)=NCj;
        j=j+1;

```

```

        NCj=1;
    end

end

if (j==1)
    ClusterLengthsMax(j)=SecNum;
end

clear SortedThisHighwayMatrix;
CumNum=1;
SortedThisHighwayMatrix=sortrows(HighwayMatrix,2);
% [a,NC]=size(ClusterLengthsMax);
NC=size(ClusterLengthsMax,2);
for j=1:NC
    OutliernessIndex=0;
    %ClusterSecNums=size(HighwayMatrix(idx==j),1);
    ThisClusterSecNums=ClusterLengthsMax(j);
    if (ThisClusterSecNums>0)
        clear OutliernessIndex;
        %Calculate OI
        %OutliernessIndex=ThisHwSecNum/ClusterSecNums;
        ThisClusterNumber=j;
        if (ThisClusterSecNums>1)
            clear ThisClusterArray;
            for k=1:ThisClusterSecNums
                % HighwayMatrix(idx==j,7)=ThisClusterNumber;
                HighwayMatrix(idx==j,7)=ThisClusterNumber;
            end
        end
        SortedThisHighwayMatrix(CumNum,8)=ThisClusterNumber;
        ThisClusterArray(k,:)=SortedThisHighwayMatrix(CumNum,3:7);
        CumNum=CumNum+1;
    end
end

end

end

%All in One Figure
ClNumForChart=SortedThisHighwayMatrix(:,8);
labels = {'2007','2008','2009','2010','2011'};
h=figure('Visible','off');

parallelcoords(SortedThisHighwayMatrix(:,3:7),'group',ClNumForChart,'la
bels',labels);
title('All Highways');
FileString= char(strcat(FiguresAddress,'\ ',ThisHighwayName,'-
','All',' .jpg'));
saveas(h,FileString);
close(h);
end

%Save Cluster Number in the database .IO field

```

```

for k=1:TotalSecNum
    if (strcmp(NetworkData(k).SIGNED_HIG,ThisHighwayName))
        %NetworkData(k).OI=OIByObjID(NetworkData(k).OBJECTID,
SortedThisHighwayMatrix );
        [ ClNum, OI, OI7,
OI7N,OI8,OI8N,OI9,OI9N,OI10,OI10N,OI11,OI11N,OI7F,OI8F,OI9F,OI10F,OI11F
] = ClNumByObjID(NetworkData(k).OBJECTID, SortedThisHighwayMatrix );
        NetworkData(k).ClNum=ClNum;
        NetworkData(k).OI=OI;
        NetworkData(k).OI7=OI7;
        NetworkData(k).OI7N=OI7N;
        NetworkData(k).OI8=OI8;
        NetworkData(k).OI8N=OI8N;
        NetworkData(k).OI9=OI9;
        NetworkData(k).OI9N=OI9N;
        NetworkData(k).OI10=OI10;
        NetworkData(k).OI10N=OI10N;
        NetworkData(k).OI11=OI11;
        NetworkData(k).OI11N=OI11N;
        NetworkData(k).OI7F=OI7F;
        NetworkData(k).OI8F=OI8F;
        NetworkData(k).OI9F=OI9F;
        NetworkData(k).OI10F=OI10F;
        NetworkData(k).OI11F=OI11F;
        if (OI11<Alpha)
            NetworkData(k).IsOI11=0;    %The point is not outlier
        else
%           if (OI11N<Beta)
%           if (OI==0)
                NetworkData(k).IsOI11=1;    %The point is likely
erronous
            else
                NetworkData(k).IsOI11=2;    %The section is truly
dissimilar
            end
        end

        if (OI10<Alpha)
            NetworkData(k).IsOI10=0;    %The point is not outlier
        else
%           if (OI10N<Beta)
%           if (OI==0)
                NetworkData(k).IsOI10=1;    %The point is likely
erronous
            else
                NetworkData(k).IsOI10=2;    %The section is truly
dissimilar
            end
        end

        if (OI9<Alpha)
            NetworkData(k).IsOI9=0;    %The point is not outlier
        else

```

```

%           if (OI10N<Beta)
%           if (OI==0)
erroneous       NetworkData(k).IsOI9=1;    %The point is likely
%           else
dissimilar     NetworkData(k).IsOI9=2;    %The section is truly
%           end
%           end
%           if (OI8<Alpha)
%           NetworkData(k).IsOI8=0;    %The point is not outlier
%           else
%           if (OI10N<Beta)
%           if (OI==0)
erroneous       NetworkData(k).IsOI8=1;    %The point is likely
%           else
dissimilar     NetworkData(k).IsOI8=2;    %The section is truly
%           end
%           end
%           if (OI7<Alpha)
%           NetworkData(k).IsOI7=0;    %The point is not outlier
%           else
%           if (OI10N<Beta)
%           if (OI==0)
erroneous       NetworkData(k).IsOI7=1;    %The point is likely
%           else
dissimilar     NetworkData(k).IsOI7=2;    %The section is truly
%           end
%           end
%           end
%           end

    100*HN/HighwayNum;
end

end
shapewrite(NetworkData,OIDetectedAddress);
shapeinfo(OIDetectedAddress)

a='Done!'

```

```
function ProbByNum=ProbByNum(SecNumDiff)

%mu=1.6790;
%sigma=0.9357;
%ProbByNum = lognpdf(SecNumDiff,mu,sigma);
B = 9.25462;
ProbByNum=raylcdf(SecNumDiff+0.5,B)-raylcdf(SecNumDiff-0.5,B);
```

```

function
[a]=RankSections(ShapefileAddress,HighwayNames,RankedSHPAAddress,DistLimit)
NetworkData=shaperead(ShapefileAddress);%,'Selector',{@(v1) (v1~='L')
,'SIGNED_RDB'});
HWNArray=size(HighwayNames);
HighwayNum=HWNArray(1,2);
TotalSecNum=size(NetworkData,1);
MaxObjID=0;
for i=1:TotalSecNum;
    ThisObjID=NetworkData(i).OBJECTID;
    if MaxObjID<ThisObjID
        MaxObjID=ThisObjID;
    end
end
for i=1:MaxObjID;
    TotalArray(i)=0;
end
for HN=1:HighwayNum
    ThisRank=1;
    ThisHighwayName=HighwayNames(HN);
    ThisHwSecNum=0;
    clear RankHWMMatrix;
    for ii=1:TotalSecNum
        if (strcmp(NetworkData(ii).SIGNED_HIG,ThisHighwayName))
            ThisHwSecNum=ThisHwSecNum+1;
            RankHWMMatrix(ThisHwSecNum,1)=NetworkData(ii).OBJECTID;
            RankHWMMatrix(ThisHwSecNum,2)=NetworkData(ii).XCoord;
            RankHWMMatrix(ThisHwSecNum,3)=NetworkData(ii).YCoord;
            RankHWMMatrix(ThisHwSecNum,4)=NetworkData(ii).BRMND_DFO;

RankHWMMatrix(ThisHwSecNum,5)=ConvertBRMTextToNum(NetworkData(ii).BRM);
        end
    end

    if (ThisHwSecNum>=6)

        clear SortedMatrixByX;
        SortedMatrixByX=sortrows(RankHWMMatrix,2); %Sorts from left to the
right of the map
        EndNodeNum=0;
        OldID=0;
        OldID2=0;
        OldID3=0;
        OldID4=0;
        OldID5=0;
        OldID6=0;
        OldID7=0;
        OldID8=0;
        OldID9=0;
        OldID10=0;
        OldID11=0;

```



```

OldID12=0;
OldID13=0;
OldID14=0;
OldID15=0;
clear EndNodeArray;
for i=1:ThisHwSecNum    %primary section = starting point
    OldID=0;
    OldID2=0;
    OldID3=0;
    OldID4=0;
    OldID5=0;
    OldID6=0;
    OldID7=0;
    OldID8=0;
    OldID9=0;
    OldID10=0;
    OldID11=0;
    OldID12=0;
    OldID13=0;
    OldID14=0;
    OldID15=0;

    if (HN==188 || HN==195)
        i
    end
    ThisPrimSecID=SortedMatrixByX(i,1);
    ThisPrimSecX=SortedMatrixByX(i,2);
    ThisPrimSecY=SortedMatrixByX(i,3);
    ThisPrimSecDFO=SortedMatrixByX(i,4);
    ThisPrimSecBRM=SortedMatrixByX(i,5);
    j=0;
    NNeighbor=0;
    ThisSectionID=ThisPrimSecID;
    ThisSectionX=ThisPrimSecX;
    ThisSectionY=ThisPrimSecY;
    ThisSectionDFO=ThisPrimSecDFO;
    ThisSectionBRM=ThisPrimSecBRM;
    NoSecFound=0;
    while (NoSecFound==0)
        NNeighbor=0;
        MinDist=9999;
        for j=1:ThisHwSecNum
            ThisNeighborID=SortedMatrixByX(j,1);
            if (ThisNeighborID~=ThisSectionID    &&
ThisNeighborID~=OldID && ThisNeighborID~=OldID2 &&
ThisNeighborID~=OldID3...
                && ThisNeighborID~=OldID4 &&
ThisNeighborID~=OldID5 && ThisNeighborID~=OldID6 &&
ThisNeighborID~=OldID7...
                && ThisNeighborID~=OldID8 &&
ThisNeighborID~=OldID9 && ThisNeighborID~=OldID10 &&
ThisNeighborID~=OldID11...

```

```

                && ThisNeighborID~=OldID12 &&
ThisNeighborID~=OldID13 && ThisNeighborID~=OldID14 &&
ThisNeighborID~=OldID15)
                ThisNeighborX=SortedMatrixByX(j,2);
                ThisNeighborY=SortedMatrixByX(j,3);
                ThisNeighborDFO=SortedMatrixByX(j,4);
                ThisNeighborBRM=SortedMatrixByX(j,5);
                ThisDist=sqrt((ThisNeighborX-
ThisSectionX)^2+(ThisNeighborY-ThisSectionY)^2);
                %if ((ThisNeighborX>ThisSectionX ||
ThisNeighborY<ThisSectionY) && (ThisDist<DistLimit ||
ThisNeighborDFO==ThisSectionDFO))
                if ((ThisDist<DistLimit ||
ThisNeighborDFO==ThisSectionDFO || ThisNeighborBRM==ThisSectionBRM))
                NNeighbor=NNeighbor+1;
                if (ThisNeighborDFO==ThisSectionDFO)
                MinDist=0;
                MinDistID=ThisNeighborID;
                MinDistX=ThisNeighborX;
                MinDistY=ThisNeighborY;
                MinDistDFO=ThisNeighborDFO;
                MinDistBRM=ThisNeighborBRM;
                elseif (ThisDist<MinDist)
                MinDist=ThisDist;
                MinDistID=ThisNeighborID;
                MinDistX=ThisNeighborX;
                MinDistY=ThisNeighborY;
                MinDistDFO=ThisNeighborDFO;
                MinDistBRM=ThisNeighborBRM;
                end
                end
                end
end
if (NNeighbor==0)
    NoSecFound=1;
    EndNodeNum=EndNodeNum+1;
    EndNodeArray(EndNodeNum,1)=ThisSectionID;
    EndNodeArray(EndNodeNum,2)=ThisSectionX;
    EndNodeArray(EndNodeNum,3)=ThisSectionY;
    EndNodeArray(EndNodeNum,4)=ThisSectionDFO;
    EndNodeArray(EndNodeNum,5)=ThisSectionBRM;

else
    OldID15=OldID14;
    OldID14=OldID13;
    OldID13=OldID12;
    OldID12=OldID11;
    OldID11=OldID10;
    OldID10=OldID9;
    OldID9=OldID8;
    OldID8=OldID7;

```

```

        OldID7=OldID6;
        OldID6=OldID5;
        OldID5=OldID4;
        OldID4=OldID3;
        OldID3=OldID2;
        OldID2=OldID;
        OldID=ThisSectionID;
        ThisSectionID=MinDistID;
        ThisSectionX=MinDistX;
        ThisSectionY=MinDistY;
        ThisSectionDFO=MinDistDFO;
        ThisSectionBRM=MinDistBRM;
    end
end

end

clear SortedEndNodeArray;
clear FinalEndNodeArray;
SortedEndNodeArray=sortrows (EndNodeArray,1);
ArraySize=size (SortedEndNodeArray,1);
%   FinalEndNodeArray(1,1)= SortedEndNodeArray(1,1);
%   FinalEndNodeArray(1,2)= SortedEndNodeArray(1,2);
%   FinalEndNodeArray(1,3)= SortedEndNodeArray(1,3);
%   FinalEndNodeArray(1,4)= SortedEndNodeArray(1,4);
%   FinalEndNodeArray(1,5)= SortedEndNodeArray(1,5);
j=1;

if (ArraySize==1)
    FinalEndNodeArray(1,1)= SortedEndNodeArray(1,1);
    FinalEndNodeArray(1,2)= SortedEndNodeArray(1,2);
    FinalEndNodeArray(1,3)= SortedEndNodeArray(1,3);
    FinalEndNodeArray(1,4)= SortedEndNodeArray(1,4);
    FinalEndNodeArray(1,5)= SortedEndNodeArray(1,5);
else
    for i=1:ArraySize-1
        if
((SortedEndNodeArray(i,1)~=SortedEndNodeArray(i+1,1)) || (i==1))
            FinalEndNodeArray(j,1)= SortedEndNodeArray(i+1,1);
            FinalEndNodeArray(j,2)= SortedEndNodeArray(i+1,2);
            FinalEndNodeArray(j,3)= SortedEndNodeArray(i+1,3);
            FinalEndNodeArray(j,4)= SortedEndNodeArray(i+1,4);
            FinalEndNodeArray(j,5)= SortedEndNodeArray(i+1,5);
            j=j+1;
        end
    end
end

EndNoneNum=size (FinalEndNodeArray,1);

OldID=0;

```

```

OldID2=0;
OldID3=0;
OldID4=0;
OldID5=0;
OldID6=0;
OldID7=0;
OldID8=0;
OldID9=0;
OldID10=0;
OldID11=0;
OldID12=0;
OldID13=0;
OldID14=0;
OldID15=0;

ThisRank=1;
for i=1:EndNoneNum
    if (HN==188 || HN==195)
        i
        end
        %primary section = starting point
        ThisPrimSecID=FinalEndNodeArray(i,1);
        ThisPrimSecX=FinalEndNodeArray(i,2);
        ThisPrimSecY=FinalEndNodeArray(i,3);
        ThisPrimSecDFO=FinalEndNodeArray(i,4);
        ThisPrimSecBRM=FinalEndNodeArray(i,5);
        j=0;
        NNeighbor=0;
        ThisSectionID=ThisPrimSecID;
        ThisSectionX=ThisPrimSecX;
        ThisSectionY=ThisPrimSecY;
        ThisSectionDFO=ThisPrimSecDFO;
        ThisSectionBRM=ThisPrimSecBRM;
        TotalArray(ThisSectionID)=ThisRank;
        ThisRank=ThisRank+1;
        NoSecFound=0;
        while (NoSecFound==0)
            NNeighbor=0;
            MinDist=9999;
            for j=1:ThisHwSecNum
                ThisNeighborID=SortedMatrixByX(j,1);
                if (ThisNeighborID~=ThisSectionID &&
ThisNeighborID~=OldID && ThisNeighborID~=OldID2 &&
ThisNeighborID~=OldID3...
                    && ThisNeighborID~=OldID4 &&
ThisNeighborID~=OldID5 && ThisNeighborID~=OldID6 &&
ThisNeighborID~=OldID7...
                        && ThisNeighborID~=OldID8 &&
ThisNeighborID~=OldID9 && ThisNeighborID~=OldID10 &&
ThisNeighborID~=OldID11...
                            && ThisNeighborID~=OldID12 &&
ThisNeighborID~=OldID13 && ThisNeighborID~=OldID14 &&
ThisNeighborID~=OldID15)

```

```

        ThisNeighborX=SortedMatrixByX(j,2);
        ThisNeighborY=SortedMatrixByX(j,3);
        ThisNeighborDFO=SortedMatrixByX(j,4);
        ThisNeighborBRM=SortedMatrixByX(j,5);
        ThisDist=sqrt((ThisNeighborX-
ThisSectionX)^2+(ThisNeighborY-ThisSectionY)^2);
        %if ((ThisNeighborX>ThisSectionX ||
ThisNeighborY<ThisSectionY) && (ThisDist<DistLimit ||
ThisNeighborDFO==ThisSectionDFO))
            if ((ThisDist<DistLimit ||
ThisNeighborDFO==ThisSectionDFO || ThisNeighborBRM==ThisSectionBRM))
                NNeighbor=NNeighbor+1;
                if (ThisNeighborDFO==ThisSectionDFO)
                    MinDist=0;
                    MinDistID=ThisNeighborID;
                    MinDistX=ThisNeighborX;
                    MinDistY=ThisNeighborY;
                    MinDistDFO=ThisNeighborDFO;
                    MinDistBRM=ThisNeighborBRM;
                elseif (ThisDist<MinDist)
                    MinDist=ThisDist;
                    MinDistID=ThisNeighborID;
                    MinDistX=ThisNeighborX;
                    MinDistY=ThisNeighborY;
                    MinDistDFO=ThisNeighborDFO;
                    MinDistBRM=ThisNeighborBRM;
                end
            end
        end
    end
end
if (NNeighbor==0)
    NoSecFound=1;
else
    OldID15=OldID14;
    OldID14=OldID13;
    OldID13=OldID12;
    OldID12=OldID11;
    OldID11=OldID10;
    OldID10=OldID9;
    OldID9=OldID8;
    OldID8=OldID7;
    OldID7=OldID6;
    OldID6=OldID5;
    OldID5=OldID4;
    OldID4=OldID3;
    OldID3=OldID2;
    OldID2=OldID;
    OldID=ThisSectionID;
    ThisSectionID=MinDistID;
    ThisSectionX=MinDistX;
    ThisSectionY=MinDistY;
    ThisSectionDFO=MinDistDFO;

```

```

        ThisSectionBRM=MinDistBRM;
        TotalArray(ThisSectionID)=ThisRank;
        ThisRank=ThisRank+1;
    end
end

    end
end
end

for k=1:TotalSecNum
    ThisObjID=NetworkData(k).OBJECTID;
    NetworkData(k).Rank=TotalArray(ThisObjID);
end

shapewrite(NetworkData,RankedSHPAddress);
shapeinfo(RankedSHPAddress)
a='Ranking Done!'

```

```

function [ LogTotalProb ] = TotalProbFunc( HighwayMatrix,
ClusterLengths )

BryanFrequencyArray={3      3      4      0      5      5      3      4      5
4      1 ...
      4      3      2      1      1      1      3      0      1      1      1 ...
      0      1      1      0      0      1      0      1      0      0      0 ...
1};

% ClusterLengths=round(ClusterLengths);
[HighwayMatrixSize,a]=size(HighwayMatrix);
% RoundingError=HighwayMatrixSize-sum(ClusterLengths);
LogTotalProb=0;
SortedMatrix=sortrows(HighwayMatrix,2);
SortedClusteringMatrix=SortedMatrix(:,3:7);
[a,ClusterNum]=size(ClusterLengths);
ClusterLengths=ArrayRounder(ClusterLengths);

CumulativeLength=0;
for i=1:ClusterNum
    ThisClusterLength=ClusterLengths(i);

ThisArray=SortedClusteringMatrix(CumulativeLength+1:CumulativeLength+Th
isClusterLength,:);

    if (size(ThisArray,1)<6)
        if (size(ThisArray,1)==5)
            ThisArray(6,1)=50.1;
            ThisArray(6,2)=51.2;
            ThisArray(6,3)=52.3;
            ThisArray(6,4)=53.4;
            ThisArray(6,5)=54.5;
        elseif (size(ThisArray,1)==4)
            ThisArray(6,1)=61.6;
            ThisArray(6,2)=62.6;
            ThisArray(6,3)=63.6;
            ThisArray(6,4)=64.6;
            ThisArray(6,5)=65.6;

            ThisArray(5,1)=51.4;
            ThisArray(5,2)=52.4;
            ThisArray(5,3)=53.4;
            ThisArray(5,4)=54.4;
            ThisArray(5,5)=55.4;
        elseif (size(ThisArray,1)==3)

            ThisArray(6,1)=61.2;
            ThisArray(6,2)=62.2;
            ThisArray(6,3)=63;
            ThisArray(6,4)=64;
            ThisArray(6,5)=65;

```

```

ThisArray(5,1)=51;
ThisArray(5,2)=52;
ThisArray(5,3)=53;
ThisArray(5,4)=54;
ThisArray(5,5)=55;

ThisArray(4,1)=40;
ThisArray(4,2)=41;
ThisArray(4,3)=42;
ThisArray(4,4)=43;
ThisArray(4,5)=44;
elseif (size(ThisArray,1)==2)
ThisArray(6,1)=60;
ThisArray(6,2)=61;
ThisArray(6,3)=62;
ThisArray(6,4)=63;
ThisArray(6,5)=64;

ThisArray(5,1)=53;
ThisArray(5,2)=54;
ThisArray(5,3)=55;
ThisArray(5,4)=56;
ThisArray(5,5)=57;

ThisArray(4,1)=46;
ThisArray(4,2)=45;
ThisArray(4,3)=44;
ThisArray(4,4)=43;
ThisArray(4,5)=42;

ThisArray(3,1)=38;
ThisArray(3,2)=37;
ThisArray(3,3)=36;
ThisArray(3,4)=35;
ThisArray(3,5)=34;
elseif (size(ThisArray,1)==1)
ThisArray(6,1)=64;
ThisArray(6,2)=65;
ThisArray(6,3)=66;
ThisArray(6,4)=67;
ThisArray(6,5)=68;

ThisArray(5,1)=53;
ThisArray(5,2)=54;
ThisArray(5,3)=55;
ThisArray(5,4)=56;
ThisArray(5,5)=57;

ThisArray(4,1)=49;
ThisArray(4,2)=48;
ThisArray(4,3)=47;
ThisArray(4,4)=46;

```



```

        ThisArray(4,5)=45;

        ThisArray(3,1)=36;
        ThisArray(3,2)=34;
        ThisArray(3,3)=33;
        ThisArray(3,4)=32;
        ThisArray(3,5)=38;

        ThisArray(2,1)=29;
        ThisArray(2,2)=27;
        ThisArray(2,3)=25;
        ThisArray(2,4)=24;
        ThisArray(2,5)=23;
    end
end

ThisArray(1,1)=ThisArray(1,1)+0.1;

ThisArray(2,1)=ThisArray(2,1)+0.14;
ThisArray(2,2)=ThisArray(2,2)+0.11;
ThisArray(2,3)=ThisArray(2,3)+0.15;
ThisArray(2,4)=ThisArray(2,4)+0.16;
ThisArray(2,5)=ThisArray(2,5)+0.146;

if (ThisClusterLength>=3)
    ThisArray(3,1)=ThisArray(3,1)+0.17;
    ThisArray(3,2)=ThisArray(3,2)+0.18;
    ThisArray(3,3)=ThisArray(3,3)+0.12;
    ThisArray(3,4)=ThisArray(3,4)+0.19;
    ThisArray(3,5)=ThisArray(3,5)+0.129;
end
if (ThisClusterLength>=4)
    ThisArray(4,1)=ThisArray(4,1)+0.111;
    ThisArray(4,2)=ThisArray(4,2)+0.112;
    ThisArray(4,3)=ThisArray(4,3)+0.113;
    ThisArray(4,4)=ThisArray(4,4)+0.13;
    ThisArray(4,5)=ThisArray(4,5)+0.123;
end
if (ThisClusterLength>=5)
    ThisArray(5,5)=ThisArray(5,5)+0.123;
end

% options = statset('Display','final');
options = statset('Display','off');
if (size(ThisArray,1)<=size(ThisArray,2))
    display 'Error!'
end

```

```

%     try
%       NmDist = gmdistribution.fit(ThisArray,1,'Options',options);
%     catch

%     end
    ProbArray = pdf(NmDist,ThisArray);
    LogYMult=0;
    for j=1:ThisClusterLength
        LogYMult=LogYMult+log(ProbArray(j));
    end
    CumulativeLength=CumulativeLength+ThisClusterLength;

%LogTotalProb=ThisClusterLength*abs(log(ProbByNum(ThisClusterLength)))+
LogTotalProb+abs(LogYMult);

%LogTotalProb=LogTotalProb+LogYMult+ThisClusterLength*log(ProbByNum(This
sClusterLength));

LogTotalProb=LogTotalProb+LogYMult+log(ProbByNum(ThisClusterLength));

end

```

APPENDIX B: AUDIT AND ORIGINAL RATINGS OF NORMAL DATA

Rating #	Highway Name	BRM	BRM Distance	Year	Original DS	Audit DS	ΔDS
1	FM0050 K	0412	0	2011	36	45	9
2	FM0050 K	0412	0.5	2011	45	57	12
3	FM0050 K	0412	1	2011	56	73	17
4	FM0050 K	0412	1.5	2011	41	62	21
5	FM0060 K	0602	0	2011	98	100	2
6	FM0060 K	0602	0.5	2011	99	100	1
7	FM0060 K	0602	1	2011	99	100	1
8	FM0060 K	0602	1.5	2011	100	100	0
9	FM0060 K	0616	0	2011	100	100	0
10	FM0060 K	0616	0.5	2011	100	100	0
11	FM0060 K	0616	1	2011	100	100	0
12	FM0060 K	0616	1.5	2011	100	100	0
13	FM0111 K	0432	0	2011	75	100	25
14	FM0111 K	0432	0.5	2011	81	99	18
15	FM0158 K	0616	0	2011	58	58	0
16	FM0158 K	0616	1	2011	73	71	2
17	FM0158 K	0616	1.5	2011	87	90	3
18	FM0158 R	0616	0.4	2011	80	78	2
19	FM0159 K	0434	1	2011	100	100	0
20	FM0159 K	0434	1.5	2011	99	100	1
21	FM0391 K	0402	0	2011	100	100	0
22	FM0391 K	0402	0.5	2011	100	100	0
23	FM0391 K	0402	1	2011	99	90	9
24	FM0391 K	0402	1.5	2011	100	69	31
25	FM0485 K	0604	0	2011	85	100	15
26	FM0485 K	0604	0.5	2011	100	100	0
27	FM0485 K	0604	1	2011	100	100	0
28	FM0485 K	0604	1.5	2011	79	100	21
29	FM0485 K	0606	0	2011	100	100	0
30	FM0485 K	0606	0.5	2011	98	100	2
31	FM0486 K	0512	0	2011	100	100	0
32	FM0486 K	0512	0.5	2011	93	99	6
33	FM0486 K	0512	1	2011	100	100	0
34	FM0486 K	0512	1.5	2011	85	100	15
35	FM0487 K	0568	0	2011	99	100	1
36	FM0487 K	0568	0.5	2011	100	100	0
37	FM0487 K	0568	1	2011	100	100	0
38	FM0487 K	0568	1.5	2011	100	100	0
39	FM0487 K	0574	0	2011	99	100	1
40	FM0487 K	0574	0.5	2011	100	100	0

Rating #	Highway Name	BRM	BRM Distance	Year	Original DS	Audit DS	ΔDS
41	FM0487 K	0574	1	2011	100	100	0
42	FM0487 K	0574	1.5	2011	100	100	0
43	FM0487 K	0586	0	2011	100	100	0
44	FM0487 K	0586	0.5	2011	100	100	0
45	FM0487 K	0586	1	2011	100	100	0
46	FM0487 K	0586	1.5	2011	100	100	0
47	FM0487 K	0588	0	2011	100	100	0
48	FM0487 K	0588	0.5	2011	100	100	0
49	FM0487 K	0588	1	2011	100	100	0
50	FM0487 K	0588	1.5	2011	74	85	11
51	FM0908 K	0574	0	2011	98	100	2
52	FM0908 K	0574	0.5	2011	78	100	22
53	FM0908 K	0574	1	2011	82	100	18
54	FM0908 K	0574	1.5	2011	92	100	8
55	FM0908 K	0586	0	2011	100	100	0
56	FM0908 K	0586	0.5	2011	100	100	0
57	FM0908 K	0586	1	2011	100	100	0
58	FM0908 K	0586	1.5	2011	100	100	0
59	FM0937 K	0374	0	2011	100	100	0
60	FM0937 K	0374	0.5	2011	100	100	0
61	FM0937 K	0374	1	2011	100	100	0
62	FM0975 K	0420	0.5	2011	73	95	22
63	FM0975 K	0420	1	2011	70	100	30
64	FM0975 K	0420	1.5	2011	63	100	37
65	FM0975 K	0424	0	2011	100	100	0
66	FM0975 K	0424	0.5	2011	100	99	1
67	FM0975 K	0424	1.5	2011	98	99	1
68	FM0979 K	0604	0	2011	100	100	0
69	FM0979 K	0604	1	2011	75	100	25
70	FM0979 K	0610	0	2011	59	67	8
71	FM0979 K	0610	0.5	2011	100	100	0
72	FM0979 K	0610	1	2011	75	89	14
73	FM0979 K	0610	1.5	2011	71	90	19
74	FM0979 K	0622	0	2011	100	100	0
75	FM0979 K	0622	1	2011	100	100	0
76	FM0979 K	0622	1.5	2011	100	100	0
77	FM0979 K	0628	0	2011	100	100	0
78	FM0979 K	0628	0.5	2011	100	100	0
79	FM0979 K	0628	1	2011	100	100	0
80	FM0979 K	0628	1.5	2011	100	100	0

Rating #	Highway Name	BRM	BRM Distance	Year	Original DS	Audit DS	ΔDS
81	FM0979 K	0630	0	2011	100	100	0
82	FM0979 K	0630	0.5	2011	100	100	0
83	FM0979 K	0630	1	2011	100	100	0
84	FM1179 K	0408	0	2011	98	100	2
85	FM1179 K	0408	0.5	2011	100	100	0
86	FM1179 K	0408	1.5	2011	100	100	0
87	FM1362 K	0418	0	2011	71	100	29
88	FM1362 K	0418	0.5	2011	99	100	1
89	FM1362 K	0418	1	2011	70	100	30
90	FM1362 K	0418	1.5	2011	94	100	6
91	FM1444 K	0582	0	2011	78	100	22
92	FM1444 K	0582	0.5	2011	70	100	30
93	FM1444 K	0582	1	2011	100	100	0
94	FM1444 K	0582	1.5	2011	100	100	0
95	FM1452 K	0634	0	2011	100	100	0
96	FM1452 K	0634	0.5	2011	100	100	0
97	FM1452 K	0634	1	2011	100	100	0
98	FM1452 K	0634	1.5	2011	100	100	0
99	FM1452 K	0646	0	2011	100	100	0
100	FM1452 K	0646	0.5	2011	100	100	0
101	FM1452 K	0646	1	2011	100	100	0
102	FM1452 K	0646	1.5	2011	100	100	0
103	FM1600 K	0402	0	2011	81	100	19
104	FM1600 K	0402	0.5	2011	75	100	25
105	FM1600 K	0402	1.5	2011	77	100	23
106	FM1600 K	0406	0	2011	94	100	6
107	FM1600 K	0406	0.5	2011	69	69	0
108	FM1600 K	0406	1	2011	100	100	0
109	FM1600 K	0406	1.5	2011	100	100	0
110	FM1600 K	0406	2	2011	100	100	0
111	FM1915 K	0400	0	2011	100	100	0
112	FM1915 K	0400	0.5	2011	100	100	0
113	FM1915 K	0400	1	2011	94	100	6
114	FM1915 K	0400	1.5	2011	92	100	8
115	FM2000 K	0416	0	2011	100	100	0
116	FM2000 K	0416	0.5	2011	100	100	0
117	FM2000 K	0416	1	2011	100	100	0
118	FM2000 K	0416	1.5	2011	100	100	0
119	FM2095 K	0594	0	2011	74	100	26
120	FM2095 K	0594	0.5	2011	100	100	0

Rating #	Highway Name	BRM	BRM Distance	Year	Original DS	Audit DS	ΔDS
121	FM2095 K	0594	1.5	2011	65	100	35
122	FM2159 K	0380	0	2011	65	100	35
123	FM2159 K	0380	0.5	2011	66	99	33
124	FM2223 K	0402	0	2011	100	100	0
125	FM2223 K	0402	1	2011	100	100	0
126	FM2223 K	0402	1.5	2011	100	100	0
127	FM2346 K	0382	0	2011	100	100	0
128	FM2346 K	0382	0.5	2011	100	100	0
129	FM2346 K	0382	1	2011	100	100	0
130	FM2346 K	0382	1.5	2011	100	100	0
131	FM2346 K	0386	0.5	2011	99	100	1
132	FM2346 K	0386	1	2011	100	100	0
133	FM2346 K	0386	1.5	2011	100	100	0
134	FM2347 L	0426	0	2011	100	100	0
135	FM2347 L	0426	0.5	2011	100	100	0
136	FM2549 K	0394	0	2011	100	100	0
137	FM2549 K	0394	0.5	2011	100	100	0
138	FM2549 K	0394	1	2011	100	100	0
139	FM2549 K	0394	1.5	2011	100	100	0
140	FM2549 K	0396	0	2011	100	100	0
141	FM2549 K	0396	0.5	2011	100	100	0
142	FM2549 K	0396	1	2011	100	100	0
143	FM2549 K	0396	1.5	2011	100	100	0
144	FM2549 K	0400	0	2011	100	100	0
145	FM2549 K	0400	0.5	2011	100	99	1
146	FM2549 K	0400	1.5	2011	100	100	0
147	FM2549 K	0402	0	2011	100	100	0
148	FM2549 K	0402	0.5	2011	100	100	0
149	FM2549 K	0402	1	2011	100	100	0
150	FM2549 K	0402	1.5	2011	100	100	0
151	FM3058 K	0568	0	2011	100	100	0
152	FM3058 K	0568	0.5	2011	100	100	0
153	FM3058 K	0568	1.5	2011	100	100	0
154	SH0007 K	0614	0	2011	100	100	0
155	SH0007 K	0614	0.5	2011	100	100	0
156	SH0007 K	0614	1	2011	100	100	0
157	SH0007 K	0614	1.5	2011	100	100	0
158	SH0021 K	0688	0	2011	100	100	0
159	SH0021 K	0688	0.5	2011	100	100	0
160	SH0021 K	0688	1	2011	100	100	0

Rating #	Highway Name	BRM	BRM Distance	Year	Original DS	Audit DS	ΔDS
161	SH0021 K	0688	1.5	2011	99	99	0
162	SH0021 R	0632	0	2011	100	100	0
163	SH0021 R	0632	0.5	2011	100	100	0
164	SH0021 R	0632	1	2011	100	100	0
165	SH0021 R	0632	1.3	2011	100	100	0
166	SH0021 R	0632	2.5	2011	100	100	0
167	SH0021 R	0632	3	2011	100	100	0
168	SH0021 R	0634	0	2011	100	100	0
169	SH0021 R	0634	0.5	2011	100	100	0
170	SH0021 R	0634	1	2011	100	100	0
171	SH0021 R	0634	1.5	2011	100	100	0
172	SH0030 K	0624	0	2011	100	100	0
173	SH0030 K	0624	0.5	2011	100	100	0
174	SH0030 K	0624	1	2011	100	100	0
175	SH0030 K	0624	1.5	2011	91	100	9
176	SH0036 K	0552	0	2011	100	100	0
177	SH0036 K	0552	1	2011	94	95	1
178	SH0036 K	0552	1.5	2011	69	93	24
179	SH0090 K	0394	0	2011	58	96	38
180	SH0090 K	0394	0.5	2011	68	71	3
181	SH0090 K	0394	1	2011	50	69	19
182	SH0090 K	0394	1.5	2011	49	81	32
183	SH0090 K	0394	2	2011	53	76	23
184	SH0090 K	0394	2.5	2011	53	80	27
185	SH0090 K	0394	3	2011	78	89	11
186	US0077 K	0412	0	2011	100	100	0
187	US0077 K	0412	0.5	2011	100	100	0
188	US0077 K	0412	1	2011	98	100	2
189	US0077 K	0412	1.5	2011	100	100	0
190	US0077 K	0416	0	2011	100	100	0
191	US0077 K	0416	0.5	2011	100	100	0
192	US0077 K	0416	1	2011	98	100	2
193	US0077 K	0416	1.5	2011	100	100	0
194	US0079 K	0474	0	2011	100	100	0
195	US0079 K	0474	0.5	2011	100	100	0
196	US0079 K	0474	1	2011	100	100	0
197	US0079 K	0474	1.5	2011	100	100	0
198	US0079 K	0482	0	2011	99	99	0
199	US0079 K	0482	0.5	2011	100	100	0
200	US0079 K	0482	1	2011	100	100	0

Rating #	Highway Name	BRM	BRM Distance	Year	Original DS	Audit DS	ΔDS
201	US0079 K	0482	1.5	2011	100	100	0
202	US0079 K	0500	0	2011	79	81	2
203	US0079 K	0500	0.5	2011	90	91	1
204	US0079 K	0500	1	2011	100	100	0
205	US0079 K	0500	1.5	2011	88	89	1
206	US0079 K	0522	0	2011	100	100	0
207	US0079 K	0522	0.5	2011	100	100	0
208	US0190 K	0622	0	2011	100	100	0
209	US0190 K	0622	0.5	2011	91	96	5
210	US0190 K	0660	0	2011	100	100	0
211	US0190 K	0660	0.5	2011	100	100	0
212	US0190 K	0660	1	2011	100	100	0
213	US0190 K	0694	0	2011	100	100	0
214	US0190 K	0694	0.5	2011	100	100	0
215	US0190 K	0694	1	2011	100	100	0
216	US0190 K	0694	1.5	2011	100	100	0
217	US0190 K	0708	0	2011	87	100	13
218	US0190 K	0708	0.5	2011	100	100	0
219	US0190 K	0708	1.5	2011	100	100	0
220	US0190 L	0672	0	2011	100	100	0
221	US0190 L	0672	0.1	2011	100	100	0
222	US0190 L	0672	0.6	2011	100	100	0
223	US0190 L	0672	1.1	2011	82	83	1
224	US0190 L	0672	1.6	2011	85	79	6
225	FM0046 K	0602	0	2010	100	100	0
226	FM0046 K	0602	0.5	2010	96	100	4
227	FM0046 K	0602	1	2010	100	100	0
228	FM0046 K	0602	1.5	2010	100	100	0
229	FM0050 K	0408	0	2010	100	100	0
230	FM0050 K	0408	0.5	2010	100	100	0
231	FM0050 K	0408	1	2010	100	100	0
232	FM0050 K	0408	1.5	2010	100	100	0
233	FM0050 K	0412	0	2010	67	96	29
234	FM0050 K	0412	0.5	2010	53	76	23
235	FM0050 K	0412	1	2010	47	50	3
236	FM0050 K	0412	1.5	2010	47	56	9
237	FM0050 K	0432	0	2010	100	100	0
238	FM0050 K	0432	0.5	2010	100	100	0
239	FM0050 K	0432	1	2010	100	100	0
240	FM0050 K	0432	1.5	2010	100	100	0

Rating #	Highway Name	BRM	BRM Distance	Year	Original DS	Audit DS	ΔDS
241	FM0060 K	0604	0	2010	100	100	0
242	FM0060 K	0604	0.5	2010	85	100	15
243	FM0060 K	0604	1	2010	68	100	32
244	FM0060 K	0604	1.5	2010	99	100	1
245	FM0060 K	0620	0	2010	100	100	0
246	FM0060 L	0636	0	2010	100	99	1
247	FM0060 L	0636	0.5	2010	100	100	0
248	FM0060 L	0636	0.8	2010	91	84	7
249	FM0111 K	0424	0	2010	66	100	34
250	FM0111 K	0424	0.5	2010	62	100	38
251	FM0111 K	0424	1.5	2010	65	100	35
252	FM0111 K	0426	0	2010	64	100	36
253	FM0111 K	0426	0.5	2010	67	100	33
254	FM0111 K	0426	1	2010	61	100	39
255	FM0149 K	0426	0	2010	100	100	0
256	FM0149 K	0432	0	2010	100	100	0
257	FM0149 K	0432	0.5	2010	85	100	15
258	FM0149 K	0432	1	2010	100	100	0
259	FM0149 K	0432	1.5	2010	82	100	18
260	FM0158 L	0620	0	2010	100	96	4
261	FM0158 L	0620	0.5	2010	99	96	3
262	FM0158 L	0620	1	2010	100	93	7
263	FM0158 L	0620	1.5	2010	100	100	0
264	FM0166 K	0604	0	2010	85	100	15
265	FM0166 K	0604	0.5	2010	94	100	6
266	FM0166 K	0604	1	2010	68	99	31
267	FM0166 K	0604	1.5	2010	96	100	4
268	FM0244 K	0416	0	2010	100	100	0
269	FM0244 K	0416	0.5	2010	100	100	0
270	FM0244 K	0416	1	2010	100	100	0
271	FM0244 K	0416	1.5	2010	100	100	0
272	FM0485 K	0580	0	2010	77	100	23
273	FM0485 K	0580	0.5	2010	53	69	16
274	FM0485 K	0580	1	2010	86	90	4
275	FM0485 K	0580	1.5	2010	86	100	14
276	FM0485 K	0586	0	2010	91	100	9
277	FM0485 K	0586	0.5	2010	98	100	2
278	FM0485 K	0586	1	2010	87	100	13
279	FM0485 K	0586	1.5	2010	100	100	0
280	FM0486 K	0502	0	2010	100	100	0

Rating #	Highway Name	BRM	BRM Distance	Year	Original DS	Audit DS	ΔDS
281	FM0486 K	0502	0.5	2010	97	99	2
282	FM0486 K	0502	1	2010	56	94	38
283	FM0486 K	0502	1.5	2010	86	100	14
284	FM0486 K	0512	0	2010	97	100	3
285	FM0486 K	0512	0.5	2010	93	100	7
286	FM0486 K	0512	1	2010	100	100	0
287	FM0486 K	0512	1.5	2010	85	100	15
288	FM0487 K	0570	0	2010	73	99	26
289	FM0487 K	0570	0.5	2010	74	99	25
290	FM0487 K	0570	1.5	2010	100	100	0
291	FM0487 K	0592	0	2010	100	100	0
292	FM0487 K	0592	0.5	2010	85	100	15
293	FM0487 K	0592	1	2010	100	100	0
294	FM0487 K	0592	1.5	2010	99	100	1
295	FM0696 K	0596	0	2010	100	100	0
296	FM0696 K	0596	0.5	2010	100	100	0
297	FM0696 K	0596	1	2010	100	100	0
298	FM0908 K	0586	0	2010	100	100	0
299	FM0908 K	0586	0.5	2010	100	100	0
300	FM0908 K	0586	1	2010	100	100	0
301	FM0908 K	0586	1.5	2010	100	100	0
302	FM0908 K	0594	0	2010	100	100	0
303	FM0908 K	0594	0.5	2010	100	100	0
304	FM0908 K	0594	1	2010	100	100	0
305	FM0908 K	0596	0	2010	100	100	0
306	FM0908 K	0596	0.5	2010	100	100	0
307	FM0908 K	0596	1	2010	100	100	0
308	FM0908 K	0596	1.5	2010	100	100	0
309	FM0908 K	0598	0	2010	100	100	0
310	FM0908 K	0598	0.5	2010	100	100	0
311	FM0908 K	0598	1	2010	100	100	0
312	FM0908 K	0598	1.5	2010	100	100	0
313	FM0908 K	0600	0	2010	100	100	0
314	FM0908 K	0600	0.5	2010	100	100	0
315	FM0908 K	0600	1	2010	100	100	0
316	FM0908 K	0600	1.5	2010	100	100	0
317	FM0908 K	0602	0	2010	98	100	2
318	FM0908 K	0602	0.5	2010	100	100	0
319	FM0908 K	0602	1	2010	100	100	0
320	FM0908 K	0602	1.5	2010	100	100	0

Rating #	Highway Name	BRM	BRM Distance	Year	Original DS	Audit DS	ΔDS
321	FM0974 K	0636	0	2010	100	100	0
322	FM0974 K	0636	0.5	2010	100	100	0
323	FM0974 K	0636	1	2010	100	100	0
324	FM0974 K	0636	1.5	2010	100	100	0
325	FM0979 K	0600	0	2010	95	100	5
326	FM0979 K	0600	0.5	2010	95	99	4
327	FM0979 K	0600	1	2010	89	99	10
328	FM0979 K	0600	1.5	2010	100	99	1
329	FM0979 K	0610	0	2010	99	99	0
330	FM0979 K	0610	0.5	2010	98	100	2
331	FM0979 K	0610	1	2010	100	100	0
332	FM0979 K	0610	1.5	2010	91	100	9
333	FM0979 K	0612	0	2010	47	54	7
334	FM0979 K	0612	0.5	2010	43	44	1
335	FM0979 K	0612	1	2010	60	100	40
336	FM0979 K	0612	1.5	2010	97	100	3
337	FM1179 K	0406	0	2010	100	100	0
338	FM1179 K	0406	0.5	2010	100	100	0
339	FM1179 K	0406	1	2010	100	100	0
340	FM1227 K	0432	0	2010	65	100	35
341	FM1227 K	0432	0.5	2010	66	100	34
342	FM1227 K	0432	1	2010	70	100	30
343	FM1227 K	0432	1.5	2010	70	100	30
344	FM1331 K	0570	0	2010	85	90	5
345	FM1331 K	0570	0.5	2010	68	100	32
346	FM1331 K	0570	1	2010	87	100	13
347	FM1444 K	0582	0	2010	85	100	15
348	FM1444 K	0582	0.5	2010	72	100	28
349	FM1444 K	0582	1	2010	100	100	0
350	FM1444 K	0582	1.5	2010	100	100	0
351	FM1486 K	0412	0	2010	100	100	0
352	FM1486 K	0412	0.5	2010	100	100	0
353	FM1486 K	0412	1	2010	100	100	0
354	FM1600 K	0402	0	2010	81	100	19
355	FM1600 K	0402	0.5	2010	71	100	29
356	FM1600 K	0402	1.5	2010	66	89	23
357	FM1644 K	0404	0	2010	100	100	0
358	FM1644 K	0404	0.5	2010	100	100	0
359	FM1774 K	0434	0	2010	71	100	29
360	FM1774 K	0434	0.5	2010	75	100	25

Rating #	Highway Name	BRM	BRM Distance	Year	Original DS	Audit DS	ΔDS
361	FM1774 K	0434	1	2010	74	100	26
362	FM1774 K	0434	1.5	2010	100	100	0
363	FM1915 K	0392	0.5	2010	88	96	8
364	FM1915 K	0392	1	2010	79	98	19
365	FM1915 K	0392	1.5	2010	59	69	10
366	FM2000 K	0418	0	2010	100	100	0
367	FM2000 K	0418	0.5	2010	100	100	0
368	FM2000 K	0418	1	2010	78	100	22
369	FM2000 K	0418	1.5	2010	78	100	22
370	FM2000 K	0420	0	2010	81	100	19
371	FM2000 K	0420	0.5	2010	64	100	36
372	FM2027 K	0386	0	2010	91	88	3
373	FM2027 K	0388	0	2010	99	99	0
374	FM2027 K	0388	0.5	2010	100	100	0
375	FM2027 K	0388	1	2010	100	100	0
376	FM2027 K	0388	1.5	2010	100	100	0
377	FM2027 K	0390	0	2010	100	100	0
378	FM2027 K	0390	0.5	2010	100	100	0
379	FM2027 K	0390	1	2010	100	100	0
380	FM2027 K	0390	1.5	2010	100	100	0
381	FM2027 K	0392	0	2010	100	100	0
382	FM2027 K	0392	0.5	2010	100	100	0
383	FM2027 K	0392	1	2010	100	100	0
384	FM2027 K	0392	1.5	2010	100	100	0
385	FM2027 K	0392	2	2010	100	100	0
386	FM2027 K	0392	2.5	2010	100	100	0
387	FM2038 K	0622	0	2010	100	100	0
388	FM2038 K	0622	0.5	2010	100	100	0
389	FM2038 K	0622	1	2010	100	100	0
390	FM2095 K	0592	1	2010	69	100	31
391	FM2154 K	0624	0	2010	99	100	1
392	FM2154 K	0624	0.5	2010	100	100	0
393	FM2154 K	0624	1	2010	100	100	0
394	FM2154 K	0624	1.5	2010	100	100	0
395	FM2154 K	0628	0	2010	100	100	0
396	FM2154 K	0628	0.5	2010	100	100	0
397	FM2154 K	0628	1	2010	100	100	0
398	FM2154 K	0628	1.5	2010	100	100	0
399	FM2159 K	0384	0	2010	100	100	0
400	FM2159 K	0384	0.5	2010	100	100	0

Rating #	Highway Name	BRM	BRM Distance	Year	Original DS	Audit DS	ΔDS
401	FM2159 K	0384	1	2010	89	100	11
402	FM2159 K	0384	1.5	2010	98	100	2
403	FM2223 K	0404	0.5	2010	98	100	2
404	FM2223 K	0404	1	2010	100	100	0
405	FM2223 K	0404	1.5	2010	100	100	0
406	FM2446 K	0618	0	2010	100	100	0
407	FM2446 K	0618	0.5	2010	100	100	0
408	FM2446 K	0618	1	2010	100	100	0
409	FM2446 K	0618	1.5	2010	100	100	0
410	FM2549 K	0402	0	2010	97	100	3
411	FM2549 K	0402	0.5	2010	96	100	4
412	FM2549 K	0402	1	2010	98	100	2
413	FM2549 K	0402	1.5	2010	100	100	0
414	FM2620 K	0406	0	2010	92	100	8
415	FM2620 K	0406	0.5	2010	77	90	13
416	FM2620 K	0406	1	2010	100	100	0
417	FM2620 K	0406	1.5	2010	100	100	0
418	FM2620 K	0410	0	2010	68	69	1
419	FM2620 K	0410	0.5	2010	100	100	0
420	FM2620 K	0410	1	2010	100	100	0
421	FM2819 K	0646	0	2010	100	100	0
422	FM2819 K	0646	0.5	2010	85	90	5
423	FM2819 K	0646	1	2010	72	100	28
424	FM2819 K	0646	1.5	2010	100	100	0
425	FM2819 K	0648	0	2010	82	100	18
426	FM2819 K	0648	0.5	2010	87	100	13
427	FM2819 K	0648	1	2010	100	100	0
428	FM2819 K	0648	1.5	2010	100	100	0
429	FM3090 K	0422	0	2010	99	100	1
430	FM3090 K	0422	0.5	2010	100	100	0
431	FM3090 K	0422	1	2010	69	69	0
432	FM3090 K	0422	1.5	2010	99	100	1
433	FM3242 K	0402	0	2010	71	100	29
434	FM3242 K	0402	0.5	2010	82	100	18
435	FM3242 K	0402	1	2010	81	100	19
436	FM3242 K	0402	1.5	2010	69	100	31
437	FM3242 K	0406	0	2010	81	100	19
438	FM3242 K	0406	1	2010	92	100	8
439	FM3242 K	0406	1.5	2010	92	100	8
440	SH0006 L	0610	0	2010	100	100	0

Rating #	Highway Name	BRM	BRM Distance	Year	Original DS	Audit DS	ΔDS
441	SH0006 L	0610	0.5	2010	100	100	0
442	SH0006 L	0610	1.1	2010	100	100	0
443	SH0006 L	0610	1.6	2010	100	100	0
444	SH0014 K	0384	0.5	2010	81	89	8
445	SH0014 K	0384	1	2010	92	100	8
446	SH0014 K	0384	1.5	2010	92	100	8
447	SH0021 K	0620	0	2010	92	88	4
448	SH0021 K	0620	1	2010	88	85	3
449	SH0021 K	0620	1.5	2010	100	93	7
450	SH0021 L	0622	0.5	2010	100	100	0
451	SH0021 L	0622	1	2010	100	100	0
452	SH0021 L	0622	1.5	2010	100	100	0
453	SH0021 L	0628	0	2010	100	100	0
454	SH0021 L	0628	1	2010	100	100	0
455	SH0021 L	0628	1.5	2010	100	100	0
456	SH0021 R	0634	0	2010	100	100	0
457	SH0021 R	0634	0.5	2010	100	100	0
458	SH0021 R	0634	1	2010	100	100	0
459	SH0021 R	0634	1.5	2010	100	100	0
460	SH0030 K	0622	0	2010	97	98	1
461	SH0030 K	0622	0.5	2010	81	81	0
462	SH0030 K	0622	1	2010	90	90	0
463	SH0030 K	0622	1.5	2010	88	87	1
464	SH0030 K	0624	0	2010	100	100	0
465	SH0030 K	0624	0.5	2010	100	100	0
466	SH0030 K	0624	1	2010	100	100	0
467	SH0030 K	0624	1.5	2010	100	100	0
468	SH0030 K	0628	0	2010	100	100	0
469	SH0030 K	0628	0.5	2010	100	100	0
470	SH0030 K	0628	1	2010	100	100	0
471	SH0030 K	0628	1.5	2010	100	100	0
472	SH0030 K	0638	0	2010	100	100	0
473	SH0030 K	0638	0.5	2010	100	100	0
474	SH0030 K	0638	1	2010	100	100	0
475	SH0030 K	0638	1.5	2010	100	100	0
476	SH0090 K	0398	0	2010	69	100	31
477	SH0090 K	0398	0.5	2010	64	100	36
478	SH0090 K	0398	1	2010	72	100	28
479	SH0090 K	0398	1.5	2010	63	90	27
480	SH0090 K	0412	0.5	2010	61	100	39

Rating #	Highway Name	BRM	BRM Distance	Year	Original DS	Audit DS	ΔDS
481	SH0090 K	0412	1	2010	66	100	34
482	SH0090 K	0412	1.5	2010	67	100	33
483	SH0105 K	0656	0	2010	100	100	0
484	SH0105 K	0656	0.5	2010	100	100	0
485	SH0105 K	0656	1	2010	100	100	0
486	SH0105 K	0656	1.5	2010	99	100	1
487	US0077 K	0428	0.5	2010	100	100	0
488	US0077 K	0428	1	2010	100	100	0
489	US0077 K	0428	1.5	2010	99	100	1
490	US0079 K	0484	0	2010	99	100	1
491	US0079 K	0484	0.5	2010	100	100	0
492	US0079 K	0484	1	2010	100	100	0
493	US0079 K	0502	0	2010	87	87	0
494	US0079 K	0502	0.5	2010	98	100	2
495	US0079 K	0502	1	2010	95	99	4
496	US0079 K	0502	1.5	2010	100	100	0
497	US0079 K	0520	0	2010	100	100	0
498	US0079 K	0520	0.5	2010	100	100	0
499	US0079 K	0520	1	2010	100	100	0
500	US0079 K	0520	1.5	2010	99	100	1
501	US0190 K	0692	0	2010	83	94	11
502	US0190 K	0692	0.5	2010	72	82	10
503	FM0039 K	0414	0	2009	73	100	27
504	FM0039 K	0414	0.5	2009	70	100	30
505	FM0039 K	0414	1	2009	82	100	18
506	FM0039 K	0414	1.5	2009	100	100	0
507	FM0039 K	0426	0	2009	70	100	30
508	FM0039 K	0426	0.5	2009	70	100	30
509	FM0050 K	0412	0	2009	76	99	23
510	FM0050 K	0412	0.5	2009	99	100	1
511	FM0050 K	0412	1	2009	93	96	3
512	FM0050 K	0424	0	2009	68	100	32
513	FM0050 K	0424	0.5	2009	100	100	0
514	FM0050 K	0424	1	2009	90	90	0
515	FM0050 K	0424	1.5	2009	82	100	18
516	FM0050 K	0430	0	2009	100	100	0
517	FM0050 K	0430	0.5	2009	100	100	0
518	FM0050 K	0430	1	2009	100	100	0
519	FM0050 K	0430	1.5	2009	100	100	0

Rating #	Highway Name	BRM	BRM Distance	Year	Original DS	Audit DS	ΔDS
520	FM0060 K	0598	0	2009	100	100	0
521	FM0060 K	0598	0.5	2009	100	100	0
522	FM0060 K	0598	1	2009	100	100	0
523	FM0060 K	0598	1.5	2009	100	100	0
524	FM0060 K	0618	0	2009	82	100	18
525	FM0060 K	0632	1	2009	61	58	3
526	FM0060 L	0636	0	2009	100	100	0
527	FM0060 L	0636	0.5	2009	86	96	10
528	FM0060 L	0636	0.8	2009	99	99	0
529	FM0149 K	0432	0	2009	100	100	0
530	FM0149 K	0432	0.5	2009	100	100	0
531	FM0149 K	0432	1	2009	100	100	0
532	FM0149 K	0432	1.5	2009	69	69	0
533	FM0158 R	0620	0	2009	100	100	0
534	FM0158 R	0620	0.5	2009	100	100	0
535	FM0158 R	0620	1	2009	100	100	0
536	FM0158 R	0620	1.5	2009	100	100	0
537	FM0159 K	0430	0	2009	90	90	0
538	FM0159 K	0430	0.5	2009	100	100	0
539	FM0159 K	0430	1	2009	100	100	0
540	FM0159 K	0430	1.5	2009	100	100	0
541	FM0166 K	0602	0	2009	77	100	23
542	FM0166 K	0602	0.5	2009	81	100	19
543	FM0166 K	0602	1	2009	100	100	0
544	FM0166 K	0602	1.5	2009	100	100	0
545	FM0244 K	0402	1.5	2009	66	100	34
546	FM0244 K	0416	0	2009	100	100	0
547	FM0244 K	0416	0.5	2009	100	100	0
548	FM0244 K	0416	1	2009	100	100	0
549	FM0244 K	0416	1.5	2009	100	100	0
550	FM0391 K	0398	0	2009	100	100	0
551	FM0391 K	0398	0.5	2009	99	99	0
552	FM0391 K	0398	1	2009	100	99	1
553	FM0391 K	0400	0	2009	100	100	0
554	FM0391 K	0400	0.5	2009	100	100	0
555	FM0391 K	0400	1	2009	100	99	1
556	FM0391 K	0400	1.5	2009	100	100	0
557	FM0437 K	0584	0	2009	100	100	0
558	FM0437 K	0584	0.5	2009	100	100	0
559	FM0437 K	0584	1	2009	100	100	0
560	FM0437 K	0584	1.5	2009	100	100	0

Rating #	Highway Name	BRM	BRM Distance	Year	Original DS	Audit DS	ΔDS
561	FM0437 K	0586	0	2009	100	100	0
562	FM0437 K	0586	0.5	2009	100	100	0
563	FM0437 K	0586	1	2009	100	100	0
564	FM0485 K	0590	0	2009	100	100	0
565	FM0485 K	0590	0.5	2009	100	100	0
566	FM0485 K	0590	1	2009	100	100	0
567	FM0485 K	0590	1.5	2009	100	100	0
568	FM0485 K	0594	0	2009	100	100	0
569	FM0485 K	0594	0.5	2009	100	100	0
570	FM0485 K	0594	1	2009	100	100	0
571	FM0485 K	0594	1.5	2009	100	100	0
572	FM0486 K	0502	0	2009	100	100	0
573	FM0486 K	0502	0.5	2009	99	99	0
574	FM0486 K	0502	1	2009	62	99	37
575	FM0486 K	0502	1.5	2009	100	100	0
576	FM0486 K	0516	0	2009	100	100	0
577	FM0486 K	0516	0.5	2009	100	100	0
578	FM0486 K	0516	1	2009	100	100	0
579	FM0486 K	0516	1.5	2009	100	100	0
580	FM0487 K	0574	0	2009	100	100	0
581	FM0487 K	0574	0.5	2009	100	100	0
582	FM0487 K	0574	1	2009	66	65	1
583	FM0487 K	0574	1.5	2009	100	100	0
584	FM0908 K	0594	0	2009	100	100	0
585	FM0908 K	0594	0.5	2009	100	100	0
586	FM0908 K	0594	1	2009	100	100	0
587	FM0937 K	0372	0	2009	100	100	0
588	FM0937 K	0372	0.5	2009	100	100	0
589	FM0937 K	0372	1	2009	100	100	0
590	FM0937 K	0372	1.5	2009	100	100	0
591	FM0974 K	0620	0	2009	100	100	0
592	FM0974 K	0620	0.5	2009	100	100	0
593	FM0974 K	0620	1	2009	100	100	0
594	FM0974 K	0620	1.5	2009	100	100	0
595	FM0975 K	0420	0.5	2009	96	96	0
596	FM0975 K	0420	1	2009	91	100	9
597	FM0975 K	0420	1.5	2009	100	100	0
598	FM0979 K	0586	0	2009	100	100	0
599	FM0979 K	0586	0.5	2009	100	100	0
600	FM0979 K	0586	1	2009	100	100	0

Rating #	Highway Name	BRM	BRM Distance	Year	Original DS	Audit DS	ΔDS
601	FM0979 K	0586	1.5	2009	100	100	0
602	FM0979 K	0596	0	2009	74	100	26
603	FM0979 K	0596	0.5	2009	74	100	26
604	FM0979 K	0608	0	2009	100	100	0
605	FM0979 K	0608	0.5	2009	78	96	18
606	FM0979 K	0608	1	2009	82	100	18
607	FM0979 K	0608	1.5	2009	100	100	0
608	FM1361 K	0618	0	2009	100	100	0
609	FM1361 K	0618	0.5	2009	100	100	0
610	FM1361 K	0618	1	2009	100	100	0
611	FM1361 K	0618	1.5	2009	100	100	0
612	FM1444 K	0582	0	2009	85	100	15
613	FM1444 K	0582	0.5	2009	85	100	15
614	FM1444 K	0582	1	2009	100	100	0
615	FM1444 K	0582	1.5	2009	100	100	0
616	FM1600 K	0402	0	2009	92	100	8
617	FM1600 K	0402	0.5	2009	94	100	6
618	FM1600 K	0402	1.5	2009	100	99	1
619	FM1644 K	0400	0	2009	100	100	0
620	FM1644 K	0400	0.5	2009	100	100	0
621	FM1644 K	0400	1	2009	100	100	0
622	FM1644 K	0400	1.5	2009	100	100	0
623	FM1696 K	0644	0	2009	61	100	39
624	FM1696 K	0644	0.5	2009	62	100	38
625	FM1696 K	0644	1	2009	65	100	35
626	FM1696 K	0644	1.5	2009	64	100	36
627	FM1696 K	0650	0	2009	68	100	32
628	FM1915 K	0396	0	2009	76	96	20
629	FM1915 K	0396	0.5	2009	76	97	21
630	FM1915 K	0396	1	2009	53	64	11
631	FM1915 K	0396	1.5	2009	88	88	0
632	FM1915 K	0398	0.5	2009	76	79	3
633	FM1915 K	0398	1	2009	75	72	3
634	FM2027 K	0388	0.5	2009	100	100	0
635	FM2027 K	0388	1	2009	100	100	0
636	FM2027 K	0388	1.5	2009	100	100	0
637	FM2095 K	0588	0	2009	100	100	0
638	FM2095 K	0588	0.5	2009	100	100	0
639	FM2095 K	0588	1	2009	100	100	0
640	FM2095 K	0588	1.5	2009	71	100	29

Rating #	Highway Name	BRM	BRM Distance	Year	Original DS	Audit DS	ΔDS
641	FM2096 K	0378	0	2009	100	100	0
642	FM2096 K	0378	0.5	2009	100	100	0
643	FM2096 K	0378	1	2009	100	100	0
644	FM2096 K	0378	1.5	2009	100	100	0
645	FM2155 K	0424	0	2009	73	100	27
646	FM2155 K	0424	0.5	2009	65	100	35
647	FM2155 K	0424	1	2009	63	100	37
648	FM2155 K	0424	1.5	2009	61	100	39
649	FM2223 K	0404	0	2009	100	100	0
650	FM2223 K	0404	0.5	2009	100	100	0
651	FM2223 K	0404	1	2009	100	100	0
652	FM2223 K	0404	1.5	2009	100	100	0
653	FM2223 K	0406	0	2009	100	100	0
654	FM2223 K	0406	0.5	2009	100	100	0
655	FM2223 K	0406	1	2009	100	100	0
656	FM2223 K	0406	1.5	2009	100	100	0
657	FM2446 K	0620	0	2009	100	100	0
658	FM2446 K	0620	1	2009	100	100	0
659	FM2446 K	0620	1.5	2009	100	100	0
660	FM2620 K	0408	0	2009	78	100	22
661	FM2620 K	0408	0.5	2009	82	100	18
662	FM2818 K	0418	1.5	2009	100	100	0
663	FM2818 R	0410	0.1	2009	100	100	0
664	FM2818 R	0410	0.6	2009	100	100	0
665	FM2818 R	0410	0.9	2009	100	100	0
666	FM2818 R	0410	1.4	2009	100	100	0
667	FM2988 K	0636	0	2009	100	100	0
668	FM2988 K	0636	0.5	2009	100	100	0
669	FM2988 K	0636	1	2009	78	96	18
670	FM2988 K	0636	1.5	2009	98	100	2
671	SH0006 R	0612	0	2009	78	79	1
672	SH0006 R	0612	1.1	2009	90	90	0
673	SH0006 R	0614	0	2009	98	100	2
674	SH0006 R	0614	0.5	2009	98	100	2
675	SH0006 R	0614	1	2009	100	100	0
676	SH0006 R	0614	1.5	2009	99	98	1
677	SH0007 K	0612	1	2009	65	100	35
678	SH0007 K	0616	0	2009	98	100	2
679	SH0007 K	0616	0.5	2009	97	97	0
680	SH0007 K	0616	1	2009	99	99	0

Rating #	Highway Name	BRM	BRM Distance	Year	Original DS	Audit DS	ΔDS
681	SH0007 K	0616	1.5	2009	100	100	0
682	SH0021 L	0632	0	2009	100	100	0
683	SH0021 L	0632	0.5	2009	100	100	0
684	SH0021 L	0632	1	2009	100	100	0
685	SH0021 L	0632	1.3	2009	100	100	0
686	SH0021 L	0632	2.5	2009	100	100	0
687	SH0021 L	0632	3	2009	100	100	0
688	SH0021 L	0638	0	2009	100	100	0
689	SH0021 L	0638	0.5	2009	100	100	0
690	SH0021 L	0638	1	2009	100	100	0
691	SH0030 K	0652	0	2009	85	100	15
692	SH0030 K	0652	0.5	2009	97	98	1
693	SH0030 K	0652	1	2009	91	99	8
694	SH0030 K	0652	1.5	2009	98	97	1
695	SH0036 K	0540	0	2009	81	93	12
696	SH0036 K	0540	0.5	2009	95	95	0
697	SH0036 K	0540	1	2009	91	92	1
698	SH0036 K	0540	1.5	2009	90	91	1
699	SH0036 K	0546	0	2009	92	92	0
700	SH0036 K	0546	0.5	2009	97	97	0
701	SH0036 K	0546	1	2009	93	94	1
702	SH0036 K	0546	1.5	2009	93	93	0
703	SH0036 L	0558	0	2009	100	100	0
704	SH0036 L	0558	0.5	2009	100	100	0
705	SH0047 L	0412	0	2009	100	100	0
706	SH0047 L	0412	0.5	2009	100	100	0
707	SH0047 L	0412	1	2009	100	100	0
708	SH0047 L	0412	1.5	2009	100	100	0
709	SH0047 L	0414	0	2009	100	100	0
710	SH0047 L	0414	0.5	2009	94	100	6
711	SH0047 L	0414	1	2009	100	100	0
712	SH0047 L	0414	1.5	2009	100	100	0
713	SH0090 K	0416	0	2009	100	100	0
714	SH0090 K	0416	0.5	2009	100	100	0
715	SH0090 K	0416	1.5	2009	71	100	29
716	SH0090 K	0426	0	2009	61	100	39
717	SH0090 K	0426	0.5	2009	61	100	39
718	SH0090 K	0426	1	2009	60	100	40
719	SH0090 K	0426	1.5	2009	62	100	38
720	SH0105 K	0646	0	2009	100	100	0

Rating #	Highway Name	BRM	BRM Distance	Year	Original DS	Audit DS	ΔDS
721	SH0105 K	0646	0.5	2009	94	97	3
722	SH0105 K	0646	1	2009	85	89	4
723	SH0105 K	0646	1.5	2009	79	80	1
724	SH0105 K	0646	1.8	2009	73	75	2
725	SH0105 K	0648	0	2009	75	71	4
726	SH0105 K	0648	0.5	2009	100	100	0
727	SH0105 K	0664	0	2009	100	100	0
728	SH0105 K	0664	0.5	2009	100	100	0
729	SH0105 K	0664	1	2009	100	100	0
730	SH0105 K	0664	1.5	2009	100	100	0
731	SH0105 K	0666	0	2009	100	100	0
732	SH0105 K	0666	1	2009	100	100	0
733	US0077 K	0420	0	2009	74	74	0
734	US0077 K	0420	0.5	2009	100	100	0
735	US0077 K	0420	1	2009	68	66	2
736	US0077 K	0420	1.5	2009	51	51	0
737	US0079 K	0478	0	2009	100	100	0
738	US0079 K	0478	0.5	2009	100	100	0
739	US0079 K	0478	1	2009	100	100	0
740	US0079 K	0478	1.5	2009	100	100	0
741	US0079 K	0492	0.5	2009	100	100	0
742	US0079 K	0492	1	2009	100	100	0
743	US0079 K	0492	1.3	2009	100	100	0
744	US0079 K	0492	1.9	2009	100	100	0
745	US0190 K	0614	0	2009	90	90	0
746	US0190 K	0614	0.5	2009	95	96	1
747	US0190 K	0614	1	2009	96	98	2
748	US0190 K	0614	1.5	2009	99	98	1
749	US0190 K	0622	0	2009	100	100	0
750	US0190 K	0622	0.5	2009	97	99	2
751	US0190 K	0634	0	2009	99	96	3
752	US0190 K	0634	0.5	2009	99	96	3
753	US0190 K	0634	1	2009	100	100	0
754	FM0046 K	0606	0	2008	81	100	19
755	FM0046 K	0606	0.5	2008	76	100	24
756	FM0046 K	0606	1	2008	92	100	8
757	FM0046 K	0606	1.5	2008	67	100	33
758	FM0046 K	0608	0	2008	70	100	30
759	FM0046 K	0608	0.5	2008	99	100	1
760	FM0046 K	0608	1	2008	100	100	0

Rating #	Highway Name	BRM	BRM Distance	Year	Original DS	Audit DS	ΔDS
761	FM0046 K	0608	1.5	2008	100	100	0
762	FM0046 K	0620	0	2008	94	100	6
763	FM0046 K	0620	0.5	2008	94	100	6
764	FM0046 K	0620	1	2008	89	100	11
765	FM0046 K	0620	1.5	2008	94	100	6
766	FM0050 K	0414	0	2008	100	100	0
767	FM0050 K	0414	0.5	2008	100	100	0
768	FM0050 K	0426	0	2008	97	100	3
769	FM0050 K	0426	0.5	2008	93	100	7
770	FM0050 K	0426	1	2008	100	100	0
771	FM0050 K	0426	1.5	2008	97	100	3
772	FM0060 K	0606	0.5	2008	81	100	19
773	FM0060 K	0614	0	2008	86	83	3
774	FM0060 K	0614	0.5	2008	96	97	1
775	FM0060 K	0614	1	2008	91	95	4
776	FM0060 K	0614	1.5	2008	86	91	5
777	FM0060 K	0622	0	2008	69	69	0
778	FM0060 L	0630	1.2	2008	95	99	4
779	FM0060 L	0630	1.7	2008	98	100	2
780	FM0060 R	0630	0.7	2008	98	99	1
781	FM0060 R	0630	1.2	2008	98	98	0
782	FM0060 R	0630	1.7	2008	95	99	4
783	FM0060 R	0634	0.3	2008	86	100	14
784	FM0060 R	0634	0.5	2008	82	98	16
785	FM0060 R	0634	1.6	2008	99	93	6
786	FM0111 K	0426	0	2008	70	99	29
787	FM0111 K	0426	0.5	2008	63	100	37
788	FM0111 K	0426	1	2008	63	100	37
789	FM0111 K	0428	0.5	2008	91	93	2
790	FM0111 K	0428	1.5	2008	98	100	2
791	FM0158 K	0614	0	2008	91	100	9
792	FM0158 K	0614	0.5	2008	96	99	3
793	FM0158 K	0614	1	2008	86	92	6
794	FM0158 R	0618	0.1	2008	93	96	3
795	FM0158 R	0618	0.6	2008	100	100	0
796	FM0158 R	0618	1.1	2008	100	100	0
797	FM0158 R	0620	0	2008	100	100	0
798	FM0158 R	0620	0.5	2008	100	99	1
799	FM0158 R	0620	1	2008	100	100	0
800	FM0158 R	0620	1.5	2008	100	100	0

Rating #	Highway Name	BRM	BRM Distance	Year	Original DS	Audit DS	ΔDS
801	FM0159 K	0428	0	2008	97	100	3
802	FM0159 K	0428	0.5	2008	92	100	8
803	FM0159 K	0428	1	2008	97	100	3
804	FM0159 K	0428	1.5	2008	94	100	6
805	FM0159 K	0434	1	2008	87	100	13
806	FM0159 K	0434	1.5	2008	78	90	12
807	FM0390 K	0460	0	2008	99	100	1
808	FM0390 K	0460	0.5	2008	97	100	3
809	FM0390 K	0460	1	2008	97	100	3
810	FM0390 K	0460	1.5	2008	97	100	3
811	FM0391 K	0400	0	2008	99	100	1
812	FM0391 K	0400	0.5	2008	99	100	1
813	FM0391 K	0400	1	2008	99	100	1
814	FM0391 K	0400	1.5	2008	100	100	0
815	FM0485 K	0586	0	2008	100	100	0
816	FM0485 K	0586	0.5	2008	100	100	0
817	FM0485 K	0586	1	2008	100	100	0
818	FM0485 K	0586	1.5	2008	100	100	0
819	FM0485 K	0606	0	2008	98	100	2
820	FM0485 K	0606	0.5	2008	97	100	3
821	FM0486 K	0500	0	2008	100	100	0
822	FM0486 K	0500	1	2008	99	100	1
823	FM0486 K	0500	1.5	2008	99	100	1
824	FM0486 K	0510	0	2008	97	100	3
825	FM0486 K	0510	0.5	2008	97	100	3
826	FM0486 K	0510	1	2008	100	100	0
827	FM0486 K	0510	1.5	2008	97	100	3
828	FM0487 K	0574	0	2008	92	100	8
829	FM0487 K	0574	0.5	2008	99	100	1
830	FM0487 K	0574	1	2008	94	96	2
831	FM0487 K	0574	1.5	2008	98	99	1
832	FM0908 K	0568	0	2008	99	100	1
833	FM0908 K	0568	0.5	2008	100	100	0
834	FM0908 K	0568	1	2008	100	100	0
835	FM0908 K	0568	1.5	2008	100	100	0
836	FM0974 K	0634	0	2008	100	100	0
837	FM0974 K	0634	0.5	2008	100	100	0
838	FM0974 K	0634	1	2008	100	100	0
839	FM0979 K	0590	0	2008	100	100	0
840	FM0979 K	0590	0.5	2008	100	100	0

Rating #	Highway Name	BRM	BRM Distance	Year	Original DS	Audit DS	ΔDS
841	FM0979 K	0590	1	2008	100	100	0
842	FM0979 K	0590	1.5	2008	100	100	0
843	FM0979 K	0600	0	2008	97	100	3
844	FM0979 K	0600	0.5	2008	97	100	3
845	FM0979 K	0600	1	2008	100	100	0
846	FM0979 K	0600	1.5	2008	99	100	1
847	FM0979 K	0608	0	2008	100	100	0
848	FM0979 K	0608	0.5	2008	97	100	3
849	FM0979 K	0608	1.5	2008	93	100	7
850	FM0979 K	0624	0	2008	89	90	1
851	FM0979 K	0624	0.5	2008	99	100	1
852	FM0979 K	0624	1	2008	100	100	0
853	FM0979 K	0624	1.5	2008	90	90	0
854	FM1155 K	0434	1	2008	99	100	1
855	FM1179 K	0418	0	2008	98	100	2
856	FM1179 K	0418	0.5	2008	98	100	2
857	FM1179 K	0418	1	2008	99	100	1
858	FM1331 K	0568	0	2008	87	100	13
859	FM1331 K	0568	0.5	2008	70	100	30
860	FM1331 K	0568	1	2008	85	100	15
861	FM1331 K	0568	1.5	2008	76	100	24
862	FM1361 K	0614	0	2008	100	100	0
863	FM1361 K	0614	0.5	2008	100	100	0
864	FM1361 K	0614	1	2008	99	100	1
865	FM1361 K	0614	1.5	2008	100	99	1
866	FM1371 K	0450	0	2008	100	100	0
867	FM1371 K	0450	0.5	2008	96	100	4
868	FM1371 K	0450	1	2008	97	100	3
869	FM1371 K	0450	1.5	2008	87	100	13
870	FM1444 K	0582	0	2008	99	100	1
871	FM1444 K	0582	0.5	2008	94	100	6
872	FM1444 K	0582	1	2008	99	100	1
873	FM1444 K	0582	1.5	2008	82	100	18
874	FM1644 K	0400	0	2008	100	100	0
875	FM1644 K	0400	0.5	2008	100	100	0
876	FM1644 K	0400	1	2008	100	100	0
877	FM1644 K	0400	1.5	2008	100	100	0
878	FM1948 K	0610	0	2008	99	100	1
879	FM1948 K	0610	0.5	2008	93	100	7
880	FM1948 K	0610	1	2008	100	100	0

Rating #	Highway Name	BRM	BRM Distance	Year	Original DS	Audit DS	ΔDS
881	FM1948 K	0610	1.5	2008	97	100	3
882	FM2038 K	0622	0	2008	100	100	0
883	FM2038 K	0622	0.5	2008	98	100	2
884	FM2038 K	0622	1	2008	100	100	0
885	FM2154 K	0624	0	2008	87	90	3
886	FM2154 K	0624	0.5	2008	87	99	12
887	FM2154 K	0624	1	2008	99	100	1
888	FM2154 K	0624	1.5	2008	100	100	0
889	FM2159 K	0380	0.5	2008	99	100	1
890	FM2223 K	0402	0	2008	100	100	0
891	FM2223 K	0402	0.5	2008	100	100	0
892	FM2223 K	0402	1	2008	100	100	0
893	FM2223 K	0402	1.5	2008	100	100	0
894	FM2269 K	0394	0.5	2008	100	100	0
895	FM2269 K	0394	1.5	2008	97	100	3
896	FM2269 K	0396	0	2008	100	100	0
897	FM2269 K	0396	0.5	2008	100	90	10
898	FM2269 K	0396	1	2008	100	100	0
899	FM2269 K	0396	1.5	2008	99	100	1
900	FM2293 K	0600	0	2008	94	100	6
901	FM2293 K	0600	0.5	2008	91	100	9
902	FM2293 K	0600	1	2008	87	100	13
903	FM2293 K	0600	1.5	2008	94	100	6
904	FM2293 K	0604	0	2008	79	100	21
905	FM2293 K	0604	0.5	2008	73	100	27
906	FM2293 K	0604	1	2008	93	100	7
907	FM2293 K	0604	1.5	2008	91	100	9
908	FM2347 L	0424	1.5	2008	63	73	10
909	FM2446 K	0610	0	2008	99	100	1
910	FM2446 K	0610	0.5	2008	100	100	0
911	FM2446 K	0610	1	2008	97	100	3
912	FM2446 K	0610	1.5	2008	99	100	1
913	FM2446 K	0616	0	2008	98	100	2
914	FM2446 K	0616	0.5	2008	99	100	1
915	FM2446 K	0616	1	2008	87	100	13
916	FM2549 K	0392	0	2008	99	100	1
917	FM2549 K	0392	0.5	2008	97	100	3
918	FM2549 K	0392	1	2008	100	100	0
919	FM2549 K	0392	1.5	2008	100	100	0
920	FM2621 K	0438	0	2008	97	100	3

Rating #	Highway Name	BRM	BRM Distance	Year	Original DS	Audit DS	ΔDS
921	FM2621 K	0438	0.5	2008	99	100	1
922	FM2621 K	0438	1	2008	99	100	1
923	FM2621 K	0438	1.5	2008	99	100	1
924	FM2818 L	0410	0.1	2008	96	93	3
925	FM2818 L	0410	0.6	2008	96	86	10
926	FM2818 L	0410	0.9	2008	100	100	0
927	FM2818 L	0410	1.4	2008	100	100	0
928	FM2818 L	0412	1.2	2008	96	100	4
929	FM2935 K	0444	0	2008	97	100	3
930	FM2935 K	0444	0.5	2008	99	100	1
931	FM2935 K	0444	1	2008	99	100	1
932	FM3058 K	0570	0	2008	94	100	6
933	FM3058 K	0570	0.5	2008	100	100	0
934	FM3061 K	0404	0	2008	100	100	0
935	FM3061 K	0404	1.5	2008	100	100	0
936	FM3242 K	0406	0	2008	90	90	0
937	FM3242 K	0406	0.5	2008	54	54	0
938	FM3242 K	0406	1	2008	54	54	0
939	FM3242 K	0406	1.5	2008	99	100	1
940	FM3242 K	0408	0	2008	100	100	0
941	FM3242 K	0408	0.5	2008	99	100	1
942	FM3242 K	0408	1	2008	100	100	0
943	SH0007 K	0610	0	2008	82	83	1
944	SH0007 K	0610	0.5	2008	74	84	10
945	SH0007 K	0610	1	2008	67	93	26
946	SH0007 K	0610	1.5	2008	67	87	20
947	SH0021 R	0626	0	2008	100	99	1
948	SH0021 R	0626	0.5	2008	100	100	0
949	SH0021 R	0626	1	2008	100	100	0
950	SH0021 R	0626	1.5	2008	99	99	0
951	SH0030 K	0622	0	2008	91	100	9
952	SH0030 K	0622	0.5	2008	95	99	4
953	SH0030 K	0622	1	2008	90	100	10
954	SH0030 K	0622	1.5	2008	92	98	6
955	SH0036 K	0548	0	2008	100	100	0
956	SH0036 K	0548	0.5	2008	98	100	2
957	SH0036 K	0548	1	2008	97	100	3
958	SH0036 K	0552	0	2008	97	100	3
959	SH0036 K	0552	1	2008	96	91	5
960	SH0036 K	0552	1.5	2008	100	99	1

Rating #	Highway Name	BRM	BRM Distance	Year	Original DS	Audit DS	ΔDS
961	SH0036 L	0558	0	2008	94	100	6
962	SH0036 L	0558	0.5	2008	94	99	5
963	US0077 K	0442	0	2008	100	100	0
964	US0077 K	0442	0.5	2008	100	100	0
965	US0077 K	0442	1	2008	100	100	0
966	US0077 K	0442	1.5	2008	100	100	0
967	US0079 K	0468	0	2008	96	100	4
968	US0079 K	0468	0.5	2008	95	97	2
969	US0079 K	0468	1	2008	94	98	4
970	US0079 K	0468	1.5	2008	88	88	0
971	US0079 K	0484	0	2008	100	100	0
972	US0079 K	0484	0.5	2008	100	100	0
973	US0079 K	0484	1	2008	100	100	0
974	US0079 K	0520	0	2008	100	96	4
975	US0079 K	0520	0.5	2008	99	98	1
976	US0079 K	0520	1	2008	99	100	1
977	US0079 K	0520	1.5	2008	100	100	0
978	US0190 K	0634	0	2008	99	100	1
979	US0190 K	0634	0.5	2008	100	100	0
980	US0190 K	0634	1	2008	99	100	1
981	US0190 K	0662	0	2008	60	100	40
982	US0190 K	0662	0.5	2008	59	100	41
983	US0190 K	0662	1	2008	57	100	43
984	US0190 K	0688	0.5	2008	100	100	0
985	US0190 K	0688	1	2008	100	100	0
986	US0190 R	0668	0.3	2008	93	100	7
987	US0190 R	0668	0.8	2008	93	99	6
988	US0190 R	0668	1.3	2008	97	95	2
989	US0190 R	0672	0	2008	82	100	18
990	US0190 R	0672	0.1	2008	83	100	17
991	US0190 R	0672	0.6	2008	90	100	10
992	US0190 R	0672	1.6	2008	88	96	8
993	US0290 L	0664	0	2008	75	100	25
994	US0290 L	0664	0.5	2008	100	100	0
995	US0290 L	0664	1.5	2008	100	100	0
996	US0290 R	0662	0	2008	91	92	1
997	US0290 R	0662	0.5	2008	87	93	6
998	US0290 R	0662	1	2008	88	89	1
999	US0290 R	0662	1.5	2008	97	89	8
1000	US0290 R	0672	0.3	2008	70	93	23

Rating #	Highway Name	BRM	BRM Distance	Year	Original DS	Audit DS	ΔDS
1001	US0290 R	0678	0	2008	100	100	0
1002	US0290 R	0678	0.5	2008	99	100	1
1003	US0290 R	0678	1	2008	99	100	1
1004	US0290 R	0678	1.5	2008	100	100	0
1005	FM0046 K	0602	0	2007	81	100	19
1006	FM0046 K	0602	0.5	2007	100	100	0
1007	FM0046 K	0602	1	2007	100	100	0
1008	FM0046 K	0602	1.5	2007	100	100	0
1009	FM0046 K	0604	0	2007	100	100	0
1010	FM0046 K	0604	0.5	2007	94	100	6
1011	FM0046 K	0620	0	2007	100	100	0
1012	FM0046 K	0620	0.5	2007	100	100	0
1013	FM0046 K	0620	1	2007	100	100	0
1014	FM0046 K	0620	1.5	2007	100	100	0
1015	FM0050 K	0426	0	2007	99	100	1
1016	FM0050 K	0426	0.5	2007	100	100	0
1017	FM0050 K	0426	1	2007	100	100	0
1018	FM0050 K	0426	1.5	2007	100	100	0
1019	FM0050 K	0432	0	2007	100	100	0
1020	FM0050 K	0432	0.5	2007	100	100	0
1021	FM0050 K	0432	1	2007	100	100	0
1022	FM0050 K	0432	1.5	2007	100	100	0
1023	FM0050 K	0436	0	2007	100	100	0
1024	FM0050 K	0436	0.5	2007	100	100	0
1025	FM0050 K	0436	1	2007	100	100	0
1026	FM0050 K	0436	1.5	2007	64	100	36
1027	FM0060 K	0594	0	2007	90	90	0
1028	FM0060 K	0594	0.5	2007	100	100	0
1029	FM0060 K	0594	1	2007	97	100	3
1030	FM0060 K	0594	1.5	2007	100	100	0
1031	FM0060 K	0636	0.9	2007	100	100	0
1032	FM0060 K	0636	1	2007	89	93	4
1033	FM0060 K	0636	1.5	2007	86	99	13
1034	FM0111 K	0426	0	2007	90	90	0
1035	FM0111 K	0426	0.5	2007	100	100	0
1036	FM0111 K	0426	1	2007	46	47	1
1037	FM0158 L	0618	0.1	2007	100	100	0
1038	FM0158 L	0618	0.6	2007	100	100	0
1039	FM0158 L	0618	1.1	2007	100	100	0
1040	FM0158 L	0618	1.6	2007	100	100	0

Rating #	Highway Name	BRM	BRM Distance	Year	Original DS	Audit DS	ΔDS
1041	FM0159 K	0436	0	2007	90	90	0
1042	FM0159 K	0436	0.5	2007	100	100	0
1043	FM0159 K	0436	1	2007	100	100	0
1044	FM0159 K	0436	1.5	2007	100	100	0
1045	FM0166 K	0604	0	2007	100	100	0
1046	FM0166 K	0604	0.5	2007	99	100	1
1047	FM0166 K	0604	1	2007	99	100	1
1048	FM0166 K	0604	1.5	2007	99	100	1
1049	FM0166 K	0608	0	2007	96	89	7
1050	FM0166 K	0608	0.5	2007	100	99	1
1051	FM0166 K	0608	1	2007	100	100	0
1052	FM0166 K	0608	1.5	2007	100	100	0
1053	FM0362 K	0424	0	2007	61	75	14
1054	FM0362 K	0424	0.5	2007	14	36	22
1055	FM0362 K	0424	1	2007	19	50	31
1056	FM0362 K	0424	1.5	2007	23	36	13
1057	FM0391 K	0398	0	2007	100	100	0
1058	FM0391 K	0398	0.5	2007	100	100	0
1059	FM0391 K	0398	1	2007	100	100	0
1060	FM0485 K	0576	0	2007	100	100	0
1061	FM0485 K	0576	0.5	2007	100	100	0
1062	FM0485 K	0582	0	2007	100	100	0
1063	FM0485 K	0582	0.5	2007	100	100	0
1064	FM0485 K	0582	1	2007	90	90	0
1065	FM0485 K	0582	1.5	2007	100	99	1
1066	FM0485 K	0604	0	2007	22	31	9
1067	FM0485 K	0604	0.5	2007	46	69	23
1068	FM0485 K	0604	1	2007	45	66	21
1069	FM0486 K	0518	0	2007	51	68	17
1070	FM0486 K	0518	0.5	2007	97	98	1
1071	FM0486 K	0518	1	2007	98	98	0
1072	FM0486 K	0518	1.5	2007	100	100	0
1073	FM0487 K	0594	0	2007	98	98	0
1074	FM0487 K	0594	0.5	2007	100	100	0
1075	FM0696 K	0598	0	2007	98	100	2
1076	FM0696 K	0598	0.5	2007	100	90	10
1077	FM0696 K	0598	1	2007	99	100	1
1078	FM0696 K	0598	1.5	2007	90	90	0
1079	FM0908 K	0590	0	2007	100	100	0
1080	FM0908 K	0590	0.5	2007	98	100	2

Rating #	Highway Name	BRM	BRM Distance	Year	Original DS	Audit DS	 ΔDS
1081	FM0908 K	0590	1	2007	97	100	3
1082	FM0908 K	0590	1.5	2007	100	100	0
1083	FM0908 K	0598	0	2007	98	99	1
1084	FM0908 K	0598	1	2007	94	91	3
1085	FM0908 K	0598	1.5	2007	99	99	0
1086	FM0974 K	0626	0	2007	100	100	0
1087	FM0974 K	0626	0.5	2007	99	100	1
1088	FM0974 K	0626	1.5	2007	98	100	2
1089	FM0976 K	0598	0.5	2007	100	100	0
1090	FM0976 K	0598	1	2007	100	100	0
1091	FM0976 K	0598	1.5	2007	100	100	0
1092	FM0979 K	0608	0	2007	58	82	24
1093	FM0979 K	0608	0.5	2007	89	78	11
1094	FM0979 K	0608	1	2007	50	64	14
1095	FM0979 K	0608	1.5	2007	80	100	20
1096	FM0979 K	0612	0	2007	88	100	12
1097	FM0979 K	0612	0.5	2007	94	100	6
1098	FM0979 K	0612	1	2007	91	100	9
1099	FM0979 K	0612	1.5	2007	100	100	0
1100	FM0979 K	0628	0	2007	100	100	0
1101	FM0979 K	0628	0.5	2007	100	100	0
1102	FM0979 K	0628	1.5	2007	100	100	0
1103	FM1227 K	0432	0	2007	99	100	1
1104	FM1227 K	0432	0.5	2007	99	100	1
1105	FM1227 K	0432	1	2007	100	100	0
1106	FM1227 K	0432	1.5	2007	99	100	1
1107	FM1227 K	0434	0	2007	100	100	0
1108	FM1227 K	0434	0.5	2007	100	100	0
1109	FM1227 K	0434	1	2007	99	100	1
1110	FM1373 K	0600	0	2007	100	100	0
1111	FM1373 K	0600	0.5	2007	100	100	0
1112	FM1373 K	0600	1	2007	100	100	0
1113	FM1373 K	0604	0	2007	100	100	0
1114	FM1373 K	0608	0	2007	100	100	0
1115	FM1373 K	0608	0.5	2007	100	100	0
1116	FM1373 K	0608	1	2007	100	100	0
1117	FM1373 K	0608	1.5	2007	100	100	0
1118	FM1600 K	0398	0	2007	98	99	1
1119	FM1600 K	0398	0.5	2007	74	100	26
1120	FM1600 K	0398	1	2007	89	99	10

Rating #	Highway Name	BRM	BRM Distance	Year	Original DS	Audit DS	ΔDS
1121	FM1600 K	0398	1.5	2007	77	100	23
1122	FM1687 K	0608	0	2007	99	100	1
1123	FM1687 K	0608	0.5	2007	100	100	0
1124	FM1696 K	0650	0	2007	100	100	0
1125	FM1696 K	0656	0	2007	99	100	1
1126	FM1696 K	0656	1	2007	100	100	0
1127	FM1696 K	0656	1.5	2007	100	100	0
1128	FM1915 K	0396	0	2007	92	64	28
1129	FM1915 K	0396	0.5	2007	65	62	3
1130	FM1915 K	0396	1	2007	96	96	0
1131	FM1915 K	0396	1.5	2007	64	65	1
1132	FM2000 K	0412	0	2007	100	100	0
1133	FM2000 K	0412	0.5	2007	100	100	0
1134	FM2000 K	0412	1	2007	100	100	0
1135	FM2000 K	0412	1.5	2007	100	100	0
1136	FM2000 K	0414	0	2007	100	100	0
1137	FM2000 K	0414	0.5	2007	100	100	0
1138	FM2000 K	0414	1	2007	100	100	0
1139	FM2000 K	0414	1.5	2007	100	100	0
1140	FM2000 K	0416	0	2007	100	100	0
1141	FM2000 K	0416	0.5	2007	100	100	0
1142	FM2000 K	0416	1	2007	100	100	0
1143	FM2000 K	0416	1.5	2007	100	100	0
1144	FM2027 K	0390	0	2007	96	100	4
1145	FM2027 K	0390	0.5	2007	100	100	0
1146	FM2027 K	0390	1	2007	100	100	0
1147	FM2027 K	0390	1.5	2007	100	100	0
1148	FM2096 K	0374	0	2007	100	100	0
1149	FM2096 K	0374	0.5	2007	100	100	0
1150	FM2096 K	0374	1	2007	100	100	0
1151	FM2096 K	0374	1.5	2007	100	100	0
1152	FM2223 K	0406	0	2007	100	100	0
1153	FM2223 K	0406	0.5	2007	100	100	0
1154	FM2223 K	0406	1	2007	100	100	0
1155	FM2223 K	0406	1.5	2007	100	100	0
1156	FM2269 K	0392	0	2007	100	100	0
1157	FM2269 K	0392	0.5	2007	100	100	0
1158	FM2269 K	0392	1	2007	100	100	0
1159	FM2269 K	0392	1.5	2007	100	100	0
1160	FM2446 K	0616	0	2007	99	100	1

Rating #	Highway Name	BRM	BRM Distance	Year	Original DS	Audit DS	ΔDS
1161	FM2446 K	0616	0.5	2007	100	100	0
1162	FM2446 K	0616	1	2007	100	100	0
1163	FM2549 K	0396	0	2007	100	100	0
1164	FM2549 K	0396	0.5	2007	100	100	0
1165	FM2549 K	0396	1	2007	100	100	0
1166	FM2549 K	0396	1.5	2007	100	100	0
1167	FM2562 K	0418	0	2007	100	100	0
1168	FM2562 K	0418	0.5	2007	77	100	23
1169	FM2562 K	0418	1	2007	100	100	0
1170	FM2562 K	0418	1.5	2007	78	100	22
1171	FM2954 K	0602	0	2007	100	100	0
1172	FM2954 K	0602	0.5	2007	100	100	0
1173	FM3090 K	0420	0	2007	100	100	0
1174	FM3090 K	0420	0.5	2007	100	100	0
1175	FM3090 K	0420	1	2007	100	100	0
1176	FM3090 K	0420	1.5	2007	100	100	0
1177	FM3090 K	0422	0	2007	100	100	0
1178	FM3090 K	0422	0.5	2007	100	100	0
1179	FM3090 K	0422	1	2007	100	100	0
1180	FM3090 K	0422	1.5	2007	100	100	0
1181	FM3090 K	0430	0.5	2007	100	100	0
1182	FM3090 K	0430	1.5	2007	100	100	0
1183	FM3242 K	0406	0	2007	100	100	0
1184	FM3242 K	0406	0.5	2007	100	100	0
1185	SH0006 K	0560	0	2007	100	100	0
1186	SH0006 K	0560	0.5	2007	100	100	0
1187	SH0007 K	0610	0	2007	97	97	0
1188	SH0007 K	0610	0.5	2007	92	100	8
1189	SH0007 K	0610	1	2007	72	100	28
1190	SH0007 K	0610	1.5	2007	78	99	21
1191	SH0014 K	0388	0	2007	96	95	1
1192	SH0014 K	0388	0.5	2007	99	99	0
1193	SH0014 K	0388	1	2007	100	100	0
1194	SH0021 K	0622	0	2007	100	100	0
1195	SH0021 K	0640	0	2007	100	100	0
1196	SH0021 K	0640	0.5	2007	99	100	1
1197	SH0021 K	0640	1	2007	67	100	33
1198	SH0021 K	0640	1.5	2007	66	100	34
1199	SH0021 K	0640	1.8	2007	63	99	36
1200	SH0021 R	0626	0	2007	83	93	10

Rating #	Highway Name	BRM	BRM Distance	Year	Original DS	Audit DS	ΔDS
1201	SH0021 R	0626	0.5	2007	100	98	2
1202	SH0021 R	0626	1	2007	99	100	1
1203	SH0021 R	0626	1.5	2007	100	100	0
1204	SH0021 R	0634	0	2007	100	100	0
1205	SH0021 R	0634	0.5	2007	100	100	0
1206	SH0021 R	0634	1	2007	100	100	0
1207	SH0021 R	0634	1.5	2007	100	100	0
1208	SH0021 R	0638	0	2007	100	100	0
1209	SH0021 R	0638	0.5	2007	98	100	2
1210	SH0021 R	0638	1	2007	99	100	1
1211	SH0036 K	0534	0	2007	100	100	0
1212	SH0036 K	0534	0.5	2007	99	97	2
1213	SH0036 K	0534	1	2007	97	95	2
1214	SH0036 K	0534	1.5	2007	94	94	0
1215	SH0036 K	0548	0	2007	100	99	1
1216	SH0036 K	0548	0.5	2007	99	100	1
1217	SH0036 K	0548	1	2007	100	100	0
1218	SH0036 K	0548	1.5	2007	100	100	0
1219	SH0047 L	0412	0	2007	100	100	0
1220	SH0047 L	0412	0.5	2007	100	100	0
1221	SH0047 L	0412	1	2007	100	100	0
1222	SH0047 L	0412	1.5	2007	100	100	0
1223	SH0090 K	0398	0	2007	100	100	0
1224	SH0090 K	0398	0.5	2007	85	100	15
1225	SH0090 K	0398	1	2007	100	100	0
1226	SH0090 K	0398	1.5	2007	100	100	0
1227	SH0090 K	0404	0	2007	100	100	0
1228	SH0090 K	0404	0.5	2007	100	100	0
1229	SH0090 K	0404	1	2007	100	100	0
1230	SH0090 K	0404	1.5	2007	100	100	0
1231	SH0090 K	0412	0.5	2007	99	99	0
1232	SH0090 K	0412	1	2007	100	89	11
1233	SH0090 K	0418	0	2007	100	100	0
1234	SH0090 K	0418	0.5	2007	100	100	0
1235	SH0090 K	0418	1	2007	100	100	0
1236	SH0090 K	0418	1.5	2007	100	100	0
1237	SH0090 K	0428	0.5	2007	100	100	0
1238	SH0090 K	0428	1	2007	82	100	18
1239	SH0090 K	0428	1.5	2007	100	100	0
1240	SH0105 K	0644	0	2007	100	100	0

Rating #	Highway Name	BRM	BRM Distance	Year	Original DS	Audit DS	ΔDS
1241	SH0105 K	0644	0.5	2007	100	100	0
1242	SH0105 K	0644	1	2007	100	100	0
1243	SH0105 K	0644	1.5	2007	100	100	0
1244	SH0105 K	0656	0	2007	96	70	26
1245	SH0105 K	0656	0.5	2007	98	68	30
1246	SH0105 K	0656	1	2007	85	98	13
1247	SH0105 K	0656	1.5	2007	99	95	4
1248	US0077 K	0416	0	2007	100	100	0
1249	US0077 K	0416	0.5	2007	100	100	0
1250	US0077 K	0416	1	2007	100	100	0
1251	US0077 K	0416	1.5	2007	100	100	0
1252	US0077 K	0446	0	2007	100	100	0
1253	US0077 K	0446	0.5	2007	100	100	0
1254	US0077 K	0446	1	2007	100	100	0
1255	US0077 K	0446	1.5	2007	99	100	1
1256	US0079 K	0508	0	2007	100	100	0
1257	US0079 K	0508	0.5	2007	100	100	0
1258	US0079 K	0508	1	2007	100	100	0
1259	US0079 K	0508	1.5	2007	99	99	0
1260	US0079 K	0526	0	2007	99	99	0
1261	US0079 K	0526	0.5	2007	100	100	0
1262	US0079 K	0526	1	2007	100	100	0
1263	US0190 K	0608	0	2007	96	98	2
1264	US0190 K	0608	0.5	2007	100	100	0
1265	US0190 K	0608	1.5	2007	100	100	0
1266	US0190 K	0634	0	2007	100	100	0
1267	US0190 K	0634	0.5	2007	100	100	0
1268	US0190 K	0634	1	2007	100	100	0
1269	US0190 L	0674	0	2007	97	93	4
1270	US0190 L	0674	0.5	2007	100	99	1
1271	US0190 L	0674	1	2007	93	91	2
1272	US0190 L	0674	1.5	2007	94	93	1

**APPENDIX C: CODE OF COMPUTER PROGRAM DEVELOPED FOR
IMPUTING PAVEMENT NETWORK REPAIR DATA**

```
# -*- coding: utf-8 -*-

import arcpy
import os
import math

#try:
    #arcpy.AddField_management("C:/Studies/PhD/Work/20090912-PMIS
prediction    preciosion/Data/Raw-Data/Shapefiles/AllData.mdb/DS10_9",
"Probab", "DOUBLE")
    #arcpy.AddField_management("C:/Studies/PhD/Work/20090912-PMIS
prediction    preciosion/Data/Raw-Data/Shapefiles/DS10_9.shp",    "Prob",
"DOUBLE")
    #arcpy.AddField_management("C:/Studies/PhD/Work/20090912-PMIS
prediction    preciosion/Data/Raw-Data/Shapefiles/DS10_9.shp",    "RNum",
"DOUBLE")
    #arcpy.AddField_management("C:/Studies/PhD/Work/20090912-PMIS
prediction    preciosion/Data/Raw-Data/Shapefiles/DS10_9.shp",    "RProb",
"DOUBLE")
    #arcpy.AddField_management("C:/Studies/PhD/Work/20090912-PMIS
prediction    preciosion/Data/Raw-Data/Shapefiles/DS10_9.shp",    "LNum",
"DOUBLE")
    #arcpy.AddField_management("C:/Studies/PhD/Work/20090912-PMIS
prediction    preciosion/Data/Raw-Data/Shapefiles/DS10_9.shp",    "LProb",
"DOUBLE")
#except:
    #print "Field already there"

#SETTINGS
ThisSecFID=0          #Delete this
count=0              #Delete this
#MyShapefile          =          "C:/Siamak/Studies/PhD/Thesis/1-
WorkAgeEstimation/Data/GISData/NDS10_9_17BMTModels2.shp"
MyShapefile          =          "C:/Siamak/Studies/PhD/Thesis/1-
WorkAgeEstimation/Data/GISData/RandomErrorFile/Shapefiles/Error0.shp"
#DS1Col = "DS2009"
#DS2Col = "DISTRESS_S"
#DS21Col = "DS10_9"
DS1Col = "EDS9"
DS2Col = "EDS10"
DS21Col = "EDS10_9"
VertexTolerance=0.0001
MinShapeLength=0.0001
TotalSecNum = 7234
CheckSecNum = 30          #Maximum number of section that are used to
calculate prob.
```

```
PickedPercent = 7.13    #Percentage of pavement sections that we think
receive M&R annually
```

```
#DEFINITION OF VARIABLES
```

```
DS1=0
DS2=0
RowNum=0
ObjIDrow=98
ObjIDrow2=98
StartX1=0
StartY1=0
EndX1=0
EndY1=0
LeftProb=0
RightProb=0
nRight=1
```

```
#clear dataset
ClearAllRows = arcpy.UpdateCursor(MyShapefile, "")
for ClearAllRow in ClearAllRows:
    ClearAllRow.CalcProb = 0
    ClearAllRows.updateRow(ClearAllRow)
print "the dataset CalcProb column is all zero now."
```

```
AllRows = arcpy.UpdateCursor(MyShapefile, "Shape_Leng > " +
str(MinShapeLength))
MyExpression = ""
HighwayName = ""
MaxProbArray = [ [ 0 for i in range(3) ] for j in range(TotalSecNum) ]
#Column1: MaxProb----Column2: ProjectLength
NotRankedArray = [ [ 0 for i in range(3) ] for j in range(TotalSecNum)
]
a1=0
a2=0
a3=0
a4=0
a5=0
a6=0
```

```
# P(R|B)
SecLength=0;
def RB(nSec):
    Length=0.5*nSec
    Probability=((Length+0.25)**2.05)/((0.84*(Length+0.25)**2.09+29.3))-
((Length-0.25)**2.05)/((0.84*(Length-0.25)**2.09+29.3))
```

```

    return Probability

# P(DS|B)
def DSB(DS):
    #Probability=((1.0-(1.00-(DS/100.00)**2.919)**0.215)-(1.00-(1.00-
    ((DS-1.00)/100)**2.919)**0.215))

    #Second Model
    Probability=1
    return Probability

# P(DS|R,B)
def DSRB(DS):
    #Probability=(1.00-(1.00-(DS/100.00)**2.956)**0.526)-(1.00-(1.00-
    ((DS-1.00)/100.00)**2.956)**0.526)

    #Second Model based on Bryan Data
    #Probability=28.916*2.718**(-0.039*DS)

    #Second Model based on Beaumont Data
    Probability=-0.0293*DS + 3.794

    return Probability

# P( $\Delta$ |DS, B)
def DeltaDSB(DS, Delta):
    #if (Delta>0):
        #Probability=((7.46/(19.72+3.75*Delta))*(1.00-(1.00-
        (DS/(100.00-Delta)**5.23)**0.49))-((7.46/(19.72+3.75*(Delta-
        1.00)))*(1.00-(1.00-((DS-1.00)/(100-(Delta-1.00)))**5.23)**0.49))
    #elif (Delta==0):
        #Probability=((7.46/(19.72+3.75*Delta))*(1.00-(1.00-
        (DS/(100.00-Delta)**5.23)**0.49))-((7.46/(19.72+3.75*(Delta)))*(1-
        (1.00-((DS-1.00)/(100-(Delta)))**5.23)**0.49))
    #else:
        #Probability=((14.01/(18.7-7.43*Delta))*(1-(1.00-
        ((DS+Delta)/(100.00+Delta)**7.25)**0.25))-((14.01/(18.7-7.43*(Delta-
        1)))*(1-(1.00-((DS-1.00+Delta-1.00)/(100.00+(Delta-
        1.00)))**7.25)**0.25))

    #Second Model
    Probability=1
    return Probability

# P( $\Delta$ |DS,R,B)
def DeltaDSRB(DS, Delta):
    #if (Delta>0):
        #Probability=((0.28*DS**2.00/(30130.00+DS))*(1-(1.00-
        ((DS+Delta)/(100.00))**0.57)**0.23))-((0.28*(DS-
```

```

1.00)**2.00/(30130.00+(DS-1.00)))*(1-(1.00-(((DS-1.00)+(Delta-
1.00))/(100.00))**0.57)**0.23))
    #elif (Delta==0):
        #Probability=((0.28*DS**2.00/(30130.00+DS))*(1-(1.00-
((DS+Delta)/(100.00))**0.57)**0.23))-((0.28*(DS-
1.00)**2.00/(30130.00+(DS-1.00)))*(1-(1.00-(((DS-
1.00)+(Delta))/(100.00))**0.57)**0.23))
    #else:
        #Probability=
0.59/(24.98-8.56*Delta)*(1.00-(1.00-
((DS+Delta)/(100.0+Delta))**6.88)**10.95)-0.59/(24.98-8.56*(Delta-
1.0))*(1.00-(1.00-((DS-1.00+Delta-1.00)/(100+Delta-1.0))**6.88)**10.95)

#Second Model
#Probability=(132.7-1.08*Delta)*2.718**(-0.032*DS)
if (Delta>=0):
    if (DS<100):
        #Based on Bryan Data:
        #Probability=(-0.17059*DS+17.65949)+ (Delta/(100.00-DS))*
(3.185795*(141.421-1.414*DS)**0.320972+0.6+0.17059*DS-17.65949)
        #Based on Beaumont Data: Not so bad, but not good
        #Probability=(-0.534*DS+54.135)+ (Delta/(100.00-DS))*
(0.07*(141.421-1.414*DS)**0.974+1.5+0.535*DS-54.13)
        #Based on Beaumont Data
        Probability=(-0.00069*DS+1.3463)+ (Delta/(100.00-DS))*
(1.4109*(141.421-1.414*DS)**0.2307+0.0069*DS-1.3463)
    else:
        #Based on Bryan Data:
        #Probability=0.6
        #Based on Beaumont Data:
        Probability=0.65

else:
    #Based on Bryan Data:
    #Probability= 0.4014*2.71828**(0.0721*Delta)
    #Based on Beaumont Data:
    #Probability= 1.6667*2.71828**(0.0921*Delta)
    #Based on Beaumont Data (after adding Zeros) : it gives high
probability to sections with lows DS but, very negative Delta
    #Probability=-0.163*math.log(-Delta)+0.5879
    #Theoretical Model
    Probability=0.6661*(-Delta)**-0.767
return Probability

def SingleProb(nSec,DS,Delta):
    Probability=0
    #try:
    Probability = RB(nSec)*(DSRB(DS)/DSB(DS))*(DeltaDSRB(DS,
Delta)/DeltaDSB(DS, Delta))
    #except:
    #print "DS= '"+ str(DS)+"' nSec '"+ str(nSec) + "' Delta='"+
str(Delta) + "' P(R|B) = '"+ str(RB(nSec)) + "' P(DS|B)= '"+
str(DSB(DS)) + "' P(DS|R,B)= '"+ str(DSRB(DS)) + "' P(Delta|DS, B)='"+

```

```

str(DeltaDSB(DS, Delta)) + "' P(Delta|DS,R,B) '" + str(DeltaDSRB(DS,
Delta))+ "' "
    return Probability

for row in AllRows:

#calculate right hand sections
if ((row.getValue(DS1Col)>0) and (row.getValue(DS2Col)>0)):

    #The section itself
    nSec=1
    ObjIDrow=row.getValue("FID")
    SecDelta= row.getValue(DS21Col)
    SecDS= row.getValue(DS1Col)

    ThisSecProb = SingleProb(nSec,SecDS,SecDelta)
    MaxProb = ThisSecProb
    nSecMaxProb = 1
    MaxProbArray[ObjIDrow][0]=MaxProb
    MaxProbArray[ObjIDrow][1]=nSecMaxProb
    MaxProbArray[ObjIDrow][2]=ObjIDrow

    EndX1=row.getValue("EndX")
    EndY1=row.getValue("EndY")
    HighwayName = row.getValue("SIGNED_HIG")
    IsSec = True
    while ((nSec<=CheckSecNum)and(IsSec)):
        MyExpression = "SIGNED_HIG = '" + str(HighwayName) + "' AND
Shape_Leng > " + str(MinShapeLength)+ " AND ABS(StartX - (" +
str(EndX1)+ "))<=" + str(VertexTolerance) + " AND ABS(StartY - (" +
str(EndY1)+")) <=" + str(VertexTolerance)
        AdjacentRows = arcpy.SearchCursor(MyShapefile, MyExpression)

        count = 0
        for AdjacentRow in AdjacentRows:
            count += 1
            ObjIDrow2=AdjacentRow.getValue("FID")
            SecDelta= AdjacentRow.getValue(DS21Col)
            SecDS= AdjacentRow.getValue(DS1Col)
            EndX1=AdjacentRow.getValue("EndX")
            EndY1=AdjacentRow.getValue("EndY")
        if count==1:
            AdjacentRows = arcpy.SearchCursor(MyShapefile, MyExpression)
            for AdjacentRow in AdjacentRows:
                nSec = nSec+1
                SecDelta= AdjacentRow.getValue(DS21Col)
                SecDS= AdjacentRow.getValue(DS1Col)
                if (SecDS>0):
                    ThisSecProb=(ThisSecProb/RB(nSec-
1))*SingleProb(nSec,SecDS,SecDelta)

```

```

        else:
            ThisSecProb=ThisSecProb

    if (ThisSecProb>=MaxProb):
        MaxProb = ThisSecProb
        nSecMaxProb = nSec
        MaxProbArray[ObjIDrow][0]=MaxProb
        MaxProbArray[ObjIDrow][1]=nSecMaxProb
    elif count==0:
        IsSec=False
        RightProb=0
        #print "Section " + str(ObjIDrow2)+ "Does not have any
neighbor section on the right"
    else:
        IsSec=True
        print str(count) + "branches in right side of row " +
str(ObjIDrow2)
        # choose the branch with the same highway name
    if (MaxProb>MaxProbArray[ObjIDrow][0]):
        MaxProbArray[ObjIDrow][0]=MaxProb
        MaxProbArray[ObjIDrow][1]=nSecMaxProb
        MaxProbArray[ObjIDrow][2]=ObjIDrow

    RowNum=RowNum+1
    if (100*round(RowNum/100)== RowNum):
        print str(RowNum)
print "Array created. Now, it is gonna update it."

ii=0
while (ii<=RowNum-1):
    NotRankedArray[ii][0]=MaxProbArray[ii][0]
    NotRankedArray[ii][1]=MaxProbArray[ii][1]
    NotRankedArray[ii][2]=MaxProbArray[ii][2]
    ii=ii+1

#Pick First 10%
TotalPickNum=TotalSecNum*PickedPercent/100.00
MaxProbArray.sort() #sort ascendingbased on the first column

iProj=1
PickedSecNum = 0
count=0
while ((PickedSecNum<=TotalPickNum-1)):
    ThisProjectProb=MaxProbArray[TotalSecNum-iProj][0]
    ThisProjectNum=MaxProbArray[TotalSecNum-iProj][1]
    ThisObjID=MaxProbArray[TotalSecNum-iProj][2]

    iProj=iProj+1

    FirstRowExpression = "FID = " + str(ThisObjID)

```



```

FirstSecRows          =          arcpy.UpdateCursor(MyShapefile,
FirstRowExpression)
iSec=0
for FirstSecRow in FirstSecRows:          #just one section

    ThisSecFID=FirstSecRow.FID
    if (FirstSecRow.CalcProb<=ThisProjectProb):
        try:
            NotRankedArray[ThisSecFID][0] = ThisProjectProb
            FirstSecRow.CalcProb = ThisProjectProb
            FirstSecRows.updateRow(FirstSecRow)
            PickedSecNum=PickedSecNum+1
            iSec=iSec+1
        except:
            print "Error in updating data!"

    EndX1=FirstSecRow.getValue("EndX")
    EndY1=FirstSecRow.getValue("EndY")
    HighwayName = FirstSecRow.getValue("SIGNED_HIG")
    nSec=0
    IsSec = True
    while ((iSec<=ThisProjectNum-1)and(IsSec)):
        MyExpression = "SIGNED_HIG = '" + str(HighwayName) +
        "' AND Shape_Leng > " + str(MinShapeLength)+ " AND ABS(StartX - (" +
        str(EndX1)+ "))<= " + str(VertexTolerance) + " AND ABS(StartY - (" +
        str(EndY1)+")) <= " + str(VertexTolerance)
        #print "My Expression=" + MyExpression
        AdjacentRows          =          arcpy.SearchCursor(MyShapefile,
MyExpression)

        count=0
        for AdjacentRow in AdjacentRows:
            count=count+1
        AdjRows          =          arcpy.UpdateCursor(MyShapefile,
MyExpression)

        #print "count =" + str(count)
        if (count>0):
            for AdjRow in AdjRows:
                #ThisSecProb=AdjRow.CalcProb
                ThisSecFID=AdjRow.FID          #To be
deleted finally

                ThisSecProb=NotRankedArray[ThisSecFID][0]
                EndX1=AdjRow.getValue("EndX")
                EndY1=AdjRow.getValue("EndY")
                HighwayName = AdjRow.getValue("SIGNED_HIG")
                if (ThisSecProb<=ThisProjectProb):
                    try:
                        NotRankedArray[ThisSecFID][0]          =
ThisProjectProb

                        AdjRow.CalcProb = ThisProjectProb
                        AdjRows.updateRow(AdjRow)

```

```
        PickedSecNum=PickedSecNum+1
    except:
        print "Error in updating data!"
    else:
        IsSec=False
        iSec=iSec+1

print "done!"
```