DATABASE DEVELOPMENT FOR AN HMA PAVEMENT PERFORMANCE ANALYSIS SYSTEM

Project 07-11
April 2008

Midwest Regional University Transportation Center
College of Engineering
Department of Civil and Environmental Engineering
University of Wisconsin, Madison

Authors:
Robert L. Schmitt, University of Wisconsin, Platteville
Samuel Owusu-Ababio, University of Wisconsin, Platteville
Kevin D. Denn, University of Wisconsin, Platteville

Principal Investigator:
Robert L. Schmitt, Associate Professor, Department of Civil and Environmental Engineering, University of Wisconsin, Platteville

Technical Report Documentation Page

1. Report No. 07-11	2. Government Accession No.	3. Recipient's Catalog No. CFDA 20.701

4. Title and Subtitle Database Development for an HMA Pavement Performance Analysis System	5. Report Date: December 2007 6. Performing Organization Code
7. Author/s Schmitt, Robert L., Owusu-Ababio, Sam , and Denn, Kevin	8. Performing Organization Report No. MRUTC 07-11
9. Performing Organization Name and Address Midwest Regional University Transportation Center University of Wisconsin-Madison 1415 Engineering Drive, Madison, WI 53706	10. Work Unit No. (TRAIS) 11. Contract or Grant No. DTRS99-G-0005
12. Sponsoring Organization Name and Address Research and Innovative Technology Administration United States Department of Transportation 1200 New Jersey Avenue, SE Washington, D.C. 20590	13. Type of Report and Period Covered Final Report [November 2006 to March 2009] 14. Sponsoring Agency Code

15. Supplementary Notes
Project completed for the Midwest Regional University Transportation Center with support from the Wisconsin Department of Transportation.

16. Abstract

The primary purpose of this report was to develop a database template, using the existing Wisconsin DOT pavement management system, from which to perform pavement performance analysis using design, construction, and performance data for hot-mix asphaltic pavements. A second purpose was to investigate appropriate numerical or statistical methods that have the potential of quantifying and establishing relationships between design, construction, and performance data. A series of tasks were conducted including a review of literature, review of Wisconsin DOT databases, database integration with emphasis on performance modeling, and recommended approaches for performance modeling.

The literature review found that data types collected for performance evaluation and modeling vary among agencies depending on needs, but the most common types include inventory, condition, traffic volume, and maintenance and rehabilitation. Common referencing systems between various data collection systems can facilitate data integration for pavement performance modeling, however, a major barrier for achieving full data integration is lack of common referencing systems compounded by the use of different data formats. To that end, Geographic Information System (GIS) was identified as an effective tool for data integration among various divisions within an organization.

Several Wisconsin DOT databases applicable to performance modeling for hot-mix asphaltic pavements were reviewed for primary data categories including construction, design, traffic, and performance. Semantic discrepancies among databases that impede integration were summarized, then recommendations were identified to enable simple or complex queries to relate data residing in the different databases. A GIS-based database integration was recommended using similar Wisconsin DOT GIS practices. A loose coupling approach, involving the transfer of data files between the GIS and other programs, was demonstrated using screen snapshots. Then, the integrated data were prepared for export into a statistical analysis package and the results imported back to the GIS for data visualization or display.

Several statistical analysis methods to develop performance models were provided, along with reference examples for ANOVA, comparison of means, and regression models. Currently, there is an on-going research study with estimated completion in 2008, NCHRP Project 9-22, *Beta Testing and Validation of HMA PRS*, that will develop software capable of developing pavement performance models.

17. Key Words Performance Models, HMA, GIS, Integration	18. Distribution Statement No restrictions. This report is available through the Transportation Research Information Services of the National Transportation Library.		
19. Security Classification (of this report) Unclassified	20. Security Classification (of this page) Unclassified	21. No. Of Pages **55**	22. Price **-0-**

Form DOT F 1700.7 (8-72) **Reproduction of form and completed page is authorized.**

DISCLAIMER

Executive Summary

The primary purpose of this report was to develop a database template, using the existing Wisconsin DOT pavement management system, from which to perform pavement performance analysis using design, construction, and performance data for hot-mix asphaltic pavements. A second purpose was to investigate appropriate numerical or statistical methods that have the potential of quantifying and establishing relationships between design, construction, and performance data. A series of tasks were conducted including a review of literature, review of Wisconsin DOT databases, database integration with emphasis on performance modeling, export of integrated database for performance modeling, and recommended approaches for performance modeling.

The literature review found that data types collected for performance evaluation and modeling vary from agency to agency depending on needs but the most common ones include inventory, condition, traffic volume, and maintenance and rehabilitation. Common referencing systems between various data collection systems can facilitate data integration for pavement performance modeling, however, a major barrier for achieving full data integration is lack of common referencing systems compounded by the use of different data formats. To that end, Geographic Information System (GIS) was identified as an effective tool for data integration among various divisions within an organization.

Several Wisconsin DOT databases applicable to performance modeling for hot-mix asphaltic pavements were reviewed for primary data categories including construction, design, traffic, and performance. Semantic discrepancies among databases that impede integration were summarized, then recommendations were identified to enable simple or complex queries to relate data residing in the different databases. A GIS-based database integration was recommended using similar Wisconsin DOT GIS practices. A loose coupling approach, involving the transfer of data files between the GIS and other programs, was demonstrated using screen snapshots from a typical integration. Then, the integrated data were prepared for export into a statistical analysis package from the GIS and the results imported back to the GIS for data visualization or display.

Several statistical analysis methods to develop performance models were provided, along with reference examples for ANOVA, comparison of means, and regression models. Currently, there is an on-going research study, NCHRP Project 9-22, *Beta Testing and Validation of HMA PRS*, that will develop software capable of developing pavement performance models. It is expected that the report and software be completed by the end of 2008.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1 Introduction

1.1 Background

In-situ pavement performance can be considered a response variable to many project input variables, such as design (i.e., material properties, engineering criteria, etc.), construction (i.e., selected materials, design targets, base condition, etc.), and both environmental and traffic loading effects. If Wisconsin DOT and Industry are to fully understand and realize the true components of in-situ pavement performance, and specify the necessary inputs through design and construction specifications to achieve that performance, quantitative relationships must be developed between the input variables and response variables through a scientific, fully-integrated pavement performance system. The existing processes used by Wisconsin DOT and other highway agencies to determine these relationships have been largely based on a collection of experience and knowledge acquired through years of pavement performance monitoring and continuous specification development. With many experienced personnel retiring this decade, a Pavement Performance Analysis System (PPAS) will provide a lasting tool to understand the complete pavement system.

1.2 Problem Statement

A comprehensive and fully-integrated data acquisition, modeling, and analysis system is necessary to quantify the relationship between design and construction inputs, and the resulting in-situ pavement performance output. Portions of the system are already in place, but research is necessary to identify new components and fully integrate the system. For example, existing construction inputs include Job Mix Formula (JMF) data, Quality Management Program (QMP) data, and construction inspection records. Examples of existing in-situ performance output components include the Pavement Distress Index (PDI), International Roughness Index (IRI), Life-Cycle Cost Analysis (LCCA), and Long-Term Pavement Performance (LTPP) measurement systems. Both of the above construction and performance measurement systems are in-place, but are not fully integrated into a PPAS.

1.3 Project Objective

The objective of this project was to develop a database template, using the existing Wisconsin DOT pavement management system, from which to perform pavement performance analysis using design, construction, and performance data for Hot-Mix Asphaltic Pavements (HMA). A second objective of this study was to investigate appropriate numerical or statistical methods that have the potential of quantifying and establishing relationships between design, construction, and performance data.

CHAPTER 2 Literature Review

The purpose of this literature review is to identify pertinent performance data types collected for PMS, data integration, and analytical methods applied, as well as corresponding outputs. Pavement performance is a measure of highway deterioration and indicates the variation in the level of service provided to the pavement user over time. It is at the center of every pavement management system (PMS) and forms the basis for determining needed maintenance and rehabilitation strategies, as well as the overall cost-effectiveness for a given highway section or network. In addition, it provides a basis for verifying design methods, in that if a pavement section performs well, more than likely the next design will follow the procedures used in the previous design to achieve the desired performance.

2.1 Pavement Management System Data Needs

AASHTO (1993) defines a pavement management system (PMS) as a set of tools or methods that assist decision-makers in finding optimum strategies for providing, evaluating, and maintaining pavements in a serviceable condition over a period of time. PMS data needs and uses have been discussed by Haas et al (1994) and can be summarized as in Table 2.1. With the exception of the policy and cost related data, all categories provide background information for pavement performance modeling and analysis, and are discussed in the following sections.

2.1.1 Section Reference and Description

Historically, different divisions within an agency often have data collection and use needs that are not totally compatible with the needs of other divisions. Therefore, it is not uncommon to see multiple methods of referencing the location of pavement sections within a highway network. The construction division, for example, may use a construction project numbering scheme, while the operations division may use route milepost method for scheduling maintenance operations. These functions need to be coordinated to create a permanent referencing system for a functional PMS. NCHRP Synthesis 335 (2004) reported a survey in which 96% of highway agencies indicated using the milepost/logpoint method for referencing, while 15% use landmarks for referencing. The milepost referencing method requires each roadway to be given a unique name and/or number, and a distance along the route from a given origin to define points along the route. The increasing use of Geographic Information System (GIS) and Global Positioning System (GPS) technology however, is propelling the use of coordinate-based referencing systems to identify points along routes. NCHRP Synthesis 335 identified 35% of surveyed agencies using longitude and latitude, and 13% using state plane coordinate or related systems.

Table 2.1 Major Classes and Component Types of Pavement Data (Haas et al., 1994)

Data Category (1)	Components (2)	Primary Uses (3) Rehabilitation	Primary Uses (3) Maintenance
Section reference and description	-	X	X
Performance Related Data	Roughness	X	
	Surface distress	X	X
	Deflection	X	
	Friction	X	X
	Layer material properties	X	
Historic Related	Maintenance history	X	X
	Construction history	X	X
	Traffic	X	X
	Accidents		
Policy Related	Budget	X	X
	Available alternatives	X	X
Geometry Related	Section dimensions	X	
	Curvature	X	
	Cross slope	X	
	Grade	X	
	Shoulder / curb	X	X
Environment Related	Drainage	X	X
	Climate (temperature, rainfall, freezing)	X	
Cost Related	Construction costs	X	
	Maintenance costs	X	X
	Rehabilitation costs	X	
	User costs	X	

2.1.2 Performance Data

Performance data relates to the current and historical condition of the pavement. Four key indicators are commonly used to characterize the condition of the pavement and include roughness, surface distress (e.g., rutting, cracking, and faulting), deflection, and surface friction (as related to safety). The four indicators are the variables that can be measured to determine whether the pavement is functioning satisfactorily. These indicators would originally be predicted at the design stage and then periodically evaluated while the pavement is in service. A survey of 51 state agencies revealed that approximately 55% of agencies used both manual and automated methods in performance data collection while 27% use automated methods only for the same purpose (NCHRP 2004). The automated data involves multipurpose data collection vans that employ technologies such as Global Positioning System (GPS), laser sensors, and video cameras to capture inventory and centerline information. Day and Lewis (2002) have summarized best practices in automated highway collection equipment.

2.1.3 Historical Data

Historical data primarily include construction, maintenance, and traffic data. Construction data includes information on the as-built quality of the materials, such as density and permeability characteristics of asphalt concrete. Significant variability in construction quality can result in poor performance, compared to pavements with uniform quality. Pavement maintenance data involve maintenance activities that impact performance (e.g., crack sealing and patching). A high level of maintenance can result in an extended life of the pavement beyond the expected service life.

Traffic data is critical for performance prediction and for priority assignment in the selection of rehabilitation projects. Performance modeling requires an estimate of the heavy vehicle traffic that causes the majority of the pavement deterioration. The estimate is adjusted to reflect traffic growth rate for the performance period under consideration. Twenty-one of 37 agencies surveyed indicated using both automated and manual methods in traffic volume data collection (NCHRP 2004).

2.1.4 Environmental data

Pavement performance can be seriously affected by environmental and drainage conditions. The common measures used as indices of environmental conditions include freeze-thaw cycles, freezing index, seasonal rainfall, Thornthwaite Index, drainage quality, and regional factors developed by an agency (Haas et al. 1994).

2.2 Data Integration

Analytical models involving structural or functional pavement performance requires information about the indicator of performance (e.g., roughness and distress type), as well as the potential variables that affect the performance indicator (e.g., traffic loads, climate, material characteristics, maintenance and rehabilitation history, and construction history). This required information is generally associated with pavement management systems (PMS) and are often kept in separate databases managed by different divisions or offices within an agency. For example, the planning division keeps traffic records, while construction and maintenance records are maintained respectively by the construction and operations divisions. The problem is further compounded by the use of different referencing systems and data formats that may be used by some divisions within an agency. To facilitate the modeling process and other PMS activities, it would be desirable to have this information centralized so that all divisions can have ready access to the needed data and also minimize duplication. One of two main methods can be used for data integration namely, data fusion and interoperable or federated databases (FHWA 2001). Data fusion combines data from multiple sources into a single database, whereas federated databases employ multiple queries to relate data residing in different databases. Although data integration is considered very important, NCHRP 335 (2004) reported on the basis of a survey that the number of agencies that have actually completed or are close to completing a full integration of the systems is very limited.

Geographical Information Systems (GIS) have been identified as an effective tool for data integration. According to AASHTO (2001), the Illinois DOT has used GIS to integrate information from disparate databases to provide information for PMS activities. Roadway, structure, and rail crossing inventory systems are tied to a link/node base, and allow the use of multiple referencing schemes of route and milepost designations.

2.3 Performance Data Analysis and Presentation

Two main types of performance data analyses are common in the literature. The first type involves description of the present status of the network, while the second type involves prediction of the future condition of the network.

2.3.1 Performance Analysis Formats for Describing Present Network Condition Status

A wide range of formats have been used in expressing the condition status of pavement networks. These include:

a) Color-coded maps indicating in a categorical manner, the condition of all pavements in the network. This is facilitated using GIS as a tool. Petzold and Freund (1990) produced one of the earliest GIS applications to display and analyze the Highway Performance Monitoring System. The Virginia Department of Transportation has used GIS to display the general pavement conditions for its road network by county, as well as illustrating sections that are above, near, or below established condition threshold values.

b) Graphical representations of pavement condition involve the use of histograms and pie charts to show the percentages of pavement in some particular condition (e.g., good, fair, or poor). These can be broken down by highway class, political jurisdiction, etc.

c) Tabular summaries are very useful when information is sought on a specific pavement section. Tables can be used to display, for example, the sequential listing of all pavement sections based on the performance indicator values, or listings sorted by the common highway name.

2.3.2 Performance Modeling for Describing Future Condition Status of Network

Knowledge of the future condition of the pavement network allows agencies to determine maintenance and rehabilitation needs, prioritization schemes, and anticipated costs to bring the network condition to a predetermined acceptable level. The future condition is determined through prediction models. The requirements for developing reliable performance prediction models have been outlined by Darter (1980) and include:

- Having an adequate database for the pavements in service;
- Consideration of all variables that affect performance;

- Selection of the appropriate functional form of the model to represent the prevailing pavement condition; and
- Measures to assess the model precision and accuracy.

Several different types of prediction models have been discussed in the literature but they can be grouped into the following categories:

a) Mechanistic Models: According to Lytton (1987), mechanistic models predict changes in some primary mechanistic response of the pavement such as strain, deflection, or stress caused by factors such as load, temperature, and pavement support.
b) Mechanistic-Empirical Models: For mechanistic-empirical models, a response parameter such as strain, stress, or deflection is related to measured structural or functional deterioration, such as distress or roughness.
c) Empirical Models: The models relate the change in condition to the age of the pavement, loadings applied, or some combination of both. Empirical models are commonly developed through the use of regression analysis. However, a newer generation of methods including artificial neural networks, genetic algorithms, and fuzzy sets has also been used for empirical model development (AASHTO 2001).
d) Probabilistic Models: This model form describes the probability that a pavement in a known condition state at a known time will change to some other condition state in the next time period. Three types of probabilistic models have been used in the literature to develop condition models (Lytton 1987) and include: Markov Models, Semi-Markov Models, and Survivor Curves.

Table 2.2 summarizes sample performance models with corresponding factors found significant in explaining the variations in a particular performance indicator. Although the mechanistic and mechanistic-empirical models use the strain or stress properties of the asphalt concrete, they do not relate the ultimate performance measure (e.g., roughness) to the mix design characteristics that produced the strain or stress. Such information will be important in defining mix specifications for yielding a certain performance output.

Table 2.2 Sample Performance Outputs and Input Variables

Model Type (1)	Performance Output (2)	Performance Inputs (3)	Pavement Type (4)	Model Source (5)
Probabilistic (survivor curves)	NYSDOT Condition Rating ranging from 1(poor) to 10 (excellent)	Age	Rigid	DeLisle et al., 2003
Empirical based on Artificial Neural Network	FDOT pavement condition Rating (PCR)	Age, maintenance cycle, crack index, rut index, ride index	Flexible, Rigid	Yang et al., 2003
Empirical based on regression	Present serviceability rating (PSR)	Transverse joint faulting, number of transverse cracks per mile, number of deteriorated joints per mile, number of full-depth repair per mile.	Rigid (jointed plain concrete)	Huang, 2004
Empirical based on regression	Rutting Rate	Surface deflection, vertical compressive stress, ESAL	Flexible	Huang, 2004
Mechanistic	Allowable load applications	Tensile strain at bottom of asphalt layer, dynamic modulus of asphalt mixture	Flexible	Asphalt Institute, 1981
Mechanistic	Tensile strain at bottom of asphalt layer	Asphalt layer thickness, deflection difference at 305 and 600 mm of the radial distance from load plate center.	Flexible	Park and Kim, 2003
Mechanistic-empirical	Roughness	Age, roadway surface type, rehabilitation state, Strain energy at bottom of asphalt layer, cumulative equivalent single axle load	Flexible	Queiroz, 1983
Mechanistic-empirical	Percent pavement area cracked	Horizontal tensile strain at bottom of asphalt layer, cumulative equivalent single axle load	Flexible	Queiroz, 1983

2.4 Summary and Conclusions

A review was conducted regarding pertinent PMS data types generally collected for modeling pavement performance. In addition, data organization and integration issues were examined, as well as analytical methods for modeling and presenting performance related data. Based on the review, the following observations are made:

- Data types collected for performance evaluation and modeling vary from agency to agency depending on needs but the most common ones include inventory, condition, traffic volume, and maintenance and rehabilitation.
- Adoption of common referencing systems between various data collection systems can facilitate data integration for pavement performance modeling. There is evidence in the literature to suggest that a major barrier for achieving full data integration by agencies is lack of common referencing systems compounded by the use of different data formats.
- A Geographic Information System is an effective tool for data integration among various divisions within an organization.
- Major types of performance models include mechanistic, mechanistic-empirical, empirical, and probabilistic. Although the mechanistic and mechanistic-empirical models use the strain or stress properties of the asphalt concrete, they do not relate the ultimate performance measure (e.g., roughness) to the mix design characteristics that produced the strain or stress. Such information will be important in defining mix specifications for yielding a certain performance output.

CHAPTER 3 Review of Wisconsin Flexible Pavement Related Databases

This chapter reviews the major flexible pavement databases for the purpose of determining available data types for performance modeling, structures, contents, and additional data needs where necessary. The major databases for Wisconsin flexible pavements include design, construction, Meta-manager, and performance. Each of the databases is briefly described in the following sections.

3.1 Design Database

The design database found within the New Construction Report is a set of Microsoft Access files organized by year from 2000 to 2004. Each file has two key datasheets, namely, *ACOffice* and *ACField*. The A*Coffice* datasheet has a total of 544 records, which show pavement location (rural or urban, district, county, termini by descriptive start and end points), construction style (reconstruction, resurfacing, rehabilitation), contract identification numbers (*contract1*, *contract2*), project length, pavement surface thickness (*Pvtthick*), milling depth, base type (DGBC, CABC, OGBC2), pavement surface paved over (*Pvdovr*), flexible pavement type, surface year (*pvmntyr*), *mix type denoted by HvMvLv*, case type (Standard, Superpave, SMA, AC Warranty), and design ESAL magnitude. The *ACField* datasheet has a total of 5,620 records that show fields representing site identification number (site), sequence number (*Sqno*), beginning reference point (RP), contract identification number *(contract2*), highway name by direction for all years except 2002, survey length (Survlen), lane, direction, Asphalt or PCC, set value, measured IRI, and rut depth (Rut) immediately after construction.

3.2 Construction Database

The construction database consists of Microsoft Excel spreadsheets organized by year from 1997 to 2004. There are two key files, a design/test log file (*GeneralLogs 1997-2004.xls*) and a mix design data file (*data 1997-2004.xls*). The design/test log file contains 2,380 records that show fields representing the highway type (STH, local, CTH, etc.), highway number or letter, surface year, aggregate sources, project location (by descriptive start and end points), county, district, project identification number, contractor, PREfix, test number, mix type, and ESAL category (E-0.3, E-1, etc) for 2001-2004. The mix design data file shows mix design data for 2,402 records. The mix design data consist of %AC, %VMA, aggregate size distribution in mix (3/4", ½", 3/8", #4, #8, #16, #30, #50, #100, #200), %RAP, Gse, Gsb, Gmm, Gmb, dryback correction, flow, stability, TSR, blows, anti-strip agent, and asphalt cement characteristics (type, source, specific gravity). In addition, the latter file lists contractor name, PRE, test number, type that shows either ESAL category together with nominal maximum aggregate size (e.g., E-3 12.5R, for 12.5-mm NMAS mixes having RAP), or mix type (e.g., MV-3, SPPV, etc.) for various records.

3.3 Meta-Manager Database

Meta-Manager is a comprehensive, integrated database system for conducting needs and performance analyses for pavements and bridges. It is updated and distributed quarterly (Javenkoski et al., 2005). It is comprised of independent databases organized by region for all five regions in Wisconsin. Each region consists of one Excel spreadsheet workbook with multiple datasheets, as well as, ArcGIS shape files and ArcInfo GIS coverage files that can be used for geographic analysis. The workbook datasheets include information on base, roadway, unimproved pavement condition, improved pavement condition, safety, pavement treatment scoping, mobility, unimproved bridge condition, and improved bridge condition.

The mobility and roadway datasheets contain projected traffic volume data relevant to pavement performance modeling. Both datasheets identify pavement segments using sequence numbers, traffic segment identification numbers, and from-and-to reference points. Other relevant fields include highway number by direction, projected 2-way AADT, and percent trucks for 1, 5, 10, 15, and 20-year periods from a base year.

3.4 Performance Indicator Database

The performance database commonly referred to as PIF, is a relational database model designed to store pavement inventory information, capture distress characteristics, and summarize continuous ride/rutting data. The data is maintained on the host and can be served out on CD-ROMs or DVDs in many formats, including Microsoft Access. It contains pavement inventory and condition data and has various customized forms to facilitate data entry. In addition, it has several datasheets for tabular summaries of data. The key datasheets include the descriptive (DESC), pavement distress index (PDI) history file, and International Roughness Index (IRI) data.

The descriptive file identifies pavement segments by sequence numbers, county name, county number, district, from-to reference points, from feature, highway number, highway direction, functional class number, national highway system designation, surface year and original construction year. In addition, the datasheet has fields for the segment length, cumulative mileage, and roadbed soil type.

The IRI datasheet contains 153,461 records representing segments tested between 1980 and 2005. Approximately 77% of the records pertain to flexible pavements. The datasheet lists fields representing the sequence number, inverse year, day-month-year segment was tested, the surface year, surface type, air temperature, average values for IRI, PSI, and Rut. In addition, it lists the speed at which tests were conducted.

The PDI history datasheet has 65,535 records for flexible pavement segments tested between 1985 and 2005. It lists the segment sequence number, inverse year, test day-month-year, surface year, distress type severity and extent for quantifying PDI.

3.5 General Observations about Flexible Pavement Databases for Performance Modeling

Integration of information from disparate databases for performance modeling requires that semantic discrepancies within and between databases be identified and alleviated. In addition, key fields must be identified within and across the databases to enable simple or complex queries to be performed in order to relate data residing in the different databases. It is also essential to identify the performance indicator(s), such as rutting, and check if all potential influential variables are adequately represented, such as air voids and asphalt content.

3.5.1 Semantic Discrepancies in Databases

The semantic discrepancies between the databases are summarized in Table 3.1 and include the use of different field names or labels that represent the same information, inconsistent formats for data entry, and redundant fields for some databases.

Roadway segment identification varies from database to database. While the performance and Meta-Manager databases use sequence numbers and reference points to identify segments, the construction and design databases use descriptive words to define the start and end points of whole projects, which in most cases may include multiple segments or sequence numbers. The design and construction databases lack consistency in project identification based on format for entering identification numbers. *Proj#* and *contract1* are identical in format but not *contract2*. One format is necessary to relate the two databases. This would require a redundant identification field (either *contract1* or *contract2*) in the design database to be removed to save valuable computer space and time for data entry.

Fields designated to help identify highways have varying interpretations. In the construction database, two fields are required to completely identify a highway without an associated direction. The design database identifies highways in a variety of ways including, specifying the highway with or without direction (e.g., 090E, 90). In both cases, an exclusive field is also provided to indicate direction (e.g., N or E). While Meta-Manager uses a single field to completely identify a highway by number and direction, the performance database uses two fields (one for the highway number and the other exclusively for direction). From efficiency and time savings considerations, it may be appropriate to identify a highway using the Meta-Manager format, which uses one field for both highway number and direction.

There is no clear distinction between ESAL and mix type as they pertain to the construction and design databases. ESAL is considered a specific value (e.g., 6 million) for design but sometimes considered as a category (e.g., E-10) representing a type of mix. Mix types have also been specified alongside NMAS (e.g., SPPV-19.0). From performance modeling point of view, the NMAS value may be required. Hence, it is appropriate to have separate fields to denote mix type and NMAS values to facilitate data retrieval for performance analysis.

Formats for entering time events, such as pavement surface year, tend to vary between the construction database and the other databases. While the construction database uses two digits to indicate a particular surface year, the performance and design database use four digits (e.g., 07 versus 2007).

Table 3.1 Database Discrepancies

Database (1)	Field label (2)	Intended meaning (3)	Comments (4)
Design	Contract1 Contract2	Project identification number	Contract1 is an 8-digit number of the form 1234-56-78; Contract2 is same as Contract1 without the dashes (12345678).
	ESALS	Design Equivalent 18-kip single axle load value	Indicated as a specific number
	Gen_loc	Project termini	Project termini specifies beginning and ending points (e.g., Siren-CTH D).
	HvMvLv	Mix type	Field label seems to suggest that only HV, MV, LV mixes are applicable but some records for the field has SMA, E-0.3 and higher mix types, as well as "warranty" labels.
	RP	Starting point for Iri and rut measurements (e.g., 323T 1.2)	-
	HWY	Highway by direction (e.g., 090E) or just the highway number (e.g., 90 without direction).	There is an additional field for direction (Dir), even though the HWY field has direction associated with the highway (e.g., as in 090E) for all years except 2002. Note also that the *AC Offic*e HWY designation field uses 2-digits (e.g., 43) compared to 3-digits (043) as in the *AC Field* HWY designation field
	Pvmntyr	Pavement surface year	Year is represented as 4-digits (e.g., 2004)
Construction	ESAL	ESAL level category	Designated as E0.3, E1 or higher
	HWY	Highway type (e.g., STH, CTH, USH)	A separate field (#) is used to denote the number or letter label associated with the highway (e.g., 41 or K).
	Proj #	Project identification number	Proj # is an 8-digit number of the form 1234-56-78
	START	Project beginning point (e.g., USH 61)	-
	END	Project end point (e.g., East Co. Line)	-
	Type	Mix type	This field has records with HV, MV, and LV with various designations e.g., MV-2, MV-2R. In addition, superpave mixes are labeled with their corresponding NMAS in the same field (e.g., SPPV-19.0).
	YR	Pavement surface year	Year is represented as the last 2-digits of the year (e.g., 04 for 2004)

Table 3.1 Database Discrepancies (cont.)

Database (1)	Field label (2)	Intended meaning (3)	Comments (4)
Meta-Manager	HWY&DIR	Highway and direction	-
	ISEQNO	PIF segment ID number	-
	PDP FRM	From RP	-
	PDP TO	To RP	-
Performance	From RP From Plus	Identification of segment start location based on a beginning reference point (BRP) plus some distance from the BRP	-
	To RP To Plus	Identification of segment end location based on an ending reference point (ERP) plus some distance from the ERP	-
	From Feature	Feature denoting beginning of segment over which measurement is to be taken (e.g., Catlin Ave Intersection)	-
	HWY No	Highway number without direction	A separate field exists for highway direction (Dir)
	Sequence No.	Identification number to locate field measurement sample segments	A unique number or field for relating all PIF components
	Surf Year	Pavement surface year	Year is represented as 4-digits (e.g., 2004); in some cases represented by 1 or 2 digits (e.g., 1 for 2001 and 91 for 1991)

3.5.2 Database File Relations

A relational database is a collection of files that are tied together by common fields (Harris 1999). Table 3.2 shows the four main databases with component files or datasheets and corresponding key fields that are used to relate the component files. Table 3.3 on the other hand, shows the databases and potential primary key fields that can be used to relate the databases. The *sequence number* key field is common for the performance and meta-manager databases in relating their component files, while the construction and design databases have significantly different key fields as indicated in Table 3.2. The project identification number in the construction database may be a better primary key field than the *Test#* field for relating component files in the construction database. This is only possible if the identification number field can be included in the *data 1997-2004.xls* file. The key to an integrated database system is the inclusion from any data file of data fields that can be used to connect other files. Table 3.3

however, suggests that the construction database in its present form can neither be directly related to the Meta-Manager nor the performance databases.

Table 3.2 Key Fields for Relating Database Component Files

Database (1)	Component files/datasheets relevant to performance (2)	Key field for relating component files/datasheets (3)
Construction	GeneralLogs 1997-2004.xls data 1997-2004.xls	Test #
Design	AC Office, AC Field	Contract2
Performance	DESC, IRI , PDI_F	Sequence number
Meta-manager	Base, Roadway, Pave_Uimp, Pave_imp, Safety, Pave_Scope, Mobility, Bridge_Uimp, Bridge_imp	Sequence number

Table 3.3 Potential Primary Key Fields for Relating Databases

Database (1)	Construction (2)	Design (3)	Performance (4)	Meta-manager (5)
Construction	-	Contract1	-	-
Design	Contract1	-	Sequence No	-
Performance	-	Sequence No	-	Sequence No, From RP-To RP
Meta-manager	-	Sequence No	Sequence No, From RP-To RP	-

3.5.3 Database Variable Deficiencies for Performance Modeling

Darter (1980) outlined that a reliable prediction model ought to have an adequate database based on in-service segments and consider all factors that affect performance. Thus, from the standpoint of pavement performance, it is essential to first identify the relevant performance indicators and potential influential variables, as well as the databases housing the relevant group of variables. Table 3.4 depicts the main performance indicators, potential influential variables by category and their corresponding WisDOT databases. Table 3.4 indicates that pavement structural layer components such as base and subbase thicknesses are not represented in any of the databases; neither is the structural number, which represents the overall indicator of the flexible pavement strength. Although the PIF database provides some information on subgrade type in terms of the pedalogical soil names, it does not directly provide the overall soil strength, for example, used in the design of the pavement. Structural layer property information may be obtained from the pavement structural design section of the Foundation and Pavement Unit and incorporated in the database.

Table 3.4 Performance Indicators and Influential Factors

Performance Indicator (1)	Influential Variable Category (2)	Potential Influential Variables (3)	Database housing Influential Variable (4)
IRI RUT PDI Distress	Structural layer Properties	Surface thickness	Design
		Base thickness	
		Base type	Design
		Subbase type	
		Subbase thickness	
		Subgrade type	PIF
		Structural number	
	Environmental	Surface year (for age determination)	Design, PIF
		Regional location	Design, PIF
		Mean monthly or annual temperature	
		Mean monthly or annual rainfall	
		Freeze-thaw cycles	
	Construction	As-built density	
		AC content	Construction
		AC type and source	Construction
		VMA	Construction
		VFB	
		Air voids	
		Aggregate size distribution	Construction
		NMAS	Construction
		TSR	Construction
	Traffic	Heavy vehicles	Meta-Manager
		AADT	Meta-Manager
		ESAL	
		Functional Class	PIF
	Maintenance	Treatment type	
		Surface condition prior to treatment	
		Treatment year	
		Treatment cost	

Environmental conditions such as rainfall, freeze-thaw cycles, and aging have an impact on pavement performance. With the exception of age, none of the other variables is currently addressed in the database. There are approximately 18 weather stations in the state, each containing about 10 years of climatic data. It is recommended that the appropriate weather data be assigned to a specific pavement segment. An alternative approach to this may be using the pavement regional location based on WisDOT's five-region demarcation of the state or location based on North, Central, and South Zones using State Plane Coordinate boundaries established by the Wisconsin State Cartographer's Office (1995). In the absence of detailed temperature or climatic data, the pavement regional location can be modeled as dummy variable to allow the effects of climate to be accounted for in performance analysis. An application and analysis using the three zones can be found in a report by Owusu and Schmitt (2003).

Traffic volume is represented in Meta-Manager as projected 2-way AADT, and percent trucks for 1, 5, 10, 15, and 20-year periods from a base year. The traffic reported in the database are projections and do not reflect the actual conditions corresponding to specific values of performance. With WisDOT's move towards the AASHTO Mechanistic-Empirical Pavement Design Guide, the use of ESAL may not be needed in performance modeling. However, the functional class may be used as a surrogate for traffic in the absence of detailed traffic counts. Poor construction practices can result in poor performing pavements. Schmitt et al. (2007) observed that PDI on Wisconsin rural arterials decreased with increased as-built density. Thus, it is essential in the modeling process to understand the relationship between the density and mix parameters and how they in turn relate to the indicator of performance. Table 3.4 indicates that some relevant construction data such as as-built density are not represented in the construction database. Elements of maintenance and rehabilitation practices generally do not exist in any of the databases, with the exception of patching and crack sealing found in the PIF database.

3.5.4 Location Referencing Indicators

In order to physically relate specific information from a database to the road network or segments, there must be some form of a location referencing indicator (LRI) to help identify the segment(s) of interest. Table 3.5 shows the LRI for each of the databases. As much as these LRI currently meet their intended purposes, they are limited in their ability to link with coordinate-based databases. Recent technological advancements in computing and highway technology suggest a shift towards global positioning system approach for determining location.

Table 3.5 Location Referencing Indicators

Database (1)	Location Referencing Indicators (2)
Performance (PIF)—Segment Rut, IRI,	From_RP, From_Plus To_RP, To_Plus; Sequence number
Performance—Segment PDI	Sequence number
Design—Structural layer properties	Contract2
Construction—Mix properties	Test #
Construction—Aggregate sources	Project #
Meta-Manager—Traffic (AADT)	Sequence number, From_RP To RP

CHAPTER 4 Database Integration for Performance Modeling

The overall purpose of data integration in this project was to facilitate pavement performance modeling. The integration process involves an understanding of several elements including data types and formats to be collected and managed, LRI, database structures and relationships, software and hardware requirements, as well as institutional issues involved in the implementation and use of the system.

The Federal Highway Administration (FHWA) described two alternatives to data integration including data fusion and interoperable databases (2001). The former combines data from multiple sources into a single database, while the latter relate data from different databases through a series of queries. A review of the WisDOT databases outlined in the previous chapter revealed several integration challenges including inadequate representation of influential variables for performance, limited capabilities in linking databases, semantic discrepancies between the databases in terms of the use of different field names or labels that represent the same information, inconsistent formats for data entry, and redundant fields for some databases. Thus, regardless of the selected alternative, these semantic differences will have to be initially addressed. The semantic discrepancies can be addressed by producing a single data dictionary for all system designers and users. Database linkages can be facilitated by identifying logical associations between databases that would result in meaningful correlations for performance modeling. Relevant missing performance influential variables identified in Table 3.4 would need to be gathered to yield more robust performance models.

4.1 Location Referencing Based on Reference Point System

A single database system modeled after the Meta-Manager system is being proposed as an alternative. It will consist of six independent tables representing the variables in columns 1 and 2 of Table 3.4. For these tables to be appropriately linked to each other (as in a relational database), a common LRI to aid in segment identification is needed. The basic LRI proposed is based on the WisDOT reference point (RP) system. This particularly involves the conversion of construction projects' termini, as well as the start and end locations of all test lots/sections in terms of the WisDOT reference point system. Once this is done, construction data measures such as as-built density, JMF, and aggregate sources associated with particular lots/section for given reference point interval can be aligned with corresponding field performance data. Table 3.5 indicates that this approach is feasible since the performance data for segments already uses the RP system for location referencing.

Figure 4.1 provides a schematic of overlaying databases for the purpose of assigning data attributes to pavement segments based on the Reference Point System. In this figure, performance data are identified by sequence numbers, while design, traffic, and environmental data are continuous across the given constructed segment. Construction data for the contractor's Job Mix Formula (JMF) and Ride data overlay the entire project. Where there is a JMF change during construction, the appropriate change can be made using the construction stationing and conversion process provided in the following sections of this report. Construction mix properties and density require the actual as-built test values, where mix properties are in the individual test result, and density is the average of 5 or 7 tests per 750-ton lot. Because of the relatively large standard deviation associated with determining density (Schmitt et al. 2006), at this time it is recommended that the average for the lot be used in assigning an as-built density to the appropriate sequence number. Further research is recommended to determine the appropriate assignment of as-built construction data to a given location reference.

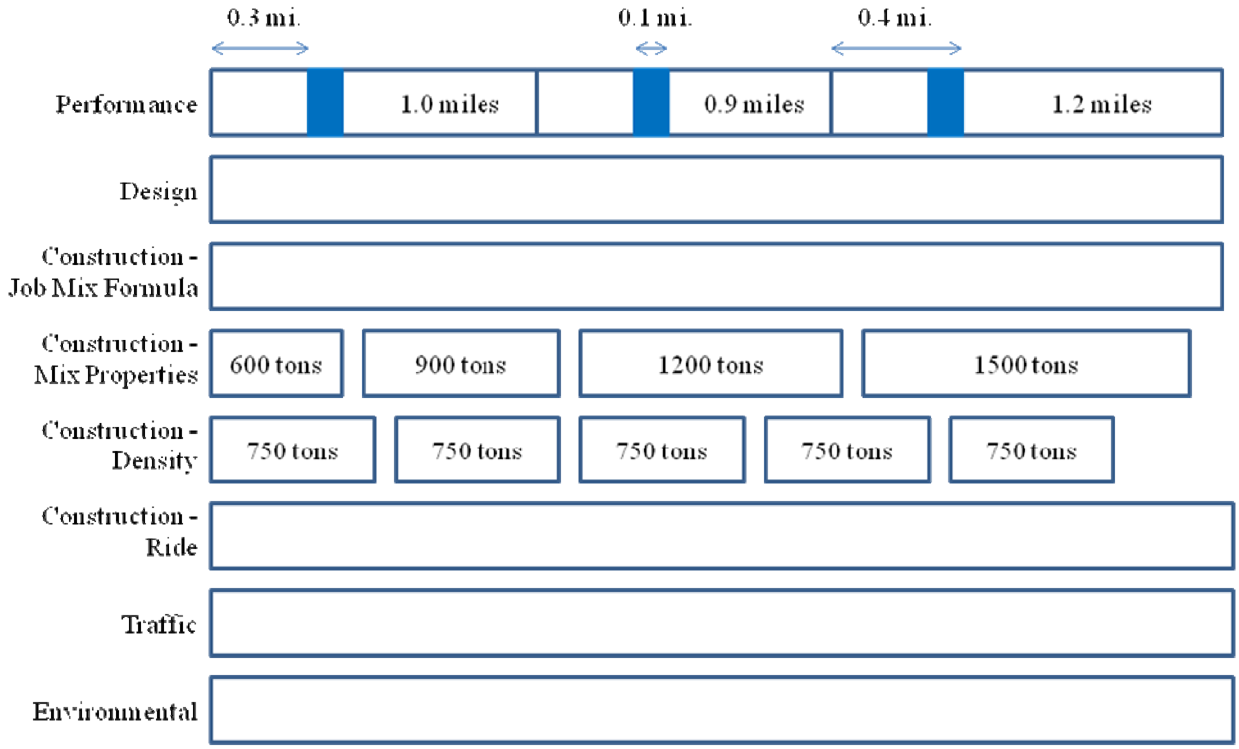

Figure 4.1 Overlay of Databases using Reference Point System

4.2 Conversion of Construction Stationing to Reference Point

At the present time, WisDOT does not have a defined procedure for relating construction stationing to the reference point system. In order to complete the alignment and integration of applicable HMA databases to model pavement performance over time, construction plan sets and field data were necessarily obtained. Plan sets for each pavement project, formatted as PDF files, were obtained via email or an ftp website (ftp://ftp.dot.state.wi.us/pub/) used by WisDOT, depending upon the method preferred by each regional office.

Using the existing reference point system, an overlay of the project stationing with the pre-existing reference point system was completed for several sample projects. In order to measure the PDI, an automated performance survey is taken continuously from an intersection or some other distinguishable feature, such as a bridge or county line. The recorded length begins 0.3 miles from a reference point for a length of 0.1 miles; therefore, as depicted by the shaded areas in Figure 4.2, the performance is recorded between 0.3 to 0.4 miles after a pre-determined point.

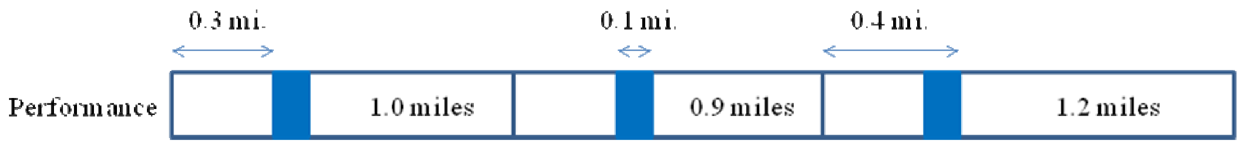

Figure 4.2 Ride Quality Measurement Methodology.

Construction plan sets were needed so that a determination could be made as to where the pre-assigned points were with respect to each project. Based on current WisDOT practice and assessment by the Pavement Management Unit, there is no way to equate an RP with project stationing without visually interpreting the plan and profile sheets for each project (Vils 2007). Not all plans have stationing; some plans use what is known as log mile, or the total cumulative mileage on a particular highway starting from where the highway begins in the state or at the state border this mileage continues until the highway ends in the state or again at the state line. When working with the plans to determine which RP may be affected, a visual inspection of the plan is necessary. The stationing and log mile are ignored, and both side road names or names of roads going over or under the mainline highway are matched with the appropriate RP.

A sample project overlay is displayed in Table 4.1. Here, the sequence number, depicted earlier as the name for the RP used in the PDI measurement, was matched up with the stationing of the project. Then the aforementioned ride-quality measurement methodology was applied to determine exactly where on the project the measurements were being taken. This data conversion ranged from 10 to 45 minutes per construction project, depending upon the length of the project and whether the stationing of the reference point was labeled or whether it needed to be obtained using a scale.

The *Length* column is the length between sequence numbers. The *Beg STA* column is the beginning project stationing from the plan sets. The *Start 0.3 – 0.4* column is the *Beg STA* column with 1,584 feet added (number of feet to reach 0.3-mile starting point). The *End 0.3 – 0.4* column is the *Beg STA* column with 2,112 feet added (number of feet to reach 0.4 miles). In this example, any data obtained between station 1040+02 and 1045+30 can be correlated with the PDI sequence number of 20820 to determine the performance of the HMA pavement over time.

Table 4.1 Sequence Number/Project Stationing Overlay (USH 18, Project I.D. 1660-04-73)

Sequence Number (1)	Length, mile (2)	Intersection (3)	Beg STA, ft. (4)	Start 0.3 - 0.4, ft. (5)	End 0.3 - 0.4, ft. (6)
20820	1.47	GRANT-IOWA CO LN	102418.0	104002.0	104530.0
20830	1.51	CTH XX INT R	110188.3	111772.3	112300.3
20840	1.02	ANDERSON LA INT	118170.1	119754.1	120282.1
20850	1.00	VICKERMAN RD INT	123523.8	125107.8	125635.8
20860	1.01	STH 80N & CTH G L	128831.6	130415.6	130943.6
20870	0.72	CTH J INT R	134135.7	135719.7	136247.7
20880	1.26	WHITSON RD INT R	137927.5	139511.5	140039.5
20890	1.01	STH 39E INT R	144617.9	146201.9	146729.9
20900	0.94	SUNNY SLOPE RD R	149949.9	151533.9	152061.9
20910	1.09	BETHLEHEM RD INT	154922.5	156506.5	157034.5
20920	1.00	CTH Q (BERG RD) R	160658.7	162242.7	162770.7
20930	0.90	TN OF DODGEVILLE	165938.7	167522.7	168050.7
20940	1.26	CTH Q (SURVEY RD)	171084.3	172668.3	173196.3
20950	0.78	USH 18W INT L	177737.1	179321.1	179849.1

4.3 Assignment of Construction Data to Sequence Numbers

The alignment of construction data measures with PDI sequence numbers is necessary to determine the effects each asphaltic concrete property has on the durability of road sections over time. Construction data measures included in the alignment were aggregate gradation, aggregate blend, bitumen data, mixture data, optimum asphalt content properties, and JMF properties. For each of the properties, field and design data was included if available.

4.3.1 Job Mix Formula

JMF data were obtained from the contractor for the project. An example of a portion of the JMF data overlay table is displayed in Table 4.2.

The data displayed under the heading column heading *JMF* are the optimum values for each of the sieve sizes as designed; these values were obtained directly from the documents provided by the contractor. However, the *Daily Average* values that are displayed were calculated. Each day, anywhere from two to five different samples was taken, and a moving average calculated for the four most recent test results. However, it should be noted that road sections were placed over two days for Sequence Numbers 61780 and 61790. In order to obtain the *Daily Average* for these cells, a weighted average was taken utilizing the daily average from each day. This weighted average was then input to Table 4.2. The *Daily Average* calculation methodology for the ½" sieve for Sequence Number 61780 is as follows: four samples were taken on 8/9/07, which had a daily average of 93.4, and three samples were taken on 8/6/07, which had a daily average of 91.9. These averages were then weighted, as shown by Equation 4.1, to provide the *Daily Average* value that was entered into Table 4.2.

Table 4.2 Sequence Number/JMF Data Overlay (USH 45, Project I.D. 9847-03-60)

Sequence Number (1)	Date Placed (2)	Daily Average, 1/2" (3)	Daily Average, 3/8" (4)	Daily Average AC Calc (5)	JMF, 1/2" (6)	JMF, 3/8" (7)	JMF AC Calc (8)	JMF Pbe (9)	JMF $P_{0.075}$/Pbe (10)	JMF Plant Mix Temp. (11)
61760	8/9/07	93.4	82.5	5.07	91.3	81.9	5.10	4.64	0.87	280-320
61770	8/9/07	93.4	82.5	5.07	91.3	81.9	5.10	4.64	0.87	280-320
61780	8/9/07 & 8/6/07	92.8	82.0	5.13	91.3	81.9	5.10	4.64	0.87	280-320
61790	8/6/07 & 8/2/07	91.6	81.3	5.20	90.7	80.9	5.20	4.70	0.81	280-320
61800	8/2/07	91.3	81.2	5.19	90.7	80.9	5.20	4.70	0.81	280-320
61810	8/2/07	91.3	81.2	5.19	90.7	80.9	5.20	4.70	0.81	280-320
61820	8/2/07	91.3	81.2	5.19	90.7	80.9	5.20	4.70	0.81	280-320
61830	8/2/07	91.3	81.2	5.19	90.7	80.9	5.20	4.70	0.81	280-320

$$DailyAverage_{1/2"} = \frac{4*93.4+3*91.9}{7} = 92.8 \quad \text{……… (Eq. 4.1)}$$

4.3.2 Density

Density data was also obtained from the documents provided by the contractor. A sample of this data is provided in Table 4.3.

Table 4.3 Sequence Number and Density Data Overlay (USH 45, Project I.D. 9847-03-60)

Sequence Number (1)	Date Placed (2)	Density Lower Lift (3)	Density Upper Lift (4)
61760	8/9/07	-	93.2
61770	8/9/07	-	92.9
61780	8/9/07 & 8/6/07	-	93.1
61790	8/6/07 & 8/2/07	-	93.3
61800	8/2/07	-	93.4
61810	8/2/07	-	92.6
61820	8/2/07	-	93.5
61830	8/2/07	-	93.7

Determining the density for the upper and lower lifts follows the same calculation procedure as the *Daily Average* described earlier, depending upon whether the lift was placed in one or two days. For this project, only a wedge and single surface layer were paved.

4.3.3 Mix Properties

Multiple mix properties were given in the documents provided by the contractor. A sample of the data provided is given in Table 4.4.

Table 4.4 Sequence Number and Mix Properties Overlay (USH 45, I.D. 9847-03-60)

Sequence Number (1)	Date Placed (2)	Daily Average Gmm (3)	Daily Average Gmb (4)	Daily Average Voids (5)	Daily Average VMA (6)	Opt. Gmm (7)	Opt. Gmb (8)	Opt. Voids (9)	Opt. VMA (10)
61760	8/9/07	2.509	2.420	3.5	14.7	2.516	2.416	4.0	14.8
61770	8/9/07	2.509	2.420	3.5	14.7	2.516	2.416	4.0	14.8
61780	8/9/07 & 8/6/07	2.509	2.416	3.7	14.9	2.516	2.416	4.0	14.8
61790	8/6/07 & 8/2/07	2.510	2.410	4.0	15.1	2.513	2.413	4.0	15.0
61800	8/2/07	2.510	2.410	4.0	15.1	2.513	2.413	4.0	15.0
61810	8/2/07	2.510	2.410	4.0	15.1	2.513	2.413	4.0	15.0
61820	8/2/07	2.510	2.410	4.0	15.1	2.513	2.413	4.0	15.0
61830	8/2/07	2.510	2.410	4.0	15.1	2.513	2.413	4.0	15.0

Data that is not displayed in Table 4.4 but is included in the spreadsheet were design aggregate blend, bitumen data, mixture data, and aggregate data. The *Daily Average* columns were again calculated as described earlier, and the *Optimum* values were recorded directly from the documents provided by the contractor.

The time consumed while completing this overlay was approximately 30 minutes. This approximation is based on the fact that the construction data measures were provided in electronic instead of paper form. An increase in the amount of time necessary to complete an overlay is likely if the construction data measures are presented in paper form. Since this project was constructed in 2007, and no other as-built construction data was entered at this time, a demonstration GIS-based integration was not possible.

4.4 Database Integration and Performance Modeling Using GIS

Star and Estes (1990) define GIS as "an information system that is designed to work with data referenced by spatial or geographic coordinates. In other words, a GIS is both a database system with specific capabilities for spatially-referenced data, as well as a set of operations for working with the data". A review of the literature suggests that GIS is an effective tool for integrating data from disparate databases that reside locally or at a remote location. Access to remote databases is made possible through the GIS database connection capabilities.

4.4.1 Basic Elements of GIS

The basic elements of GIS include data, a computer system, and GIS software. The data consist of geospatial (e.g., maps) and nonspatial data (e.g., data tables that relate to features on a map). The computer system includes the computer and operating system to run GIS. The software consists of the program and the user interface (e.g., menus, icons, command lines and scripts) for driving the hardware. In GIS, a common data source includes a reference or base map, which enables specific queries about related data tables to be visualized on a dynamic map. A sample base map is shown in Figure 4.3 for Wisconsin's network of roads. The map was derived from five geographic shape files (geo.shp) representing the 2005 STH network for the five regions of Wisconsin. According to Javenkoski et al. (2005), these shape files have been produced in accordance with the requirements for the North American Datum of 1983 High Accuracy Reference Network (NAD83 HARN). The shape files were imported into GIS software (Arcview^{TM}) from the Meta-Manager database. To make the map as a single theme rather than five separate themes, the maps were combined using the Geoprocessing wizard tool in Arcview^{TM}.

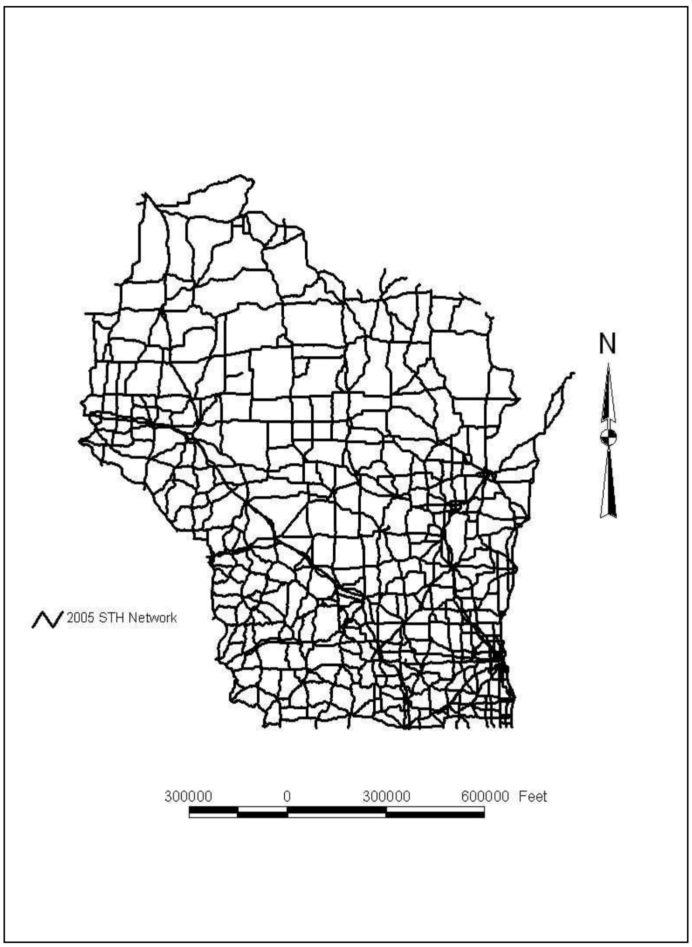

Figure 4.3 STH Network Base Map Generated in GIS

4.4.2 The Role of GIS in Performance Modeling

The ability of GIS to process, display, and integrate different data sources makes it an invaluable tool in the modeling process. For example, data reporting is a key means by which pavement management data is ultimately presented to decision-makers at all management levels within an agency and also to the public. FHWA (1991) has suggested that to meet the objective of data presentation to a given audience, the report has to be tailored for that particular audience. This is where GIS can be used as a powerful tool in modern day pavement management. For example, with the aid of GIS the highway network can be color coded according to segment condition score. This allows the intended audience to develop an immediate sense of the condition of an entire network. Once prioritization lists have been compiled, they too can be color coded by treatment type in a map-based format. Furthermore, the modeled relationship between network deterioration with varying funding levels can be graphically shown. This enables decision-makers to see what happens to the network for various funding level scenarios.

The process of modeling may take place in a GIS or require the linking of a GIS to other computer programs. There are GIS packages, such as ArcGIS or IDRISI, that have analytical functions for modeling. However, it is advised that a GIS package cannot accommodate statistical analysis as well as a commercial statistical analysis packages, or perform dynamic simulation efficiently. In those cases, it may be necessary to link the GIS to a statistical analysis package or a simulation program.

Various researchers (Corwin et al. 1997; Brimicombe 2003) have described scenarios for linking a GIS to other computer programs. These scenarios fall into three main categories including loose coupling, tight coupling, and an embedded system. The loose coupling involves the transfer of data files between the GIS and other programs. This scenario requires data files to be manipulated to be exported or imported unless the interface has already been established between the GIS and the target program. Thus, performance data can be exported into a statistical analysis package from the GIS and the results imported back to the GIS for data visualization or display. The tight coupling gives the GIS and other programs a common user interface. An embedded system bundles the GIS and other programs with shared memory and a common interface. In ArcGIS for example, the Geostatistical Analyst extension provides geostatistical functions embedded into a GIS environment.

In this project, a loose coupling approach is proposed; the reason being that a GIS package cannot accommodate statistical analysis as well as a statistical analysis package. The loose coupling approach is depicted in the framework shown in Figure 4.4. The performance-related inputs from the different databases are organized and imported into a GIS environment for basic data display and statistical analysis. The data is further processed for transfer into advanced statistical software for statistical modeling. Model results are fed back into the GIS and further displayed based on a simple or compound contextual query.

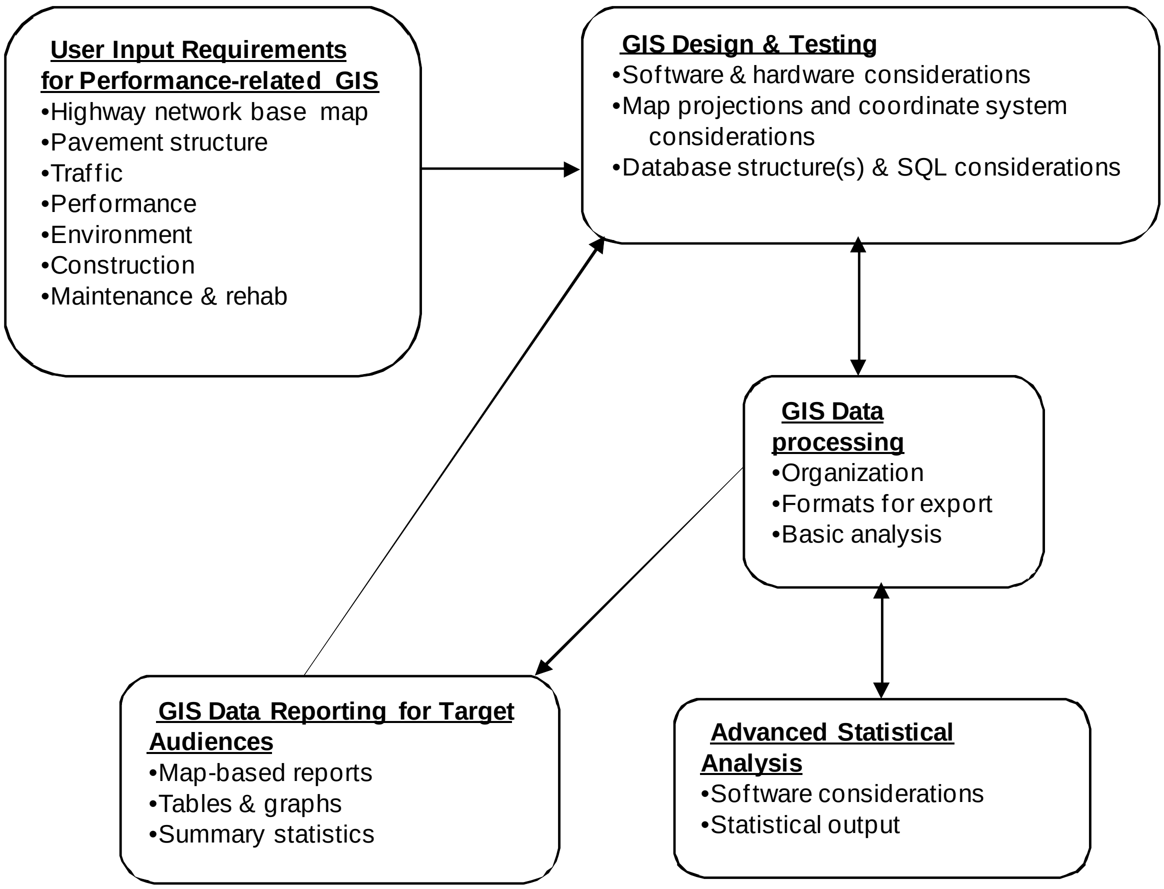

Figure 4.4 Framework for Linking GIS with Statistical Package for Performance Modeling

4.4.3 Database Integration Example Using GIS

The integration process in the demonstration GIS involved the inclusion of data elements that can be used to connect other database files. Figure 4.5 shows an example of how map, performance, design, and layer properties data have been integrated. The attribute table of the map (Attributes of 2005 STH Network) is linked with the performance data (IRI history) using the two fields that describe identical sequence number for segments. The *IRIhistory* is in turn linked to the design database (dsgacfield2000to2004.dbf) using the same sequence number description fields. The design database on the other hand, is linked with the layer properties database *dsgacoffice2000to2004.dbf* through fields that describe the contract number for projects (Contract2). After integrating the different databases, a query to identify all flexible pavements in the Southwest Region surfaced between 2000 and 2005 was performed. This query was performed only on the *IRIhistory* table, and since all the databases were integrated, the selected pavement segments (highlighted in yellow) could be tracked in all database tables as shown in Figure 4.5.

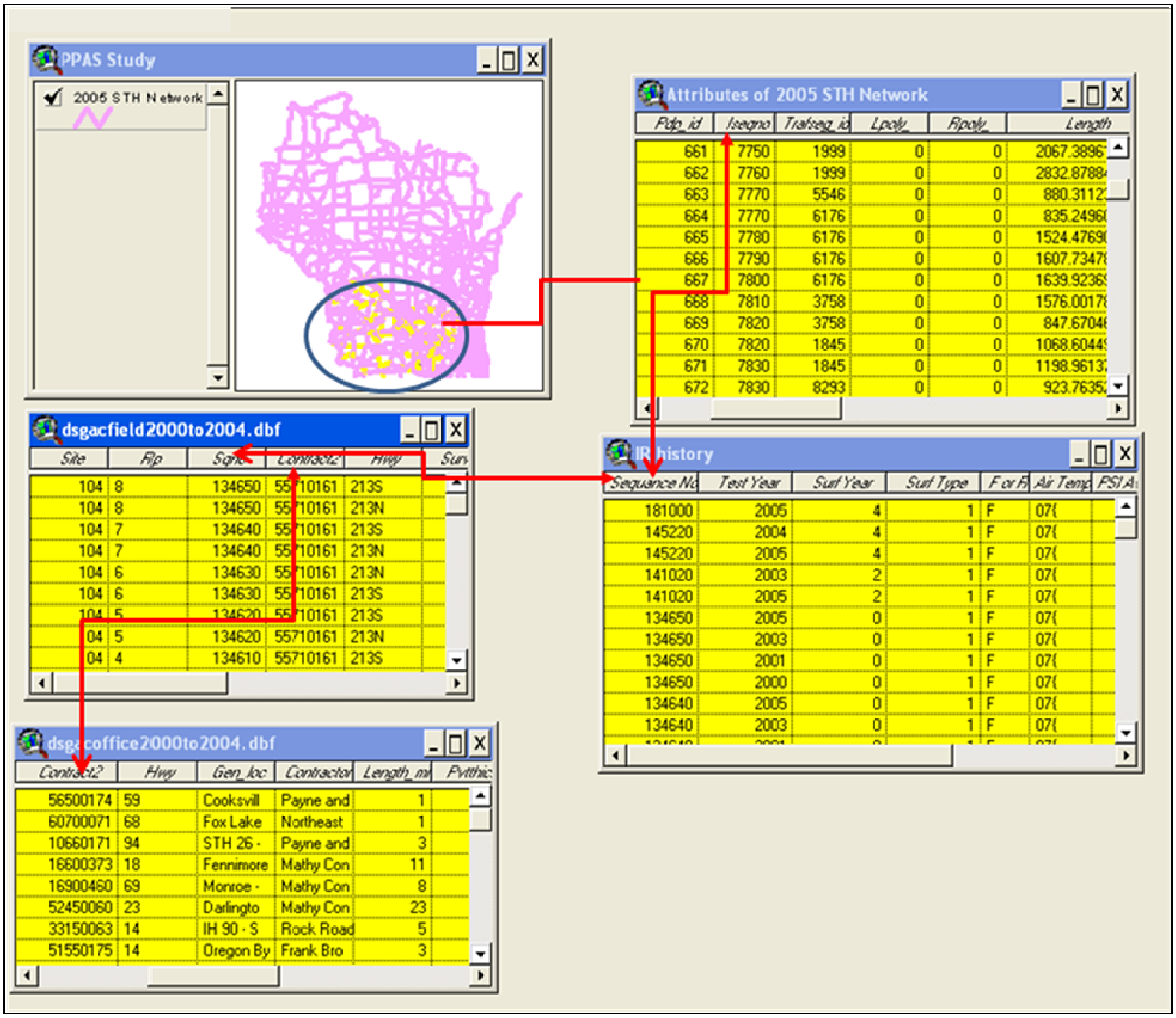

Figure 4.5 Integrated Databases for Flexible Pavements in Southwest Region

4.4.3.1 GIS Data Processing

The framework presented in Figure 4.4 requires GIS data to be processed by organizing, formatting, conducting basic analysis in GIS, and exporting the formatted data to advanced statistical software for modeling. The organization and formatting of the selected records (highlighted in yellow in Figure 4.5) involved two stages. First, the results in each database table were exported as a separate *dbase* file and then brought back into the GIS. The second stage involved combining all the files into a single table using a "Join" operation tool in GIS. At this time, fields not relevant to the performance modeling process were turned off.

4.4.3.1.1 Basic Data Analyses in GIS

Sample basic analyses conducted in GIS include simple bar charts, basic statistics, and tabular summaries of data. A bar chart and summary statistics of rut depth values for newly constructed flexible pavements in the Southwest Region are respectively shown in Figures 4.6 and 4.7. Lower ESAL pavements, along with either HV or MV pavements, had a greater mean rut depth. Although these simplified plots lack statistical rigor, they provide a basis from which to start. Figure 4.8 is a summary of age and thickness characteristics of pavements constructed between 2000 and 2005.

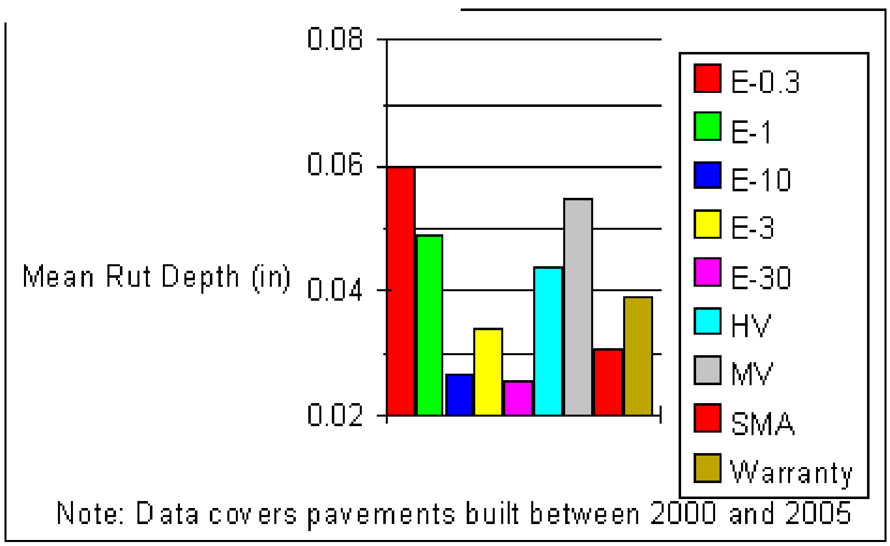

Figure 4.6 Average Rut Depths for Newly Constructed Pavements

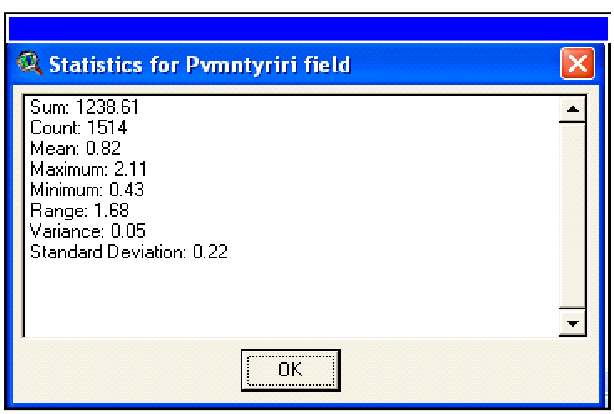

Figure 4.7 Basic Statistics for Newly Constructed Pavement IRI

sum5.dbf

Mix Category	Mean Age	Mean Surface Thickness
E-0.3	1.20	3.60
E-1	1.54	4.03
E-10	1.37	4.63
E-3	1.60	4.83
E-30	1.12	4.00
HV	2.23	4.20
MV	2.24	4.65
SMA	1.38	3.00
Warranty	1.19	5.09

Figure 4.8 Age and Surface Thickness Summary

4.4.3.1.2 Data Analysis Using Advanced Statistical Software

The combined table described earlier was exported as a dbase file into advanced statistical software (STATGRAPHICS™) for performance modeling. Sample basic plots are shown in Figures 4.9 and 4.10, while a model relating rut depth progression and other influential factors (age, surface thickness, flexible pavement type, and measured rut depth immediately after construction) is shown in Table 4.5. The model was further imported back into the GIS to generate map-based reports such as presented in Figure 4.11, which shows segments in the Southwest Region that have average rut depth greater than 1/8 inch at age six years.

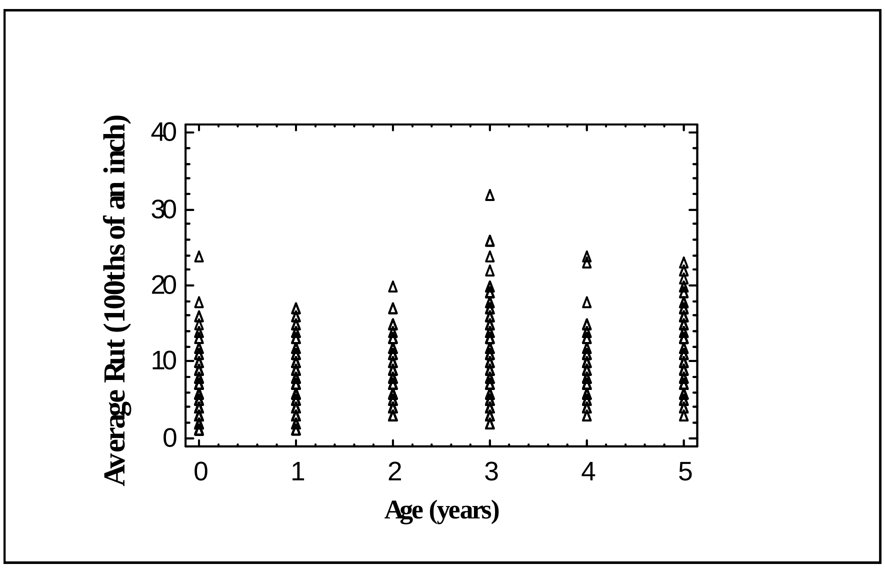

Figure 4.9 Mean Rut Depth Variations with Surface Age

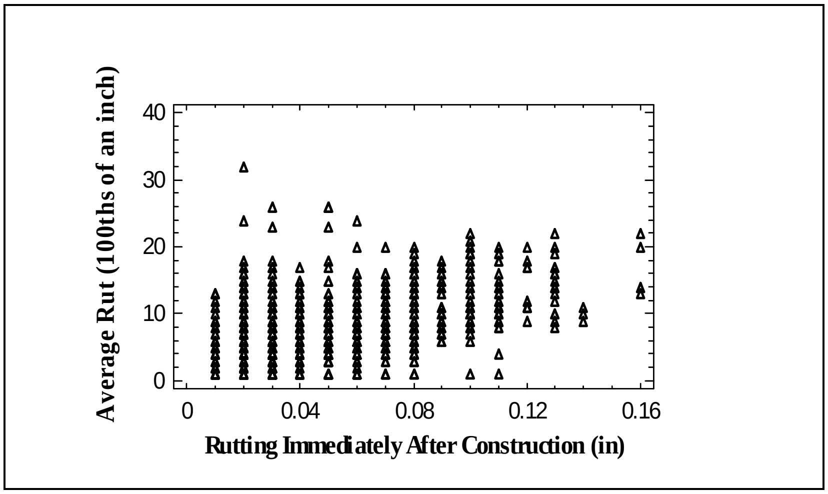

Figure 4.10 Rutting Progression Based on Initial Rut Depth at Construction

Table 4.5 Performance Model for Flexible Pavements in the Southwest Region

Model Form (1)	Adj. R^2, % (2)	DF (3)
RUT_AVG = 3.08024 + 1.35802**Age* + 68.4144**IRUT* -0.31183**h* - 1.63926**STyp*	49.1	1506
RUT_AVG = Average rut depth (100ths of an inch) *Age* = Pavement age in years *IRUT* = Construction year RUT depth (inches) *h*= Pavement surface thickness (inches) *STyp* = flexible pavement type (1 = type 3, 0 = type 1) *DF* = Degrees of Freedom to develop model		

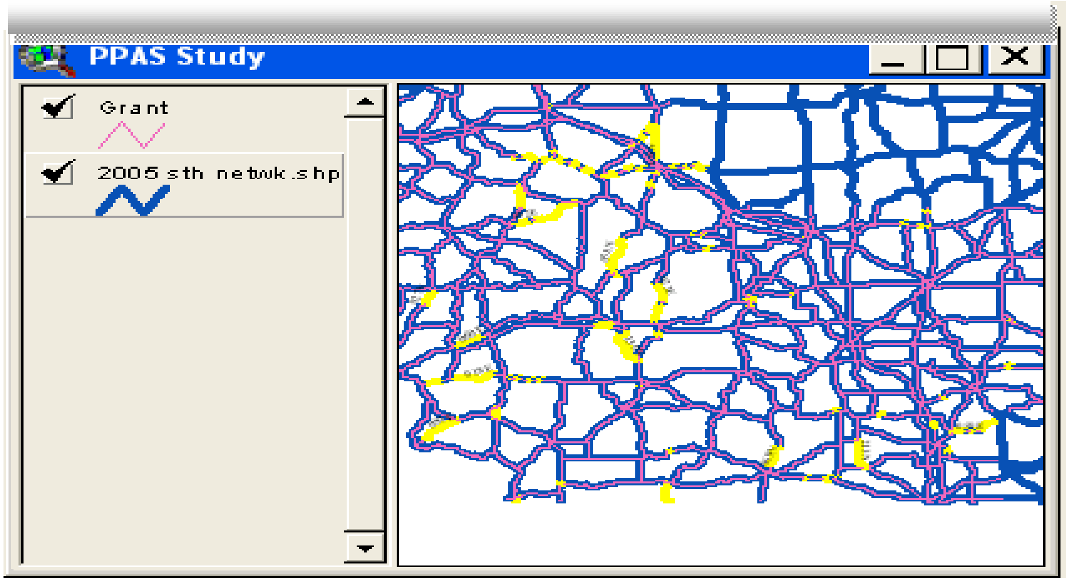

Figure 4.11 Segments with Rut Depth in Excess of 1/8-in at Age 6 Years

CHAPTER 5 Guidelines for Development of Performance Models

Performance models are used as a tool by pavement engineers to determine the present and future conditions of a network of pavements. It is common practice for agencies to set target performance values for their pavements, and periodically, determine the proportions that meet, exceed or fall below the set target. This information is further used to determine appropriate maintenance and rehabilitation options, as well as plan program work load and corresponding budget.

The purpose of this chapter is to provide guidelines for developing performance models. The guidelines identify a set of issues that must be addressed prior to engaging in any active modeling of performance. Traditional statistical methods relevant to performance modeling are reviewed and sample analyses pertaining to the statistical methods are presented. All sample analyses are based on flexible pavement performance data gathered between 2000 and 2005 in the Southwest Region of Wisconsin.

5.1 Considerations for Performance Model Development

Issues involving performance of pavements often differ in focus and scope depending on the agency (i.e., state, county, and local) and the management level involved. Knowledge about the pertinent issues can provide a basis for identifying performance data needs and analytical methods to support decision-making at any of three organizational levels including technical, administrative, and legislative levels. The administrative and legislative organizational levels tend to emphasize justification for budget requests, while the technical level focuses on the data requirements for decision-making at the various levels. Prior to developing performance models, it is essential to consider a number of issues such as:

a. Indicator of performance (IRI, PDI, specific distress, deflection etc).

   Performance can be classified as functional or structural. Functional performance relates to the ability of pavements to provide a smooth safe ride, whilst structural performance deals with the ability to withstand traffic and environmental loads. If functional performance is of interest then an indicator such as IRI or skidding friction will be required compared to deflection or a specific load related distress where structural inadequacy of pavements is what is of interest to the modeler. In these instances, the influential factors, and hence, data needs might be different.

b. Intended purpose of the performance model.

   If the intended purpose of the model, for example, is to describe the existing network condition, then analytical methods such as summary statistics or basic plots for the performance indicator may be all that is needed and no effort is required in acquiring data that may explain the condition. If future network conditions are required, then a full performance model dependent on an extensive database will be required.

c. Type of administrative, legislative, and technical decisions that model will support.

   At the administrative or legislative level, the developed model must be capable of addressing various funding scenarios on the status of the network or determine the funding level to keep network in some specified condition. In addition, the model must be capable of addressing the short-term and long-term impacts on the present and future status of the network if maintenance is deferred by funding budget request at a future year. At the technical level, the model must be capable of predicting network performance values for comparison with prescribed target values to enable appropriate maintenance or rehabilitation strategies to be developed. The decision categories listed point to the fact that a full performance model dependent on an extensive data will be required.

d. Availability and adequacy of data to characterize a model to meet the intended purpose.

   The more data there are, the better the model explains variations in performance. Data sufficiency can be explored using statistical methods that focus on sample size determination.

e. How to assess model effectiveness (precision and accuracy).

   Statistical and model validation methods can be employed to judge model effectiveness. Randomly select a portion of the data to develop a model, then apply the remaining data to the model to assess goodness of fit.

f. How model-generated results will be presented to intended audience.

   It is essential that reports or information be well communicated to facilitate the understanding of the target audience. Depending on the intended audience, formats such as tabular summaries of data, graphical representations, and color-coded maps can be considered.

5.2 Model Development Framework

Once pertinent issues such as those listed under section 5.1 have been considered, the modeling process can begin. The general framework for performance modeling is shown by Equation 5.1, where performance, as measured by a particular indicator (IRI, PDI etc.), is treated as a dependent variable, which is dictated by design, construction, traffic, and environmental parameters.

Performance = f {Design, Construction, Traffic, and Environment} …… (5.1)

The modeling process involves the various ways in which the parameters on the right hand side of Equation 5.1 can be formulated and to check adequacy of the resulting model to describe the data. There is no single approach to developing performance models since issues differ in focus and in scope. Hence, the modeler has to adapt to features of the data and attempt to best describe the relationships between variables. The literature review identified the major types of performance models to include mechanistic, mechanistic-empirical, empirical, and probabilistic. Each of these model forms was calibrated using some sort of statistical approach.

Hence, in the following sections, relevant statistical methods applicable to modeling are presented and sample practical applications are demonstrated.

5.3 Review and Application of Relevant Statistical Methods for Performance Modeling

There are a wealth of textbooks on statistical science and the application of statistics to highway pavement design and construction. Several statistical methods available to understand the data and develop performance models, include: (1) analysis of variance, (2) comparison of means, (3) detection of outliers, and (4) regression analysis. Table 5.1 presents specific applications and the recommended statistical tool(s) recommended for a specific application. A brief description of these methods as they apply to HMA performance modeling follows.

Table 5.1 Statistical Applications and Recommended Approaches

Index (1)	Application (2)	Statistical Approach (3)	Equation (4)

1	*Determining whether data is enough to develop a performance model (i.e. how many pavement sample units/segments are needed)?*	*Statistical distributions and statistical parameters* can assist in determining sample size. Any number of samples is valid, provided they were randomly chosen. The t-distribution is used for small sample sizes when the population is not known (n<30), while the normal distribution (z statistic) is used for larger sample sizes when the population statistics are known (or reasonably well known). Precision and confidence interval estimation equations can provide an interrelationship of sample size, standard deviation, and confidence level.	Population parameters unknown and only estimated
			$n = \left(\frac{t(s)}{PREC}\right)^2 \quad n<30$
			Population parameters known
			$n = \left(\frac{z(\sigma)}{PREC}\right)^2 \quad n>30$
			Where,
			n = required sample size;
			t = standard sample variate with significance level α for a one-sided test or $\alpha/2$ for a two-sided test;
			s =sample standard deviation;
			PREC = desired level of precision.
			
			z = standard normal variates with significance level α for a one-sided test or $\alpha/2$ for a two-sided test;
			σ = assumed known (or reasonably well known) standard deviation.

Table 5.1 (Cont.)

2	*Developing confidence in an estimate (e.g., what proportion of pavement segments meet a targeted performance value in a specific year)?*	*Confidence Intervals* can be constructed around a target value using the chosen level of confidence (i.e., 95%), underlying variability, and sample size.	$C.I. = \overline{Mean} \pm \left(Z * \frac{\sigma}{\sqrt{n}} \right)$ Where, C.I. = confidence interval z = standard normal variates σ = assumed known (or reasonably well known) standard deviation. n = sample size
3	*Stratifying data in terms of pavement features*	*Analysis of Variance* can detect whether there is a difference between features of data (age group, region, pavement type, PG Binder Grade, etc.), while incorporating the variability into the determination. Significant differences between features can be detected using Analysis of Variance.	$F_{Feature} = \frac{\text{MS (Between Feature)}}{\text{MS (Within Feature)}}$ Where, MS Between = Mean Square between data groups; MS Within = Mean Square between data groups;
4	*Comparing results among feature categories*	*F-tests* and *t-tests* can provide a statistical comparison of means. *Paired-sample t-tests* are used when data are collected from the same pavement from year-to-year, while a *two-sample t-test* is used when the pavement segments are independent of each other.	$\left\| \bar{X}_1 - \bar{X}_2 \right\| < t_{\alpha,v} \sqrt{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2} \right)}$ Where: $\bar{X}_i$ = sample mean i; t = the value of t for the significance level α and the degrees of freedom v; $v = (n_1 + n_2 - 2)$; and n_i and S_i^2 = represent the respective size and variance for sample i.
5	*Looking for trouble signs in performance data*	*Outliers*, or data points that are abnormal from a distribution, can detect trouble signs. Several standard tests for outliers exist, and the chi-square or other goodness-of-fit tests can be used to check normality.	$Z_i = \frac{X_i - \bar{X}}{S}$ Where, X_i = data value i; $\bar{X}$ = the sample mean; and S = the sample standard deviation.
6	*Reporting pavement performance data*	Beginning with *simple fundamental statistical measures* is always the best start (*plot* the data, calculate the *average* and *standard deviation*, etc.). The *sampling design* largely drives if/how a statistically-valid analysis can proceed, so effort must be placed on sampling design at the beginning.	Line graph, bar chart. pie chart, table, and box-and-whisker plots.

5.3.1 Analysis of Variance

An analysis of variance provides the best tool to determine whether there is a difference between features within any stratum of interest. With factual-based knowledge of a difference, necessary action can then be taken to yield more consistent level of performance. In effect, all statistical computer software packages have routines that perform ANOVA. In Figure 5.1, an ANOVA is conducted using STATGRAPHICS PLUS™ to compare average rut depth progression for pavements falling into six age groups (0, 1, 2, 3, 4, 5 years). The ANOVA results are plotted as intervals around mean rut depth for each level of age. The intervals are based on least significant difference (LSD) procedure. They are constructed in such a way that if two means are the same their intervals will overlap 95% of the time. The results suggest that the different age groups exhibit different performance levels.

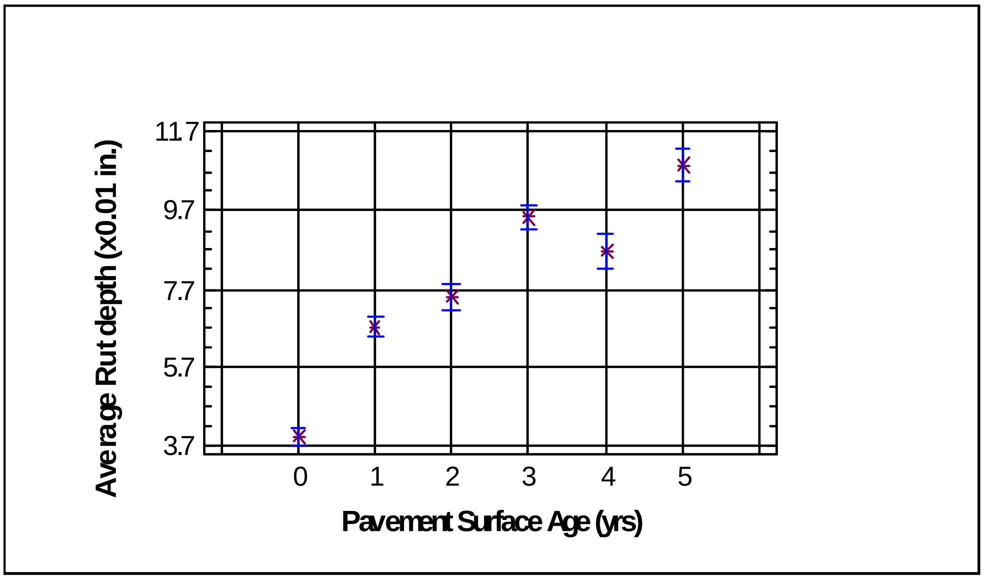

Figure 5.1 Means and 95% LSD Intervals for Rut Depth Progression by Age

5.3.2 Comparison of Means

Comparison of means can allow a determination if the mean level of one data feature is different than other features. In concept, this is similar to ANOVA, however the level of detection can be enhanced for split sample and independent sample t-tests. To ascertain whether significant differences exist in performance between different classes within a pavement system, a comparison of means test can be applied. The test focuses on the sampling distribution of the difference between sample means. According to Greenshields and Weida (1978), the difference in the population means can be tested by Equation 5.2.

$$\left| \bar{X}_1 - \bar{X}_2 \right| < t_{\alpha,v} \sqrt{\left(\frac{S_1^{\,2}}{n_1} + \frac{S_2^{\,2}}{n_2} \right)} \qquad (5.2)$$

Where:

t = the value of t for the significance level α and the degrees of freedom v;
v = ($n_1 + n_2 - 2$); and
n_i and S_i^2 = represent the respective size and variance for sample i.

If the inequality in Equation 5.2 holds, then the means represented by the two samples being compared may be considered equal with a level of confidence (1- α), otherwise the hypothesis that the means are equal is rejected. The squared-root term in Equation 5.2 represents the standard error of the difference between the two means, or the pooled standard deviation. If differences exist between the categorical data sets, a decision can be made whether the data has to be stratified by category and modeled separately or combined and modeled together using dummy variables to represent categories in the overall model developed using regression methods.

In the following example, it is examined whether the rutting progression of flexible pavement surface Type 1 and Type 3 is the same or not. Type 1 flexible pavements are placed over a flexible base, while Type 3 is placed over a rigid base. The basic statistics for the two surface types are shown in Table 5.2 and the t-test results shown in Table 5.3. Based on the results in Table 5.3, it can be concluded that the performance of the two surface types are different.

Table 5.2 Summary Statistics for Rutting Progression

Measure (1)	Type 1 (2)	Type 3 (3)
Count	1389	396
Average	7.24	5.24
Variance	21.35	7.77
Standard deviation	4.60	2.79
Minimum	1	1
Maximum	32	16
Range	31	15

Table 5.3 Test of Means for Surface Type Rutting Progression

$\lvert \text{Mean}_{\text{TYPE 1}} - \text{Mean}_{\text{TYPE 3}} \rvert$ (1)	Degrees of freedom, v (2)	T-value for 95% confidence and v = 1783 (3)	$\sqrt{\left(\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}\right)}$ (4)	Column (3) x Column (4) (5)	Are the means significantly different based on Columns (1) and (5)? (6)
7.24 - 5.24 =2.0	1389 + 396 - 2 = 1783	1.97	0.187	0.37	Yes

5.3.3 Looking for Trouble signs in Data

A potential problem in any data set is the presence of outliers. An outlier is an observation or measurement that is usually large or small relative to other values in a data set. Outliers typically are attributable to causes such as:

a) The measurement is observed, recorded, or entered into the computer wrongly.
b) The measurement is correct, but represents a rare event.
c) The measurement comes from a different population.

Each case is treated on an individual basis, with an appropriate course of action dependent upon the severity or error of the individual data point.

Identification of outliers is a useful tool for checking the validity of data. A common method used for outlier detection is the Z-score represented by Equation 5.3. It measures how many standard deviations each observed value deviates from a model fitted using all of the data except that observation. Data values are classified as outliers when the Z-score is greater than 3 or less than –3 . Data values that fall in these ranges can be reviewed for accuracy and a decision can be made whether they belong in the data set or not. With the availability of statistical software, such as STATGRAPHICS PLUS™, analyses can easily be performed to assess the validity of the data.

$$Z_i = \frac{X_i - \bar{X}}{S} \quad \text{....................(5.3)}$$

Where:

X_i = data value i;
$\bar{X}$ = the sample mean; and
S = the sample standard deviation

In the following example, outlier detection concepts are illustrated using a combination of basic plots and simple regression techniques. Basic plots are created to visually examine the nature of the relationship between the indicator of performance and each individual potential explanatory variable. By visual inspection of a plot, one can deduce whether the relationship is linear or nonlinear. The basic plot is often used in conjunction with simple regression techniques to aid in the type of data transformation to be made. The relationship between average rut depth progression and rut depth immediately after construction is examined through the plot shown in Figure 5.2. By inspection, the plot seems to suggest that rut depth progression increases in a linear or exponential fashion with increasing rut depth immediately after construction. The exact relationship can be explored using simple regression techniques, beginning with a linear relationship and making a comparison of the linear model with alternative models such as shown in Table 5.4. The model comparison suggests that the linear model is the best model based on the R^2 value. However, note that the transformed models, those having the regressor or independent variables converted to a multiple order polynomial or mathematical function, have a similar R^2 value. At this point, the modeler will initially select one of the models; however, the decision should not be based solely on the R^2 value. The simplicity of an untransformed model has the distinct advantage of being more easily interpreted by a practitioner. Thus, both accuracy and simplicity are important considerations in model selection.

To improve the selected exploratory linear model in Table 5.4, it is necessary to check for outliers by plotting the Z-scores (represented as the studentized residuals). In the example shown in Figure 5.3, the studentized residuals range from 2.8 to 7.1 (in absolute terms). The high-end indicates extreme outliers exist within the data. By purging the data to eliminate all outliers, a new studentized residual plot is shown in Figure 5.4 with corresponding comparison of alternative models shown in Table 5.5. By eliminating the outliers, the linear model R^2 value improved more than 4% as indicated in Table 5.5.

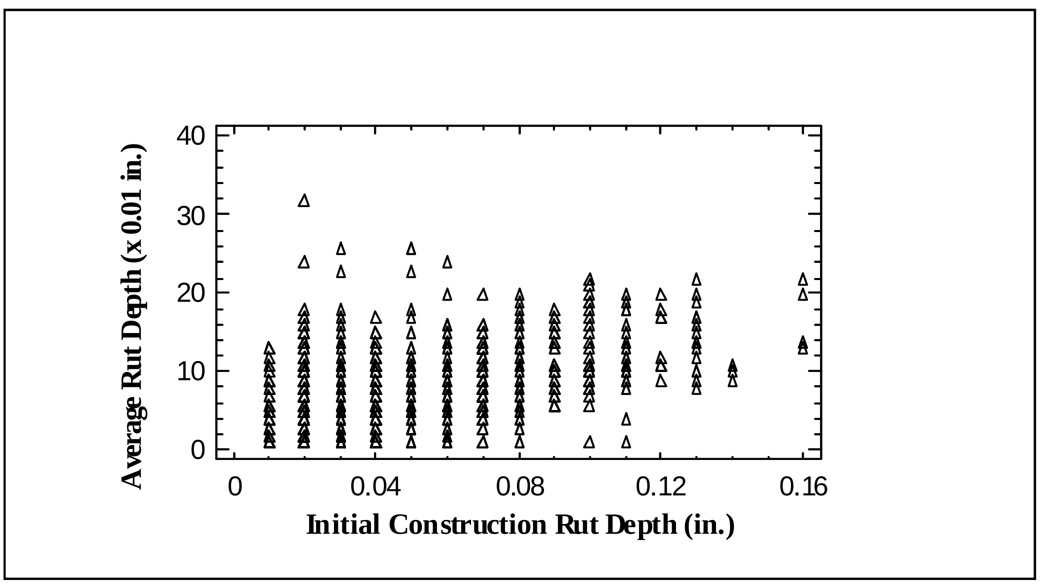

Figure 5.2 Basic Scatter Plot of Rut Progression after Initial Construction

Table 5.4 Candidate Models Prior to Accounting for Outliers

Model (1)	Correlation, coeff. (2)	R-Squared, % (3)
Linear	0.4704	22.1
Square root-Y	0.4696	22.1
Square root-X	0.4513	20.4
Exponential	0.4375	19.1
Logarithmic-X	0.4183	17.5
Multiplicative	0.4162	17.3
S-curve	-0.3397	11.5
Reciprocal-X	-0.319	10.2
Double reciprocal	0.2896	8.4
Reciprocal-Y	<no fit>	<no fit>
Logistic	<no fit>	<no fit>
Log probit	<no fit>	<no fit>

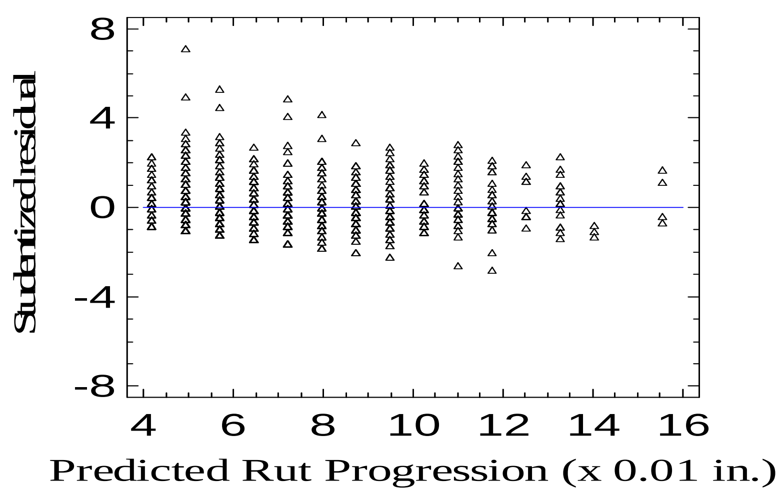

Figure 5.3 Studentized Residual Plot Prior to Accounting for Outliers.

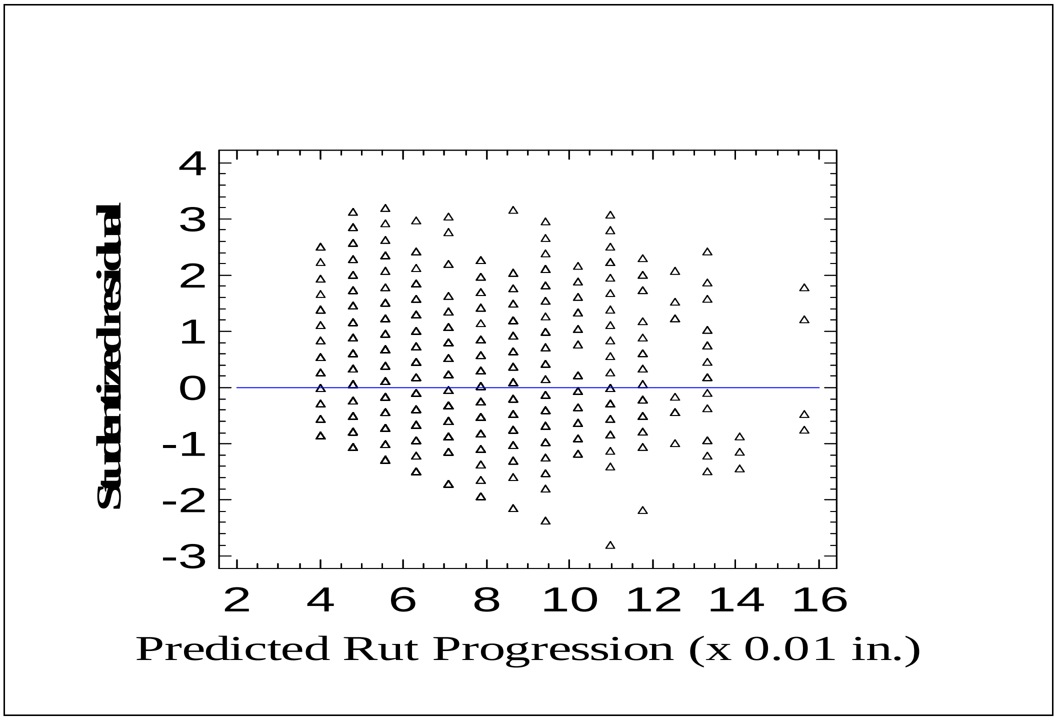

Figure 5.4 Studentized Residual Plot After Accounting for Outliers.

Table 5.5 Candidate Models after Accounting for Outliers

Model (1)	Correlation, coeff. (2)	R-Squared, % (3)
Linear	0.5132	26.3
Square root-Y	0.4947	24.5
Square root-X	0.492	24.2
Exponential	0.4553	20.7
Logarithmic-X	0.4509	20.3
Multiplicative	0.4283	18.3
S-curve	-0.3479	12.1
Reciprocal-X	-0.3443	11.8
Double reciprocal	0.2922	8.5
Reciprocal-Y	<no fit>	<no fit>
Logistic	<no fit>	<no fit>
Log probit	<no fit>	<no fit>

5.3.4 Performance Modeling Using Regression

Once the previously described exploratory data analyses have been performed, multiple linear regression (MLR) procedures can be used to model relationships between changes in key variables. MLR models are the most fundamental form of developing performance models, and have been exclusively used to develop AASHTO pavement design equations (AASHTO 2001). The basic equation for MLR models is shown in Equation 5.4.

$$Y = \beta_0 + \beta_1 * X_1 + \beta_2 * X_2 + \ldots + \beta_{p-1} * X_{p-1} + e \ldots\ldots(5.4)$$

Where,

Y = response variable (performance);
β_0 = regression constant (intercept for linear regression);
$\beta_{1,\,2,\,p-1}$ = variable constant (slope for linear regression); and
$X_{1,\,2,\,p-1}$ = predictor variables (e.g. density, permeability, traffic, region, age.); and
e= random error component of the model

In the MLR model building, the main objectives are:

a) To hypothesize the form of the linear model.
b) Estimate the unknown β-parameters.
c) Check whether the fitted model is useful for predicting the response variable, Y.

5.3.5 Forms of Multiple Linear Regression Models

Model forms for MLR could be described as first-order, second-order or quadratic, and interaction model. Equation 5.4 depicts a first-order form. A practical example involving a first order MLR is shown in Table 5.6, where average rut depth progression is related to pavement surface age, surface thickness, surface type, and rut depth immediately after construction. Further practical first-order examples that have been used to improve design procedures for paved shoulders adjacent to PCC pavements can be found in a WHRP report by Owusu-Ababio and Schmitt (2003).

Table 5.6 Sample First-Order Model Form

MLR Model Form (1)	Adj. R^2, % (2)	DF (3)
RUT_AVG = 3.08024 + 1.35802**Age* + 68.4144**IRUT* -0.31183**h* - 1.63926**STyp*	49.1	1506
RUT_AVG = Average rut depth (100ths of an inch) *Age* = Pavement age in years *IRUT* = Construction year RUT depth (inches) *h*= Pavement surface thickness (inches) *STyp* = flexible pavement type (1 = type 3, 0 = type 1) *DF* = Degrees of Freedom to develop model		

An interaction model is one in which the relationship between a response variable and an independent variable, depends on another independent variable held fixed. The form of an interaction model is as given in Equation 5.5 for two quantitative variables X_1 and X_2.

$$Y = \beta_0 + \beta_1 * X_1 + \beta_2 * X_2 + \beta_3 * X_1 * X_2 + e \text{ …… } (5.5)$$

Where,

Y = response variable (performance);
$\beta_1 + \beta_3 X_2$ = change in Y for every unit increase in X_1 holding X_2 fixed; and
$\beta_2 + \beta_3 X_1$ = change in Y for every unit increase in X_2 holding X_1 fixed.

The second order or quadratic model can be considered a special case of MLR where the model includes two terms, each including a single independent variable X. The form of this model is given in Equation 5.6.

$$Y = \beta_0 + \beta_1 * X + \beta_2 * X^2 + e \text{…….} (5.6)$$

Technically, the quadratic model contains only one variable, X, but can be considered linear in two independent variables with $X_1 = X$ and $X_2 = X^2$. The quadratic term (X^2), enables curvature to be hypothesized in the graph of the response model relating Y to X. A practical example involving a quadratic model is shown in Table 5.7.

Table 5.7 Second-Order Model Form

MLR Model Form (1)	Adj. R^2, % (2)	DF (3)
RUT_AVG = = 4.08707 + 2.4145*AGE-0.236814*AGE^2	28.6	1784
RUT_AVG = Average rut depth (100ths of an inch) *Age* = Pavement age in years *DF* = Degrees of Freedom to develop model		

5.3.6 Managing Qualitative Variables in Multiple Linear Regression Modeling

Qualitative variables, such as pavement surface type or geographical region, cannot be measured on a numerical scale. Hence, they have to be coded as values or levels before a model can be fit. The coded qualitative variables are called dummy variables since the numbers assigned to the various levels are arbitrarily selected. For a qualitative variable at two levels, a value of 1 will be assigned to one of the levels and a value of 0 to the other. Earlier, Table 5.6 presented a model with two qualitative variable levels for flexible pavement, where an indicator of 1 is assigned to type 3 and 0 assigned to type 1.

For models that involve qualitative independent variables at more than two levels, additional dummy variables must be created. In general, the number of dummy variables used to describe a qualitative variable will be one less than the number of levels of the qualitative variable.

5.3.7 Testing Model Adequacy

Once a MLR model has been fit, it is necessary to determine whether the model is useful for predicting the response variable (performance). A common way is to simultaneously conduct a test of hypothesis involving all the β-parameters, except the constant term, β_0. The test is defined in terms of the F statistic based on the null (H_0) and alternative (H_a) hypotheses, shown by Figure 5.5.

H_0: $\beta_1 = \beta_2 = \ldots\ldots\ldots = \beta_k = 0$

H_a: At least one of the parameters, $\beta_1, \beta_2, \ldots\ldots\ldots, \beta_k$ differs from zero.

Test statistic: F=Mean Square for Model/Mean square for error

$$= \left(\frac{R^2}{1-R^2}\right)\left[\frac{n-(k+1)}{k}\right]$$

Rejection region: $F>F_\alpha$

Where, n = Number of observations

k= number of parameters in the model (excluding β_0)

R^2 = multiple coefficient of determination

α= significance level

Figure 5.5 Test for Model Adequacy

If the overall model is determined useful for prediction using the F test, the modeler may elect to conduct one or more t-tests on the individual β-parameters. However, the test(s) should be decided prior to fitting the model. Sincich (1993) makes the following suggestions regarding t-tests for β coefficients associated with interaction or higher order models:

> "For interaction and higher order models, t-tests should be conducted only on the β coefficients associated with the interaction and higher order terms; no t-tests should be conducted for β coefficients associated with first order terms in the models. These terms should be kept in the model regardless of the magnitude of the p-values."

5.3.8 Detecting and Managing Multicollinearity in Regression Models

Multicollinearity occurs when two or more independent variables used in a regression model are correlated with each other. High correlations among the independent variables increase the likelihood of rounding errors in regression coefficients and standard errors. Sincich (1993) outlines some of the conditions that suggest multicollinearity in a model, including:

a. Significant correlations between pairs of independent variables in the model.
b. Nonsignificant t-tests for all (or nearly all) of the individual coefficients when the F-test for overall model adequacy is significant.
c. Opposite signs (from what is expected) in the estimated coefficients.

Approaches to dealing with multicollinearity according to SAS (2005) include:

a. Exclude redundant independent variables.
b. Redefine independent variables.
c. Use biased regression techniques such as ridge regression or principal component regression.
d. Center the independent variables in polynomial regression models.

Biased estimations produced by ridge regression or incomplete principal component regression are alternatives to ordinary least squares estimation performed by most statistical software packages. The basic idea behind ridge regression is to reduce the variances of the parameter estimates by applying a shrinkage parameter. Principal component regression is another biased regression technique. With principal component regression, the linear combinations of independent variables are dropped from the model predictor variables. A downfall of using the principal component regression is loss of simplicity and interpretation, however, the model is statistically correct.

In the previous examples, a simple approach would be to drop one or more of the correlated independent variables from the final model. If none of the variables is dropped, avoid making inferences about the individual coefficients based on the t statistic. Additionally, restrict predicted values to values of the independent variables falling within the range of the sample data.

5.3.9 Residual Analysis

Residuals are the differences between the observed response variable values and their corresponding predicted values based on the MLR model. Analysis of the residuals can provide information that can lead to modifications and improvements in MLR models. The modifications may be necessary if for example, the model has been misspecified or the data used to fit the model contain one or more unusual values (outliers).

To check whether a model has been misspecified, a plot of the residuals against each independent variable may be needed. A curvilinear trend detected in a plot, for example, implies that a higher order term (e.g., quadratic term) for that particular independent variable will probably improve model adequacy.

For outlier detection, a plot of the residuals against the predicted value is required. The criterion for outlier detection was presented earlier. Prior to eliminating an outlier from the analysis, it is necessary to conduct an investigation to determine its cause. If the outlier is the result of a recording error, it has to be fixed or removed. Otherwise it is necessary to determine how influential the outlier is before deciding whether to include or exclude it. If no outliers exist but the plot exhibits a pattern, then an appropriate variance-stabilizing transformation may be considered for the response variable. For example, if a plot reveals that the range in values of the residuals increases as the predicted values increase then consideration may be given to the use of logarithmic transformation on the response variable.

5.4 FHWA Research Products

Currently, there is an on-going research study, NCHRP Project 9-22, *Beta Testing and Validation of HMA PRS*, that will develop software specifically created to develop pavement performance models (NCHRP 2007). The objective of NCHRP 9-22 is to develop HMA PRS software and validate it with QC/QA data from actual field pavement construction projects.

The alpha version of the HMA PRS includes two application levels. Level I is based on material and construction properties (e.g., asphalt content; gradation; field-mixed, laboratory-compacted volumetrics; in-place air voids; and ride quality) currently obtained by public agencies for materials-and-method, end-result, and QC/QA types of specifications. Direct regression equations relating these properties to pavement performance (specifically, permanent deformation and fatigue cracking) that were exhibited in the WesTrack experiment are the primary basis for calculating pay factors in the Level I HMA PRS. The Level II HMA PRS uses a more sophisticated, mechanistic-empirical analysis of the results of laboratory performance tests, as well as the WesTrack property-performance relationships, to determine pay factors. Regardless of whether the Level I or Level II performance model is used, the HMA PRS calculates pay factors by comparing the life-cycle cost of the as-designed and as-built projects. This method is a significant improvement over current specifications, as the HMA PRS provides tools for objective calculation of equitable, consistent pay factors.

At the time of this report, the NCHRP 9-22 research team is working on PRS software developed from the M-E Pavement Design Guide. NCHRP expects the project to be completed by the end of 2008. Project deliverables are expected to advance the understanding of HMA inputs and resulting performance.

CHAPTER 6 CONCLUSIONS AND RECOMMENDATIONS

6.1 Conclusions

The primary purpose of this report was to develop a database template, using the existing Wisconsin DOT pavement management system, from which to perform pavement performance analysis using design, construction, and performance data for hot-mix asphaltic pavements. A second purpose was to investigate appropriate numerical or statistical methods that have the potential of quantifying and establishing relationships between design, construction, and performance data.

The literature review found that data types collected for performance evaluation and modeling vary from agency to agency depending on needs but the most common ones include inventory, condition, traffic volume, and maintenance and rehabilitation. A major barrier for achieving full data integration is lack of common referencing systems compounded by the use of different data formats. Many agencies have used Geographic Information System as an effective tool to integrate data.

Several Wisconsin DOT databases applicable to performance modeling for hot-mix asphaltic pavements were reviewed for primary data categories including construction, design, traffic, and performance. Semantic discrepancies exist among databases that impede integration were summarized, and an example relating the physical location of as-built construction properties to the reference point system was presented.

A GIS-based data integration example was provided using several WisDOT databases. A loose coupling approach, involving the transfer of data files between the GIS and other programs, was demonstrated using screen snapshots from a typical integration. Then, the integrated data were prepared for export into a statistical analysis package from the GIS and the results imported back to the GIS for data visualization or display.

Several statistical analysis methods to develop performance models were provided, along with reference examples for ANOVA, comparison of means, and regression models. Currently, there is an on-going research study, NCHRP Project 9-22, *Beta Testing and Validation of HMA PRS*, that will develop software capable of developing pavement performance models. It is expected that the report and software will be completed by the end of 2008.

6.2 Recommendations

From the findings and research in this study, the following recommendations are made:

1. Geographic Information System (GIS) is an effective tool for data integration among various divisions within an organization.

2. Databases can be integrated using a loose coupling approach, involving the transfer of data files between the GIS and other programs. Then, the integrated data can be prepared for export into a statistical analysis package from the GIS and the results imported back to the GIS for data visualization or display.

3. Further research is recommended to determine the appropriate assignment of as-built construction data to a given reference point location or sequence number location on the highway network.

4. With approximately 18 weather stations in the state, each containing about 10 years of climatic data, it is recommended that an investigation determine the appropriate assignment of weather station data to specific pavement segments.

5. Begin to develop performance models for WisDOT using approaches provided in this report. Examples were provided for HMA pavement rutting, however, numerous other performance measures can be modeled as well.

6. Monitor developments in NCHRP Project 9-22, *Beta Testing and Validation of HMA PRS*, that will develop software capable of developing pavement performance models – expected date December 2008.

References

AASHTO, “Pavement Management Guide”, AASHTO, Washington, D.C., 2001

Asphalt Institute “Thickness Design—Asphalt Pavements for Highways and Streets”, Manual Series No. 1, 1981.

Brimicombe, A., “GIS-Environmental Modeling and Engineering” , Taylor & Francis, London, 2003.

Corwin, D.L., Vaughan, P.J., Loague, K., “Modeling Nonpoint Source Pollutants in the Vadose Zone with GIS”, Environmental Science and Technology 31:2157-75, 1997.

Darter, M.I., “Requirements for Reliable Predictive Pavement Models”, Transportation Research Record 766, Transportation Research Board, Washington, D.C., 1980.

Day, J. and Lewis S., “Automated Highway Data Collection: Best Practices Field Review,” Presented at the GIS for Transportation Symposium, Atlanta, Ga., Mar. 25-27, 2002.

DeLisle, R.R., Sullo, P., and Grivas D.A., “Network-Level Pavement Performance Prediction Model Incorporating Censored Data”, Transportation Research Record 1853, TRB, National Research Council, Washington, D.C., 2003.

Federal Highway Administration, “Data Integration Primer,” Office of Asset Management, Federal Highway Administration, Washington, D.C., August 2001.

Federal Highway Administration, “An Advanced Course in Pavement Management Systems,” FHWA, Washington, D.C., 1991.

Greenshields, B.D. and Weida, F.M. (1978). “Statistics with Applications to Highway Traffic Analyses”, ENO Foundation for Transportation, Inc., Westport, CT.

Haas R., Hudson W. R., and Zaniewski J., “Modern Pavement Management”, Krieger Publishing Company, Malabar, FL., 1994

Harris, David, “Systems Analysis and Design for the Small Enterprise”, Second Edition. Orlando, FL: Harcourt Brace & Company, 1999.

Huang, Y.H., “Pavement Analysis and Design”, Prentice Hall, Upper Saddle River, NJ, 2004.

Lytton, R.L., “Concepts of Pavement Performance Prediction and Modeling.” Vol. 2, Proceedings, Second North American Conference on Managing Pavements, Ontario Ministry of Transportation, Toronto, Canada, 1987.

National Cooperative Highway Research Program Project 9-22, *Beta Testing and Validation of HMA PRS*, http://www.trb.org/TRBNet/ProjectDisplay.asp?ProjectID=958, Washington, D.C., November 2007.

Owusu-Ababio, S., and Schmitt, R.L., "Paved Shoulders Adjacent to Concrete Pavements: Final Report", Wisconsin Department of Transportation, Wisconsin Highway Research Program Report 03-03, http://www.whrp.org/Research/Rigid/rigid_0092-02-05/index.htm, Madison, WI, 2003.

Park, Hee M., and Kim, Y.R., "Prediction of Remaining Life of Asphalt Pavement with Falling-weight Deflectometer Multiload-level Deflections Transportation Research Record 1860, TRB, National Research Council, Washington, D.C., 2003.

Petzold, R.G. and Freund, D.M., "Potential for Geographic Information Systems in Transportation Planning and Highway Management," Transportation Research Record 1261, TRB, National Research Council, Washington, D.C., 1990.

Queiroz, C., "A Mechanistic Analysis of Asphalt Pavement Performance in Brasil," J. of Assoc. of Asphalt Paving Technology, Vol. 52, 1983.

SAS (2005). "Modeling with ANOVA and Regression, Training Manual", SAS Institute, Cary, NC.

Schmitt, R.L., Rao, C., and Von Quintus, H.L., "Non-Nuclear Density Testing Devices and Systems to Measure In-Place Asphalt Pavement Density," Final Report 06-12, http://www.whrp.org/Research/Flex/flex_0092-05-10/index.htm, Wisconsin Highway Research Program, Madison, WI, July 31, 2006.

Schmitt, R.L., Owusu-Ababio, S., Crovetti, J., and Cooley, L.A., Jr., *Development of In-Place Permeability Criteria for HMA Pavements in Wisconsin*, Final Report No. 06-15, http://www.whrp.org/Research/Flex/flex_0092-06-02/index.htm, Wisconsin Highway Research Program, Madison, WI, March 1, 2007.

Sincich, T., "Statistics by Example", Macmillan Publishing Company, New York, NY, 1993.

Star, J. and Estes, J.E., "Geographic Information Systems: An Introduction", Prentice Hall, Upper Saddle, NJ, 1990.

Transportation Research Board, "Pavement Management Applications Using Geographic Information Systems", NCHRP Synthesis 335, Washington, D.C. 2004.

Vils, Craig. Personal email correspondence. November 14, 2007.

Virginia Department of Transportation, "State of the Pavement—2002", Interstate and Primary Highways, Maintenance Division, Pavement Management Program, VDOT, Richmond, June 2002

Wisconsin State Cartographer's Office, "Wisconsin Coordinate Systems", State Cartographer's Office, Madison, WI., 1995.

Wisconsin Department of Transportation, "Facilities Development Manual," http://roadwaystandards.dot.wi.gov/standards/fdm/index.htm, Madison, Wisconsin, 2007.

Yang, J., Lu, J.J., Gunaratne, M., and Xiang Q., "Forecasting Overall Pavement Condition with Neural Networks-Application on Florida Highway Network", Transportation Research Record 1853, TRB, National Research Council, Washington, D.C., 2003.