TTI: 0-6835

Test Procedure for Validation of Automated Distress Data: Technical Report

Technical Report 0-6835-R1

Cooperative Research Program

TEXAS A&M TRANSPORTATION INSTITUTE
COLLEGE STATION, TEXAS

in cooperation with the
Federal Highway Administration and the
Texas Department of Transportation
http://tti.tamu.edu/documents/0-6835-R1.pdf

Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
FHWA/TX-16/0-6835-R1		

4. Title and Subtitle
TEST PROCEDURE FOR VALIDATION OF AUTOMATED DISTRESS DATA: TECHNICAL REPORT

5. Report Date
Published: July 2019

6. Performing Organization Code

7. Author(s)
Andrew Wimsatt, Emmanuel Fernando, Tom Freeman, Nasir Gharaibeh, and Francisco Dalla Rosa

8. Performing Organization Report No.
Report 0-6835-R1

9. Performing Organization Name and Address
Texas A&M Transportation Institute
College Station, Texas 77843-3135

10. Work Unit No. (TRAIS)

11. Contract or Grant No.
Project 0-6835

12. Sponsoring Agency Name and Address
Texas Department of Transportation
Research and Technology Implementation Office
125 E. 11th Street
Austin, Texas 78701-2483

13. Type of Report and Period Covered
Technical Report:
January 2015–August 2016

14. Sponsoring Agency Code

15. Supplementary Notes
Project performed in cooperation with the Texas Department of Transportation and the Federal Highway Administration.
Project Title: Center for Automated Distress Data Collection
URL: http://tti.tamu.edu/documents/0-6835-R1.pdf

16. Abstract
This project developed a performance-based specification for automated distress data collections systems based on analysis of data provided by the Texas Department of Transportation (TxDOT), analysis of data collected under this research study, and discussions with three experienced providers of distress data collection. The research team also assisted TxDOT in preparing a detailed scope of work or specifications to procure equipment and services. The research team recommends that this specification be revisited when the Department obtains RFP responses from the vendors.

17. Key Words
Pavement Management Information System, PMIS, Quality Assurance, Pavement Condition Survey, Rut Measurements

18. Distribution Statement
No restrictions. This document is available to the public through NTIS:
National Technical Information Service
Alexandria, Virginia
http://www.ntis.gov

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	72	

TEST PROCEDURE FOR VALIDATION OF AUTOMATED DISTRESS DATA: TECHNICAL REPORT

by

Andrew Wimsatt
Research Engineer
Texas A&M Transportation Institute

Emmanuel Fernando
Senior Research Engineer
Texas A&M Transportation Institute

Tom Freeman
Engineering Research Associate
Texas A&M Transportation Institute

Nasir Gharaibeh
Associate Professor of Civil Engineering
Texas A&M University

and

Francisco Dalla Rosa
Universidade de Passo Fundo, Brazil

Report 0-6835-R1
Project 0-6835
Project Title: Center for Automated Distress Data Collection

Performed in cooperation with the
Texas Department of Transportation
and the
Federal Highway Administration

Published: July 2019

TEXAS A&M TRANSPORTATION INSTITUTE
College Station, Texas 77843-3135

DISCLAIMER

This research was performed in cooperation with the Texas Department of Transportation (TxDOT) and the Federal Highway Administration (FHWA). The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the data presented herein. The contents do not necessarily reflect the official view or policies of the FHWA or TxDOT. This report does not constitute a standard, specification, or regulation.

This report is not intended for construction, bidding, or permit purposes. The engineer in charge of the project was Andrew J. Wimsatt, P.E. #72270.

The United States Government and the State of Texas do not endorse products or manufacturers. Trade or manufacturers' names appear herein solely because they are considered essential to the object of this report.

ACKNOWLEDGMENTS

This project was conducted in cooperation with TxDOT and FHWA. The authors thank Dr. Magdy Mikhail (TxDOT Maintenance Division) and Ms. Sonya Badgley (TxDOT Research and Technology Implementation Office) for their support and assistance during this study.

TABLE OF CONTENTS

Page

LIST OF FIGURES

Page

LIST OF TABLES

Page

CHAPTER 1. INTRODUCTION

The Texas Department of Transportation (TxDOT) currently has several contractors to manually collect distress data over the state highway network at a cost of over $2,000,000 annually. This survey is conducted by driving along the shoulder, in the lane, or in the grass adjacent to the roadway. The survey is labor intensive, takes about four months to complete, and exposes the survey crew as well as the driving public to accident risks.

Clearly, the need exists to examine the use of automated systems to collect pavement condition for the department's pavement management information system (PMIS) and other roadway geometric information for managing other elements of the state highway infrastructure. This project developed a performance-based specification for automated distress data collections systems based on analysis of data provided by TxDOT, analysis of data collected under this research study, and discussions with three experienced providers of distress data collection. The research team also assisted TxDOT in preparing a detailed scope of work or specifications to procure equipment and services.

This report consists of five chapters:

- Chapter 2 is a summary of quality assurance practices in automated and semi-automated pavement condition surveys.
- Chapter 3 contains an analysis of pavement distress data collected by four vendors. The vendors' data is compared to data collected by the researchers.
- Chapter 4 is an evaluation of automated rut measurements for network level collection of pavement rut depths.
- Chapter 5 contains a summary of the research and presents implementation recommendations.

CHAPTER 2. SUMMARY OF QUALITY ASSURANCE PRACTICES IN AUTOMATED/SEMI-AUTOMATED PAVEMENT CONDITION SURVEYS

INTRODUCTION

State departments of transportation (DOTs) collect pavement cracking data on a regular basis. This data, along with data on other condition indicators, is used to measure pavement performance and support pavement maintenance and rehabilitation decisions. Many state DOTs have changed their manual pavement distress survey to automated or semi-automated procedures. The reasons for this migration are linked to reduction in field staff (i.e., raters), coupled with increase in surveyed areas (e.g., 100 percent coverage) and safety risks of manual surveys (Fu et al., 2014; Vavrik et al., 2013).

Generally, these automated/semi-automated systems consist of image capturing technology (i.e., hardware) and image processing and analysis algorithms (i.e., software). These systems have evolved in the private sector as companies competing for automated distress survey contracts developed their own proprietary systems. The basic concepts of these systems are well-understood. However, the details (e.g., mathematical algorithms) are proprietary and may vary among commercial systems (Wang and Smadi 2011).

Highway agencies are concerned about the performance and quality of these automated/semi-automated surveys, and with the different procedures applied to vendor pre-qualification and Quality Assurance (QA). A review of quality assurance practices in automated/semi-automated pavement condition surveys is presented here.

DATA QUALITY ASSURANCE

This section summarizes current data quality assurance practices in terms of three categories:

- Quality of Images.
- Accuracy and Repeatability of Measurements.
- Data Delivery Requirements.

FHWA (Pierce et al., 2013) suggests that as soon as the agencies have defined the resolution, accuracy and repeatability requirements, the vendors should be invited to regularly re-evaluate their image system in order to make sure that the equipment represents the current technology and ensure the reliability of acquired data.

Quality of Images

Since visual pavement distress surveys were first performed using 35 mm roll film equipment, highway agencies and vendors have improved their data quality standards and requirements. While major advances have been made in the hardware systems of automated surveys, (Wang and Smadi, 2011) robust algorithms and software systems for identifying and classifying pavement distress types are lacking. Thus, these concerns need to be considered in vendor pre-qualification processes.

Commonly, agencies specify minimum requirements for the resolution of both downward and forward perspective images. These requirements are typically specified in terms of a combination of the following parameters:

- Minimum Image resolution (e.g., Caltrans, Louisiana, Georgia, and New Mexico).
- Minimum crack width from downward looking cameras (e.g., Caltrans, Louisiana, Georgia, and New Mexico).
- Minimum crack width from forward looking cameras (e.g., Caltrans, Virginia, North Carolina).
- Picture frames per mile.

Ranges for these parameters are shown in Table 1 for a sample of highway agencies. Forward perspective and right-of-way images can be used to obtain useful data that may not be captured by downward images (e.g., horizontal curves, grades, shoulders, and overall view of the roadway).

Table 1. Minimum Resolution Requirement for Collected Images.

Agency	Minimum Image Resolution for Forward Looking Cameras	Minimum Crack Width from Downward Looking Cameras	Minimum Crack Width from Forward Looking Cameras	Picture Frames per Mile for Forward Looking Cameras
California DOT Caltrans (2014)	1920 × 1080 pixels	1/7th inch	1/4 inch and wide cracks on the pavement 88 ft ahead of the survey vehicle	200 frames/mile (i.e., 26.4-ft interval)
Virginia DOT VDOT (2012)	NA	1/8th inch	1/4th inch	NA
British Columbia Ministry T. I. BCMOTI (2012)	1300 × 1030 pixels	NA	NA	176 frames/mile (i.e., 30-ft interval)
Louisiana LDOTD (2010)	1920 × 1080 pixels	1/13th inch (2 mm)	NA	200 frames/mile
Georgia GDOT (2012)	1920 × 1080 pixels	1/10th inch	NA	200 frames/mile
New Mexico NMDOT (2014)*	1920 × 1080 pixels	1/25th inch (1 mm)		200 frames/mile
North Carolina NCDOT (2011)	NA	1/8th inch	1/4th inch	NA

*Downward looking camera is required to have a minimum horizontal resolution of 4096 pixels wide across the 12 ft.

Other examples of acceptance criteria pertain to the quality of the camera's operation that has been specified by several agencies. Commonly, aspects such as clarity, brightness/darkness, dry pavement, replay and missing images are checked as specified, for example, in a request for proposal (RFP) issued by the Louisiana DOTD (LADOTD 2010).

Accuracy and Repeatability of Measurements

Highway agencies are using varying acceptance criteria for data delivered by vendors. Normally, these criteria are based on the accuracy and repeatability (precision) of distress values and/or overall index. Table 2 provides a sample of these requirements for cracking and distress ratings only. Other distress types, such as rutting and faulting, are not included in this table. McGhee (2004) detailed the concepts of accuracy and repeatability for quality management of automated pavement distress surveys. Accuracy and precision are normally evaluated based on the use of control and verification sites, which will be discussed later.

Table 2. Minimum Acceptance Criteria for Data.

Agency	Data Item (Cracking or Distress Index)	Accuracy	Repeatability
California DOT CALTRANS (2014)	Individual distress types	At least 85 percent of all the quantitative distress measurement and condition indicator values must be within the allowed tolerance. The analysis is based on the crew manually surveyed sections and desktop survey to measure cracks.	NA
Virginia DOT VDOT (2012)	Pavement condition indices & Individual distress types	The accuracy analysis is based on the desktop analysis of images provided by contractors. When 90% of the Contractor and agency determined indices for randomly selected sections are within 10 points. Or, the value determined from a D2S[(1)] evaluation of contractor and VDOT data.	NA
British Columbia Ministry T. I. BCMoTI (2012)	Pavement Distress Index (PDI)[(2)]	± 1 PDI value of manual survey.	± 1 standard deviation of the PDI values for five runs
Pennsylvania PennDOT (Pierce *et al.* 2013)	Individual distress types	To establish ground truth, three Pennsylvania DOT raters perform distress ratings and the ratings are averaged. ± 10 percent compared to PennDOT's survey.	± 5 percent run to run for three repeat runs
Oklahoma (ODOT 2010)	Distress rating	± 10 percent compared to Oklahoma DOT ratings from manual survey.	NA

(1) D2S: Difference two-sigma limit (d2s) is the difference between two individual test results that would be equaled or exceeded in the long run in only 1 case in 20 in the normal and correct operation of the method (ASTM C670-10).

(2) PDI is a modified version of the ASTM D6433 PCI method.

While uncommon in current practices of automated-semi-automated surveys, parameters commonly used for evaluating the reliability of binary (True/False) classifiers provide an additional method for assuring the quality of these surveys. As shown in Table 3, there are four possible outcomes for such a classifier: true positive (TP), true negative (TN), false positive (FP), and false negative (FN). These outcomes will be used to measure the classifier performance, as shown in Table 4.

Table 3. Outcomes of a Binary True/False Classifier or Detector.

		Detected	
		Yes	No
Actual	Yes	True Positive (TP)	False Negative (FN)
	No	False Positive (FP)	True Negative (TN)

Table 4. Parameters for Measuring Classifier Performance (Drosg 2009).

Parameter	Definition	Formula*
Sensitivity (also called Recall)	Proportion of actual positives which are correctly identified as such.	TP/(TP+FN)
Specificity	Proportion of negatives which are correctly identified as such.	TN/(TN + FP)
False Alarm Rate (also called false positive rate)	The complementary of specificity.	FP/(TN+FP)
Precision	Probability that a positive detection is correct.	TP/(TP + FP)

*TP=True Positive, TN=True Negative, FP=False Positive, FN=False Negative

Data Delivery Requirements

Typically, highway agencies request that data be delivered in standard industry file format or non-proprietary format, which are specified in the RFP issued to the vendors. For example, CALTRANS (Fu et al. 2013) has required that all measurements must be delivered in a software-neutral fashion to avoid software incompatibilities and data loss should CALTRANS decide to change vendors and/or software. On the other hand, NCDOT (2011) as well as Virginia DOT (VDOT, 2012) require the vendor to provide a semi-automated visual review of all pavement sections prior to submitting results. This assessment will both check the reasonableness of the automated ratings and the findings of crack or joint sealant, patching, bleeding and other distress that an automated system may not identify.

The Iowa DOT (Iowa DOT 2010) has separate requirements for delivering data and images. For images, Iowa DOT requires delivering images that represent 5 percent (close to 250 miles) of the annual mileage collected on the state's roads. The process consists of 1-mile segments that are randomly selected by the agency. In addition, the agency has required that the vendor provide access to the software for agency staff to manually check and identify distress types from the images to compare with results from automatically quantified distress types and severity using the same software. For attribute data, the quality control requirements include:

- A minimum of 98 percent of the collectable miles should be delivered to the DOT. Areas closed off for construction are not considered collectable miles.
- Of the delivered data, 100 percent of the description items are populated and accurate. Description items include: system, route, direction, and location (begin and end latitude/longitude).
- Of the delivered data, 98 percent of the sections are completely populated with data values, not including any expected limitations, such as IRI in low speed areas.

CALIBRATION AND VERIFICATION REQUIREMENTS

Control and verification sites are useful for assessing data quality and detecting data errors in pavement condition surveys. Normally, corrective actions are taken on a case-by-case basis, depending on the nature, severity, extent and cause of the problem (Pierce et al. 2013).

Equipment Calibration

Automated/semi-automated pavement condition survey systems include various hardware components, such as Distance Measuring Instrument (DMI), global positioning system (GPS), cameras, and accelerometers. These components are sensitive to environmental and operational factors (e.g., humidity, improper operation, vehicle bouncing), and thus should be checked and calibrated systematically. The calibration of equipment and confirmation of the right set-up is necessary to provide adequate assurance of the quality of the collected data (both attribute data and imagery).

Generally, if the agency owns and operates the equipment, it has more control on quality management. On the other hand, if the data collection service is provided through vendors, the agencies need to work closely with them in order to define the minimum requirements for calibration, routine checks, and documentation of quality management plans.

Normally, agencies require the vendors to periodically check and calibrate equipment as part of the quality management process. The procedure has focused on checking the correct functionality of DMI, GPS, cameras, and accelerometers. To determine whether the cameras (a key component of the hardware system) are working correctly, the effects of lighting conditions, shadows, reference synchronization, vehicle speed, and other related factors, need to be checked before the vendor starts the field services. Mraz et al. (2006) recognized that image quality is significantly influenced by the lighting conditions on both asphalt and Portland cement concrete pavements, in both semi- and fully automated systems.

In this process, the automated/semi-automated survey results are compared with those from the manual data acquisition. Agencies usually have chosen weekly calibrations. In some cases, the maximum period between each calibration has been a month. Table 5 shows some examples of current practice.

Table 5. Example Calibration Requirements.

Agency	Calibration Interval	Other Calibration Requirements
Virginia DOT VDOT (2012)	Weekly	The calibration schedule and record shall be furnished to the VDOT Project Manager on a weekly basis to demonstrate that the equipment is collecting within accepted variances.
GEORGIA (GDOT, 2012)	Monthly	Depending on the equipment condition/age, the calibration interval may be reduced to weekly basis.
FHWA LTPP Program	Monthly	Documentation includes daily equipment checks (tires pressures, bounce tests, buffer warm-ups, and so on), monthly and annual calibrations, problem reports, and daily operation reports.

Verification

Agencies have established specifications for control sites (used in pre-qualification processes) and blind verification sites (used for verification of actual production) as part of quality management of automated/semi-automated pavement visual distress surveys.

Table 6 shows examples of control site specifications established by a number of agencies for pre-qualification purposes. Normally, these sites are also used for equipment calibration purposes. In addition, these sites have been used to train and validate pavement distress raters. These sites are located on roadway segments where the agency or third-party personnel measure the pavement condition to define the reference values or ground truth.

The control site testing procedures are normally based on multiple runs (typically three runs at each site). For services that are outsourced, the agencies may require that its representatives be present each time the control sites are run or request that the results be sent to them electronically within a specified timeframe.

Table 6. Examples of Agencies' Control Site Requirements Used in Pre-qualification Processes.

Agency	Number of Sites	Site Length	Other Details
British Columbia MoTI (BCMoTI, 2012)	4 (AC[1])	0.5 mi (0.8 km)	Selected using prior year's survey data or control sections
Louisiana DOTD (LADOTD, 2012)	4 (AC) 4 (CRCP[2]) 4 (JCP[3])	0.5 mi (0.8 km)	Service provider is required to evaluate the control prior to proceeding to the next district.
IOWA DOT (IOWA, 2010)	4 (AC) 4 (PCC[4])	0.28 mi (0.46 km)	Each vendor will collect data on a maximum of 50 miles of pavements representing a cross section of Iowa pavement types. The test sections will be manually rated according to the new SHRP manual to provide a baseline with which to compare the automated results.
Pennsylvania DOT (Pierce et al. 2013)	4 (AC) 2 (JCP)	~ 0.5 mi (0.8 km)	Service provider must run each testing vehicle prior to acceptance for production testing.
OHIO DOT (Vavrik et al., 2013)	14 (AC) 11 (PCC[4]) 19 (AC/PCC)	≥ 1.0 mi (1.61 km)	Test sites are visually inspected, shortened as necessary, marked at the beginning and end, and located using GPS coordinates.

[1]AC – Asphalt Concrete
[2]CRCP – Continuously Reinforced Concrete Pavement
[3]JCP – Jointed Concrete Pavement
[4]PCC – Portland Cement Concrete

Generally, the procedures to verify the data quality from pavement condition surveys are based on the ground truth data to assess the accuracy and precision of distress measurements. These measurements are collected by trained raters and automated/semi-automated systems. The ground truth data are used by several agencies in order to assess the performance of vendors during the contract period. In addition, the ground truth surveys have helped to recognize whether or not new automated technologies are better than existing survey methods.

Similar to the control sites, verification sites have been used on subsequent cycles of data collection, where normally, the agencies have a history of the reference values at the specific sites. While the control sites are frequently located at a central region, the verification sites are usually spread throughout the geographic inventoried area. It is common for agencies to require that vendors check the control and verification sites at periodic intervals (e.g., weekly or daily) during the contract work as specified in the quality control (QC) plan. These periodic inspections may also identify the need for calibration of the equipment/method. Table 7 shows some examples of agency use of verification sites.

Although not fairly as common, agencies may also use blind sites whose locations are not disclosed to the data collection team in advance. As collection is completed in an area containing one of these unknown or blind sites, the agency requests the data for that segment of the network. The agency will have rated the distresses or manually measured the sensor data elements in advance to establish the reference values. The data collection team then submits the data, which is checked by an agency lead or QC rater. As an example, the British Columbia Ministry T.I. (BCMoTI, 2012) has specified the blind site number and location based on the contract quantities and Contractors routing schedule. The blind sites are usually programmed every three days during the surveys. Louisiana DOT (LADOTD 2010) requires that a verification site shall be selected and run on the first week of data collection in each new district. Oklahoma DOT (ODOT 2010) requires that evaluation of verification or control sites occurs once a week (which results in 6 to 10 evaluations per survey year).

PAVEMENT CONDITION ASSESSMENT PROTOCOLS

Protocols for pavement condition surveys are essential to achieving comparable results across agencies. Pierce et al. (2013) found that there are pronounced differences in distress ratings determined from the different protocols used by highway agencies. Thus, efforts have been made to standardize pavement distress definitions and ranking procedures. Accordingly, Table 7 summarizes some protocols presented by FHWA (Pierce et al., 2013). Table 7 illustrates this issue for only the definitions of longitudinal cracking and alligator cracking, as examples. Similar concerns exist for other cracking types.

Table 7. Differences in Existing Definitions of Pavement Cracking (Longitudinal and Alligator Cracking in Flexible Pavement as Examples).

Crack Type	AASHTO PP67-14-UL[(a)]	LTPP Distress Identification Manual[(b)]	ASTM D 6433-07[(c)]	PMIS Rater's Manual
Longitudinal	A crack at least 12 inch long and with a crack orientation between ±10 degrees	Cracks predominantly parallel to pavement centerline. Location within the lane (wheel path versus non-wheel path) is significant.	A crack parallel to the pavement's centerline or laydown direction.	Cracks which parallel the centerline may be rated as longitudinal cracks if they: are at least 1/8 inch wide, or show evidence of spalling or pumping, or have been sealed. Longitudinal cracking that is greater than 2.0 inches wide or faulted greater than 2.0 inches is to be rated as a failure.
Alligator	PP67 does not provide a specific definition for alligator crack. However, it provides a definition for a pattern crack, as follows: A crack that is part of a network of cracks that form an identifiable grouping of shapes. This includes cracks that are not defined as transverse or longitudinal.	Occurs in areas subjected to repeated traffic loadings (wheel paths). Can be a series of interconnected cracks in early stages of development. Develops into many-sided, sharp-angled pieces, usually less than 0.3 m on the longest side, characteristically with a chicken wire/alligator pattern, in later stages. Must have a quantifiable area.	A series of interconnecting cracks. The cracks connect, forming many sided, sharp-angled pieces that develop a pattern resembling chicken wire or the skin of an alligator. The pieces are generally less than 0.5 m on the longest side. Alligator cracking occurs only in areas subjected to repeated traffic loading, such as wheel paths.	Alligator cracking consists of interconnecting cracks which form small, irregularly shaped blocks that resemble the patterns found on an alligator's skin. Blocks formed by alligator cracks are less than 1 ft by 1 ft (0.3 m by 0.3 m). Larger blocks are rated as block cracking. Severe alligator cracking is rated as a failure if the base layer is exposed.

a) AASHTO Standard Practice for Quantifying Cracks in Asphalt Pavement Surfaces from Collected Images Utilizing Automated Methods. AASHTO Provisional Standards, 2014.
b) Distress Identification Manual for the Long-Term Pavement Performance Project. SHRP Report 338, 2003
c) ASTM D 6433–07 "Standard Practice for Roads and Parking Lots Pavement Condition Index Surveys." 2007.

EXAMPLES OF PAST SIMILAR STUDIES

Ohio DOT Experience

The Ohio Department of Transportation - ODOT (Vavrik et al., 2013) published in 2013 a pilot study on the feasibility of moving from manual to semi-automated pavement distress data collection. The study used 44 representative test sites (including asphalt and concrete pavements) where distress measurements were collected by ODOT raters and three participating vendors. The findings from the ODOT study indicate moderate agreement between survey results from ODOT raters and results from the automated/semi-automated data collection vendors. However, the distress/severity and distress/severity/extent data between ODOT raters and vendors showed a low correlation. Based on discussions between the ODOT raters and vendors, it was concluded that better communication, training, and further field optimization could result in better correlations between manual and automated distress ratings in future surveys.

The ODOT study compared the benefits and risks of automated/semi-automated surveys (see Table 8). Form a cost standpoint, the study indicates that ODOT's manual survey is more cost-effective than automated/semi-automated surveys. However, if ODOT decides to migrate to automated distress data collection, it should consider the option of purchasing the equipment, and conducting the survey using in-house forces. However, the study recognized that this option would require a survey period of 50 weeks, compared to a survey period of 28 weeks by vendors.

Table 8. Summary of Benefits and Risks of Automated/Semi-Automated Surveys Based on ODOT Study (Vavrik et al., 2013).

Benefits	Risks
Increased rater safety	Losing the ability to directly correlate with some historical PCR data
Improved data accuracy for certain distresses	Becoming tied to technological evolution that forces early equipment replacement.
Enhanced timeliness of data collection and processing	Increased annual collection and processing costs
Ability to easily track, review and reprocess historical data and images	Difficulties associated with operational change
Ability to collect data compatible with HPMS requirements	Loss of control due to dependence on a single vendor
District access to vendors for ancillary data collection	Potential variability of vendor results year to year.
Consistent, well defined methods for future automated DSE identification	Additional initial costs and personnel demands associated with procurement, calibration, and implementation of system
Ability to combine IRI, rutting, and asset collection with pavement distress ratings	Breakdowns and long repair delays for ODOT-purchased equipment
	Additional costs associated with modifying the DSE ratings, distress manual, decision trees, pavement performance models, and PMS software.

CALTRANS Experience (CALTRANS 2009)

Caltrans has used automated/semi-automated surveys for its annual Pavement Condition Survey (which includes both distress and ride quality) since July 2009.

Caltrans has planned with the selected consultant to survey every lane-mile of the road network in the first year. In succeeding years, a partial survey was completed. The agency requested that the automated image and ride quality vehicles use a Differential Global Positioning System (DGPS) to record the time of data collection and local position of each data element.

Furthermore, Caltrans invites vendors to participate in an automated/semi-automated data collection demonstration (rodeo) on 10 short test sections, each 540-ft long. This demonstration is normally done on different pavement types, traffic volumes and representative levels of structural integrity and distresses that are commonly seen in California. From the analysis of the data from these sections, Caltrans is able to evaluate the data survey activities in the field (safety, efficiency, etc.), data processing and management, accuracy of reference system (DGPS), quality of images, pavement distress/conditions defined from the images and profiles, the data viewer supplied by the vendors, and the technical competence of the suppliers.

This Caltrans study found that vendors who use visible lighting and cameras have matched the results from the recent line-scan technology, resulting in images of sufficient quality to be used in automated/semi-automated visual distress systems.

CHAPTER 3. ANALYSIS OF AUTOMATED DISTRESS METHODS

INTRODUCTION

Researchers first conducted an analysis of data collected for TxDOT by two vendors in the Bryan and Houston Districts. TxDOT provided the data. That analysis resulted in a draft pilot specification that was submitted as Product P1 in March 2015.

This analysis examines the variability and accuracy of four automated distress methods and compares them to traditional manual PMIS and walk-along methods. Four vendors collected data as part of this study in late 2015 and early 2016. The variability analysis focuses on 48 asphalt pavement sections for which replicate (2 or more) ratings were conducted for all six methods. The accuracy analysis examines a larger collection of 239 asphalt and concrete sections to identify any significant bias in condition and distress ratings among the automated methods when compared to a ground truth (walk-along) ratings method.

The variability analysis identified several general characteristics; among them

- The wide range of standard deviation values seen within a given condition class of distress should be taken as a warning that the variability estimates are suspect in small sample ($2 \leq n \leq 3$) conditions.
- The automated methods for the most part display much lower variability than manual (PMIS and WALK) methods.
- One vendor (C3) appears to have the most variability among the automated methods, exhibiting variation comparable to the manual methods.

The accuracy analysis revealed several additional items of note:

- For asphalt pavements, two automated methods (C2 and C4) compare reasonably closely in ratings to the manual methods (PMIS and WALK).
- Poor condition pavement sections were underrepresented – this casts doubt on any trends observed across ranges of pavement conditions.
- Too few concrete segments were available to do detailed CRCP and JCP comparisons – this is reserved for asphalt pavements only.
- Automated ratings for JCP pavement show much greater inconsistency between different distress types than for asphalt or CRCP pavement types. This phenomenon may be an artifact of insufficient sample sizes for concrete pavements.

ANALYSIS OF MEASUREMENT VARIABILITY

This analysis examines the measurement variability in the condition scores and individual distress scores associated with asphalt pavement. This analysis is restricted to asphalt because only this pavement type had replicate ground truth (WALK) ratings performed. The overall condition score and six individual distress scores were examined – Patching, Failures, Block Cracking, Alligator Cracking, Longitudinal Cracking, and Transverse Cracking.

The primary focus of this analysis was the characterization of variability between two manual rating methods (Walk along, PMIS) and automated ratings from four vendors (referred to as C1, C2, C3, and C4) within the following Distress and Condition Score Class Descriptions:

Table 9. Condition Score Classes.

Condition Score	Class	Description
90–100	A	Very Good
70–89	B	Good
50–69	C	Fair
35–49	D	Poor
1–34	F	Very Poor

Table 10. Distress Score Classes by Type.

Distress Score	Class	Description	Patching (%)	Failures	Block Cracking (%)	Alligator Cracking (%)	Transverse Cracking (ft/100')	Longitudinal Cracking (ft/100')
90–100	A	Very Good	0–6	0–1	0–6	0–4	0–5	0–84
80–89	B	Good	7–13		7–11	5–8	6–8	85–124
70–79	C	Fair	14–25		12–20	9–14	9–12	125–172
60–69	D	Poor	26–95	2	21–50	15–29	13–19	173–235
1–59	F	Very Poor	> 95	> 2	> 50	> 29	> 19	> 235

Out of 230 total asphalt segments, only 48 segments had two replicate ratings for either manual method. Most of the automatic methods had three replicate ratings per segment. Therefore, when comparing the measurement variability or repeatability of each method, we will restrict the analysis to the 48 segments for which replicate measurements are available for the manual methods. Table 11 lists these segments.

Table 11. Segments Used for Measurement Variability Comparison.

Section	County	Highway	Subsection	Direction	Lane
3	Travis	FM 969	3.1	EB	K1
			3.2	EB	K1
			3.3	EB	K1
			3.4	EB	K1
			3.5	EB	K1
			3.6	EB	K1
			3.7	EB	K1
			3.8	EB	K1
			3.9	EB	K1
4	Travis	FM 969	4.1	WB	K6
			4.2	WB	K6
5	Travis	State Loop 111	5.1	NB	K6
			5.2	NB	K6
6	Travis	State Loop 111	6.1	NB	K6
			6.2	NB	K6
			6.3	NB	K6
			6.4	NB	K6
17	Williamson	US 183 Frontage Roads	17.1	NB	X1
			17.2	NB	X1
			17.3	NB	X1
			17.6	SB	A1
			17.7	SB	A1
25	Brazos	SH 47	25.1	SE	R1
			25.2	SE	R1
			25.3	SE	R1
			25.4	SE	R1
			25.5	SE	R1
			25.7	NW	L1
			25.8	NW	L1
			25.9	NW	L1
			25.91	NW	L1
			25.92	NW	L1
			25.93	NW	L1
26	Brazos	FM 60	26.1	EB	K1
			26.2	EB	K1
			26.3	WB	K6
			26.4	WB	K6
44	Milam	US 79	44.1	SW	K1
			44.2	SW	K1
			44.3	SW	K1
			44.4	SW	K1
			44.5	SW	K1
			44.6	SW	K1
			44.7	SW	K1
			44.8	SW	K1
			44.9	SW	K1
47	Milam	US 77	47.2	SW	K1
			47.3	SW	K1

Since the number of replications per method within each segment was never more than three, this limits the utility of any type of quantitative analysis of variance. Statistics that describe variability such as the standard deviation or variance are weakly consistent estimators; that is, they require larger sample sizes in order to insure repeatable results. Even the Median Absolute Deviation (MAD) is not wholly immune to the effects of extremely small sample sizes. Therefore, the discussion accompanying Table 12 through Table 18 will necessarily be more qualitative in nature.

Table 12. Two Measures of Average Variability for Condition Scores.

Class	Method	Average Standard Deviation	Average MAD	Number of Segments
A	C1	0.16	0.00	11
	C2	0.39	0.18	11
	C3	9.13	0.73	11
	C4	2.13	0.33	6
	PMIS	9.05	5.82	10
	WALK	0.99	0.32	5
B	C1	4.52	0.00	6
	C2	1.87	0.83	6
	C3	6.82	3.00	6
	C4	2.16	1.50	6
	PMIS	7.78	5.50	6
	WALK			0
C	C1	1.38	0.00	18
	C2	1.22	0.50	18
	C3	18.57	2.22	18
	C4	2.88	1.33	15
	PMIS	12.54	7.39	15
	WALK	24.16	11.39	12
D	C1	1.59	0.00	8
	C2	1.39	0.25	8
	C3	12.88	1.63	8
	C4	2.97	2.00	5
	PMIS	3.65	1.94	6
	WALK	13.67	3.63	3
F	C1	0.00	0.00	2
	C2	2.29	1.50	2
	C3	13.49	2.50	2
	C4	1.15	0.00	2
	PMIS	5.30	3.75	2
	WALK			0

Table 12 through Table 18 show two measures of variability in the condition and individual distress scores for the 48 segments, averaged within class. The Median Absolute Deviation (MAD) is a robust measure of variability, and is included along with the Standard Deviation because it is less susceptible to "blowing up" in the presence of outlying or anomalous data. This can be clearly seen in the Table 12 through Table 18, where in some cases the average standard deviation is 10 or more times larger than the average MAD.

Examining Table 12, one notices that PMIS, WALK, and C3 condition scores often show the greatest variability of the six rating methods. Note that the various condition classes are relatively well populated.

Examining Table 13, it is again apparent that PMIS, WALK, and C3 Percent Patching scores often show the greatest variability of the six rating methods. Here we see that segments with distress classifications below very good (Class A) are not very well represented in data. This makes it difficult to discern any valid trends in accuracy across distress classes.

Table 13. Two Measures of Average Variability for Percent Patching.

Class	Method	Average Standard Deviation	Average MAD	Number of Segments
A	C1	1.44	0.00	39
	C2	0.97	0.41	39
	C3	11.63	1.46	39
	C4	2.22	1.10	29
	PMIS	9.98	6.15	34
	WALK	17.23	5.94	19
C	C1	0.00	0.00	1
	C2	3.21	1.00	1
	C3	35.10	7.00	1
	C4	1.00	1.00	1
	PMIS	0.71	0.50	1
	WALK			0
D	C1	3.46	0.00	3
	C2	1.15	0.00	3
	C3	18.00	5.33	3
	C4	2.89	0.00	2
	PMIS	7.78	3.67	2
	WALK	8.49	2.00	1
F	C1	0.00	0.00	2
	C2	4.28	2.00	2
	C3	31.49	2.00	2
	C4	7.51	4.00	2
	PMIS	1.06	0.75	2
	WALK			0

In Table 14, the preceding variability characteristics are shown again for Failures among very good (Class A) segments. The alarming increase in variability seen for the manual methods may well be an anomaly resulting from small sample sizes.

Table 14. Two Measures of Average Variability for Number of Failures.

Class	Method	Average Standard Deviation	Average MAD	Number of Segments
A	C1	1.44	0.00	41
	C2	1.25	0.51	41
	C3	14.14	1.98	41
	C4	2.63	1.23	30
	PMIS	7.13	4.30	35
	WALK	4.15	1.15	16
D	C1	0.58	0.00	1
	C2	0.00	0.00	1
	C3	0.00	0.00	1
	C4	1.53	1.00	1
	PMIS	26.87	19.00	1
	WALK	66.47	47.00	1
F	C1	2.31	0.00	3
	C2	0.58	0.00	3
	C3	8.58	1.00	3
	C4	1.90	1.00	3
	PMIS	27.11	19.17	3
	WALK	67.65	47.83	3

Table 15 shows the ratings variability for Percent Block Cracking. Note that as block cracking increases, the manual methods (especially WALK) increase markedly. We will see later in the accuracy analysis that this most likely results from the fact that the ground truth method reports much higher block cracking for poorer condition segments than the automated methods do.

Table 15. Two Measures of Average Variability for Percent Block Cracking.

Class	Method	Average Standard Deviation	Average MAD	Number of Segments
A	C1	1.53	0.00	32
	C2	1.48	0.63	32
	C3	12.89	2.47	32
	C4	2.70	1.28	25
	PMIS	6.39	4.09	29
	WALK	2.59	0.52	9
B	C1	4.62	0.00	1
	C2	1.73	0.00	1
	C3	1.53	1.00	1
	C4	2.83	2.00	1
	PMIS	0.00	0.00	1
	WALK			0
D	C1	1.73	0.00	3
	C2	0.77	0.00	3
	C3	16.26	0.33	3
	C4	3.71	1.50	2
	PMIS	27.22	12.83	2
	WALK	35.12	24.83	3
F	C1	0.83	0.00	9
	C2	0.18	0.11	9
	C3	15.88	0.33	9
	C4	1.39	0.67	6
	PMIS	16.87	9.28	7
	WALK	25.90	16.28	8

Table 16 shows the ratings variability for Percent Alligator Cracking. As was the case for block cracking, note that increasing alligator cracking causes the variability of manual methods (especially WALK) to increase markedly. This most likely results from the fact that the automated methods significantly undercount the percent alligator cracking compared to that from the ground truth method.

Table 16. Two Measures of Average Variability for Percent Alligator Cracking.

Class	Method	Average Standard Deviation	Average MAD	Number of Segments
A	C1	1.46	0.00	28
	C2	0.91	0.32	28
	C3	15.85	2.07	28
	C4	2.86	1.24	21
	PMIS	8.43	5.32	25
	WALK	1.48	0.41	11
B	C1	0.14	0.00	4
	C2	2.37	1.25	4
	C3	9.36	0.75	4
	C4	0.58	0.00	1
	PMIS	9.19	3.25	2
	WALK	13.67	7.25	3
C	C1	4.04	0.00	2
	C2	1.63	0.50	2
	C3	2.08	1.50	2
	C4	3.06	2.00	1
	PMIS	23.33	8.25	1
	WALK	33.94	24.00	2
D	C1	2.42	0.00	5
	C2	0.55	0.20	5
	C3	10.88	0.80	5
	C4	1.71	1.00	5
	PMIS	16.97	12.00	5
	WALK	67.41	28.60	3
F	C1	0.77	0.00	6
	C2	2.03	0.83	6
	C3	10.98	2.67	6
	C4	2.30	1.33	6
	PMIS	3.42	2.42	6
	WALK	8.49	1.00	1

Table 17 shows the ratings variability for Longitudinal Cracking. Looking at the average MAD results, it appears that the variability of the manual methods remains relatively constant across classes. Once again, the C3 method is comparable in variability to the manual methods.

Table 17. Two Measures of Average Variability for Longitudinal Cracking.

Class	Method	Average Standard Deviation	Average MAD	Number of Segments
A	C1	0.75	0.00	30
	C2	1.00	0.43	30
	C3	15.07	0.90	30
	C4	2.77	1.29	21
	PMIS	9.38	5.75	26
	WALK	16.69	5.90	15
B	C1	2.39	0.00	7
	C2	1.58	0.29	7
	C3	8.41	4.29	7
	C4	1.95	0.60	5
	PMIS	9.76	4.93	5
	WALK	25.46	7.71	3
C	C1	5.43	0.00	5
	C2	0.88	0.20	5
	C3	7.67	3.00	5
	C4	2.47	1.80	5
	PMIS	9.48	6.70	5
	WALK	4.60	1.30	2
D	C1	0.00	0.00	3
	C2	2.55	1.67	3
	C3	18.78	4.00	3
	C4	1.97	0.67	3
	PMIS	5.89	4.17	3
	WALK			0

Table 18 shows the variability seen among the various methods for Transverse Cracking. Note that poorer condition segments are not very well represented. Looking at the average MAD results, the manual rating methods are consistently two to three times more variable than even the C3 method.

Table 18. Two Measures of Average Variability for Transverse Cracking.

Class	Method	Average Standard Deviation	Average MAD	Number of Segments
A	C1	1.56	0.00	36
	C2	1.04	0.44	36
	C3	14.03	1.89	36
	C4	2.56	1.04	25
	PMIS	9.50	5.60	30
	WALK	12.07	3.79	16
B	C1	1.01	0.00	8
	C2	1.96	0.63	8
	C3	12.48	2.00	8
	C4	2.41	1.63	8
	PMIS	9.02	6.38	8
	WALK	35.71	12.63	4
D	C1	2.31	0.00	1
	C2	0.00	0.00	1
	C3	0.58	0.00	1
	C4	2.83	2.00	1
	PMIS	0.71	0.50	1
	WALK			0

After the preceding discussion, one notices the following general characteristics among the six rating methods:

- The wide range of standard deviation values seen within a given condition class of distress should be taken as a warning that the variability estimates are suspect in small sample conditions.
- The automated methods for the most part display much lower variability than manual (PMIS and WALK) methods.
- C3 appears to have the most variability among the automated methods, exhibiting variation comparable to the manual methods.

ACCURACY ANALYSIS

This section examines the accuracy of the automated and PMIS rating methods with respect to ground truth (WALK). This section of the analysis restricted the number of segments to those which had at least one ground truth rating. Each pavement type (asphalt, CRCP, JCP) was analyzed separately. For these analyses, pairs of differences between the five other rating methods and ground truth were obtained for each segment's condition and individual distress scores. Any segments which had replicate scores for a given rating method had these scores averaged before differences were taken. Taking the differences between rating methods on the same section of roadway should, under optimal conditions, create a quantity which should be symmetric about zero and remove any extraneous factors which may give rise to heterogeneity between differences within each pairing and class group. This would imply that the distributions of these differences are unbiased (neither reporting higher nor lower scores with respect to each other) and that any nonzero differences that arise are due largely to random chance. The non-parametric Two-sided Wilcoxon Signed Rank test was used to determine if there were overall statistically significant differences between each method and ground truth.

The total numbers of sections for each pavement type having a ground truth (WALK) rating were:

- 204 segments for asphalt pavement types.
- 9 segments for CRCP pavement types.
- 26 segments for JCP pavement types.

Asphalt Accuracy Results

Table 19 shows the average percent difference in condition scores for 204 sections of asphalt pavement relative to ground truth (WALK method). For all methods, including PMIS, the average percent difference increases sharply once the condition score declines to poor or very poor, indicating that the 5 methods are overestimating the condition ratings in the case of poor or worse asphalt segments. This is countered somewhat at the opposite end, where most methods underestimate the condition score of fair or better segments. Note that C1 substantially overestimates condition scores for all ratings classed below very good. C2 and C4 condition scores remain largely within 10 percent of ground truth, with the exception of very poor condition segments. The last column in Table 19 reports the overall average percentage differences. Bolded numbers in this column indicate a statistically significant (p-value < 0.05) difference as determined by the Wilcoxon Signed Rank Test.

Table 19. Percent Differences in Condition Scores Relative to Ground Truth for Asphalt.

Class	C1	C2	C3	C4	PMIS	N
A	0.51	-10.64	-17.96	-9.72	-2.65	76
B	20.71	5.31	-29.15	-2.62	4.01	21
C	39.43	-4.46	-14.30	-9.12	-2.29	50
D	103.14	11.43	13.78	10.58	29.21	31
F	280.14	56.75	50.36	112.15	54.79	26
All Classes	**63.37**	4.46	**-4.69**	11.79	12.01	204

(Bolded Differences for All Classes are Statistically Significant).

From the last row of Table 19, in terms of overall condition scores, the C2, C4, and PMIS methods perform close enough to ground truth, C1 significantly overestimates overall condition scores, and C3 significantly underestimates them.

Table 20 shows the average differences in Alligator Cracking distress scores for 204 sections of asphalt pavement relative to ground truth (WALK method). Percent differences were not used for individual distress types since these distresses can have zero values. All automated methods either significantly underestimated (C1, C2, C3) or overestimated (C4) Alligator Cracking. The

PMIS method performed close enough to ground truth overall. All methods performed the worst for poor segments (Class F) having high percentages of Alligator Cracking.

Table 20. Differences in Alligator Cracking Scores Relative to Ground Truth for Asphalt.

Class	C1	C2	C3	C4	PMIS	N
A	0.81	0.60	0.52	10.55	0.75	145
B	-3.05	-1.83	-2.17	8.00	3.00	11
C	-0.64	-0.08	-0.14	19.47	18.13	6
D	-6.86	-10.22	-15.03	16.02	12.33	12
F	-39.11	-46.05	-50.61	-15.10	-26.63	30
All Classes	**-5.83**	**-7.12**	**-8.16**	**6.74**	-2.97	204

Table 21 shows the average differences for Block Cracking for 204 sections of asphalt pavement relative to ground truth (WALK method). Surprisingly, all other methods including PMIS were significantly different from ground truth. PMIS, C1 and C4 significantly underestimated Block Cracking. C2 and C3 significantly overestimated block cracking overall. Comparing the two groups, it appears that the underestimation group does better with better quality pavement (lower percentages of block cracking), while the overestimation group does better with poorer condition pavement (higher percentages of block cracking).

Table 21. Differences in Block Cracking Scores Relative to Ground Truth for Asphalt.

Class	C1	C2	C3	C4	PMIS	N
A	0.35	4.30	9.08	-0.07	-0.09	172
B	-6.33	6.78	18.44	-10.00	-10.00	3
D	-22.67	-3.57	23.20	-34.85	-35.44	10
F	-50.34	-1.10	-9.22	-63.52	-45.75	19
All Classes	**-5.65**	**3.44**	**8.20**	**-7.46**	**-7.43**	204

Table 22 shows the average differences in Failures distress scores for 204 sections of asphalt pavement relative to ground truth (WALK method). C1 significantly underestimates the number of failures, while C2 and C3 significantly overestimate them. C4 and PMIS methods perform close enough to ground truth. Since the overwhelming majority of segments are in very good condition with respect to failures (195 out of 204 segments), the test for overall differences is essentially judging methods based on their false positive rate for failures.

Table 22. Differences in Failure Scores Relative to Ground Truth for Asphalt.

Class	C1	C2	C3	C4	PMIS	N
A	-0.08	0.71	0.62	0.33	0.16	195
B	-2.00	-0.33	-1.33	-1.67	-0.83	3
F	-3.83	-1.50	-2.00	-3.67	-2.00	6
All Classes	**-0.22**	**0.63**	**0.51**	0.16	0.10	204

Table 23 shows the average differences in Longitudinal Cracking distress scores for 204 sections of asphalt pavement relative to ground truth (WALK method). Most of the methods overall are close enough, with only C1 significantly and consistently underestimating Longitudinal Cracking.

Table 23. Differences in Longitudinal Cracking Scores Relative to Ground Truth for Asphalt.

Class	C1	C2	C3	C4	PMIS	N
A	-18.09	17.36	22.95	43.33	20.40	144
B	-91.56	-23.97	-18.11	-27.92	20.94	22
C	-129.28	-53.93	-47.67	-64.72	-35.94	21
D	-181.78	-89.76	-128.33	-94.12	-81.43	15
F	-227.83	-119.83	-180.83	-139.00	-67.00	2
All Classes	**-51.88**	-3.86	-2.11	10.33	2.75	204

Table 24 shows the average differences in Patching distress scores for 204 sections of asphalt pavement relative to ground truth (WALK method). For this type of distress, only the PMIS method does not significantly underestimate the percent patching overall.

Table 24. Differences in Patching Scores Relative to Ground Truth for Asphalt.

Class	C1	C2	C3	C4	PMIS	N
A	0.20	0.13	0.17	0.10	0.62	184
B	-4.33	-3.50	-11.00	-11.00	N/A	2
C	-9.00	-13.08	-10.17	-13.00	4.00	4
D	-38.11	-45.08	-41.42	-40.90	-25.40	12
F	-100.00	-100.00	-100.00	-100.00	0.00	2
All Classes	**-3.30**	**-3.84**	**-3.61**	**-3.65**	-1.09	204

Table 25 shows the average differences in Transverse Cracking distress scores for 204 sections of asphalt pavement relative to ground truth (WALK method). For this type of distress, both C1 and C2 significantly underestimates the extent of Transverse Cracking; all other methods perform close enough to ground truth.

Table 25. Differences in Transverse Cracking Scores Relative to Ground Truth for Asphalt.

Class	C1	C2	C3	C4	PMIS	N
A	-0.54	-0.38	0.86	0.98	1.02	165
B	-5.61	-2.50	-0.35	-1.05	0.65	19
C	-10.00	-7.69	-7.36	-7.26	-5.92	14
D	-13.67	-10.33	-6.89	-7.88	-4.33	6
All Classes	**-2.06**	**-1.38**	-0.06	-0.18	0.15	204

CRCP Concrete Accuracy Results

Out of 26 total CRCP segments, only 9 segments had a ground truth (WALK) rating. Most of the automatic methods had three replicate ratings per segment. Therefore, when comparing accuracy of methods to ground truth, we must restrict the analysis to the 9 segments for which ground truth ratings exist. Table 26 lists these segments.

Table 26. Segments Used for CRCP Accuracy Comparison.

Section	County	Highway	Subsection	Direction	Lane
14	Williamson	SH 130 (Toll Road)	14.1	SB	R1
			14.2	SB	R1
			14.3	SB	R1
16	Williamson	US 183 Main Lanes	16.8	SB	R2
			16.9	SB	R2
47	Milam	US 77	47.4	SW	K1
			47.5	SW	K1
			47.6	SW	K1
			47.7	SW	K1

Table 27 shows the average percent difference in condition scores for 9 sections of CRCP concrete pavement relative to ground truth (WALK method). Note that overall no significant percentage differences were seen. This may be the consequence of having so few sections.

Table 27. Percent Differences in Condition Scores Relative to Ground Truth for CRCP.

Class	C1	C2	C3	C4	PMIS	N
A	-18.34	-25.05	-11.72	-18.86	0.00	5
D	25.75	51.22	22.76	97.56	-41.46	3
F	-69.44	-8.33	0.00	316.67	0.00	1
All Classes	-9.33	2.23	1.08	75.87	-10.37	9

Table 28 shows the overall average differences for five types of individual CRCP concrete distresses for 9 sections of pavement relative to ground truth (WALK method). Again, because of the small number of sections, the ratings from almost all the methods could not be judged significantly different from ground truth. The exception to this was the crack spacing rating for the C2 method, which significantly underestimated the spacing.

Table 28. Overall Differences for Individual Distress Scores Relative to Ground Truth for CRCP.

Distress Type	C1	C2	C3	C4	PMIS	N
AC Patch	0.00	0.11	0.11	0.50	0.00	9
Avg. Crack Space	-0.85	**-5.78**	-1.93	1.75	0.50	9
PCC Patch	1.78	0.11	0.00	0.00	0.00	9
Punchout	-0.33	0.00	0.11	-0.67	0.25	9
Spalls	0.15	0.67	0.00	-0.17	0.25	9

(Bolded Differences for All Classes are Statistically Significant)

JCP Concrete Accuracy Results

All 26 JCP segments had a ground truth (WALK) rating, although only one segment had a PMIS rating. Most of the automatic methods had three replicate ratings per segment. Table 29 lists the JCP segments.

Table 29. Segments Used for JCP Accuracy Comparison.

Section	County	Highway	Subsection	Direction	Lane
3	Travis	FM 969	3.1	EB	K1
			3.2	EB	K1
			3.3	EB	K1
			3.4	EB	K1
			3.5	EB	K1
			3.6	EB	K1
			3.7	EB	K1
			3.8	EB	K1
			3.9	EB	K1
4	Travis	FM 969	4.1	WB	K6
			4.2	WB	K6
5	Travis	State Loop 111	5.1	NB	K6
			5.2	NB	K6
6	Travis	State Loop 111	6.1	NB	K6
			6.2	NB	K6
			6.3	NB	K6
			6.4	NB	K6
17	Williamson	US 183 Frontage Roads	17.1	NB	X1
			17.2	NB	X1
			17.3	NB	X1
			17.6	SB	A1
			17.7	SB	A1
25	Brazos	SH 47	25.1	SE	R1
			25.2	SE	R1
			25.3	SE	R1
			25.4	SE	R1
			25.5	SE	R1
			25.7	NW	L1
			25.8	NW	L1
			25.9	NW	L1
			25.91	NW	L1
			25.92	NW	L1
			25.93	NW	L1
26	Brazos	FM 60	26.1	EB	K1
			26.2	EB	K1
			26.3	WB	K6
			26.4	WB	K6
44	Milam	US 79	44.1	SW	K1
			44.2	SW	K1
			44.3	SW	K1
			44.4	SW	K1
			44.5	SW	K1
			44.6	SW	K1
			44.7	SW	K1
			44.8	SW	K1
			44.9	SW	K1
47	Milam	US 77	47.2	SW	K1
			47.3	SW	K1

PMIS results are excluded in the following tables, since only one JCP segment had a PMIS rating. Table 30 shows the average percent difference in condition scores for 26 sections of JCP concrete pavement relative to ground truth (WALK method). Note that all methods significantly underestimated pavement condition. Note also that virtually all the sections are classified as very good (Class A). As with the CRCP results, the small sample size and top heavy distribution of good condition segments undermines the validity of any quantitative comparison.

Table 30. Percent Differences in Condition Scores Relative to Ground Truth for JCP.

Class	C1	C2	C3	C4	N
A	-28.65	-32.39	-10.98	-39.29	23
B	-57.14	-24.29	-71.43	-57.14	1
C	-84.02	-60.88	7.99	-16.58	2
All Classes	**-34.01**	**-34.27**	**-11.85**	**-38.23**	26

(Bolded Differences for All Classes are Statistically Significant)

Table 31 shows the overall average differences for six types of individual JCP concrete distresses for 26 sections of pavement relative to ground truth (WALK method). The results for JCP segments are surprisingly diverse, with some methods doing great with respect to some distress types while performing poorly at others. It would appear that JCP distress types pose challenges for automated ratings methods; then again, these results could be an artefact of the small number of segments available for evaluation.

Table 31. Overall Differences for Individual Distress Scores Relative to Ground Truth for JCP.

Distress Type	C1	C2	C3	C4	N
Fails	0.19	**-1.12**	**-1.50**	**-1.27**	26
FJC	-0.23	0.71	0.87	0.19	26
Avg. Jt Space	**-2.18**	**60.00**	**-1.05**	0.35	26
PCC Patch	-0.04	0.12	0.00	0.00	26
Shat Slab	-0.17	0.04	0.12	**2.77**	26
Slab Long	-0.08	-0.35	**4.23**	**9.08**	26

CHAPTER 4. EVALUATION OF AUTOMATED RUT MEASUREMENTS FOR NETWORK LEVEL COLLECTION OF PAVEMENT RUT DEPTHS

Automated pavement condition survey systems evaluated in this research project also provided rut measurements computed from transverse profiles collected using scanning lasers mounted on the test vehicles. These sensors are typically mounted on the rear of the vehicle as illustrated in Figure 1 and Figure 2. To evaluate the rut depths determined from automated pavement condition survey vehicles, researchers set up test sections on in-service pavements where test vehicles (such as those illustrated in Figure 1 and Figure 2) collected rut measurements. Participating service providers processed the data collected from their systems and provided rut depths computed from the scans made on specific routes surveyed during this evaluation. Researchers then assessed the repeatability and accuracy of rut depth statistics determined from these measurements. This chapter presents the findings from this evaluation.

TEST ROUTES FOR RUT EVALUATION

Researchers used TxDOT's Pavement Management Information System (PMIS) database to identify candidate routes to evaluate rut depths determined from automated measurement systems. Using TxDOT's ArcMap utility, researchers identified candidate routes that exhibited various levels of rutting as reported in the PMIS database. Figure 3 illustrates a map of candidate routes from a query made on Bryan District data. The color-coding in the map identifies different levels of rutting, with green showing segments with no rutting and red identifying segments with severe rutting. From this database search, researchers decided to use the following routes in Burleson County for evaluating rut measurements collected with automated survey vehicles:

1. FM 166 from the SH 36 junction to FM 50.
2. FM 908 from SH 21 to the Milam County Line.

Table 32 shows the specific segments on these routes where providers of automated pavement condition test services collected rutting data. Each participating service provider made three runs and submitted rutting data on each segment. Test runs were made in both directions of the highway.

Figure 1. Scanning Lasers Mounted on Each Side of D-Vision Test Vehicle.

Figure 2. Scanning Lasers Mounted on Top Left and Top Right of ARAN Unit.

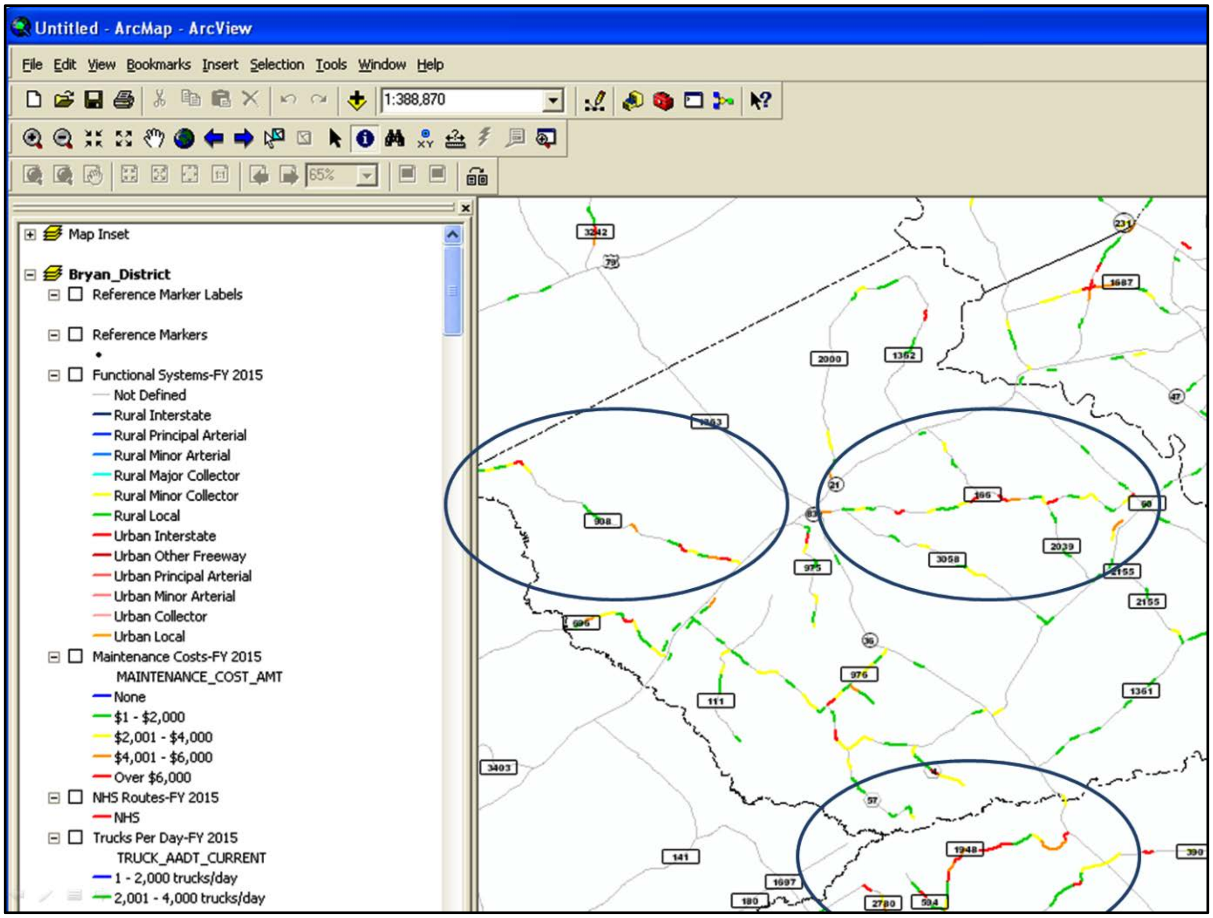

Figure 3. Illustration of PMIS Query Results on Bryan District Data.

Table 32. Test Segments for Evaluating Automated Rut Measurement Systems.

Route	Texas Reference Marker (TRM)		Beginning GPS Coordinates (degrees)		Ending GPS Coordinates (degrees)		Comment
	Begin	End	Latitude	Longitude	Latitude	Longitude	
FM 166	602	614	30.53618	-96.65568	30.53849	-96.47828	TRM 602 is about 2 miles east of the junction of FM 166 and SH 36 in Caldwell
FM 908	590	594	30.56684	-96.94633	30.55176	-96.89301	TRM 590 is at the Burleson/Milam County Line
FM 908	600	604	30.51535	-96.80962	30.49685	-96.74789	TRM 604 is near the Shell gas station at the junction of FM 908 and SH 21.

Given that data reported in the PMIS are based on the time TxDOT collected the measurements, researchers contracted with a service provider to get up-to-date rutting data on FM 166 and FM

908. These measurements were collected with a 7-point laser system, with wheel path rut depths reported at 1-ft intervals in lieu of the 0.1-mile interval used to report average rut depths in the PMIS. Researchers processed the data from these measurements to determine the percentages of the following levels of rutting over a 528-ft segment length in accordance with TxDOT practice:

1. Rut depths less than 0.25-inch.
2. Shallow rutting greater than or equal to 0.25-inch and less than 0.5-inch.
3. Deep rutting greater than or equal to 0.5-inch and less than 1.0-inch.
4. Severe rutting greater than or equal to 1.0-inch and less than 2.0-inch.
5. Failure rutting greater than or equal to 2.0-inch.

In lieu of reporting the percentages at 528-ft intervals, researchers determined the above levels of rutting over a 528-ft continuous interval to assess the rutting along each route in more detail. In this continuous interval analysis, the percentages of different levels of rutting are first determined for the beginning 528-ft section. Then, the process goes 1 sample interval forward and computes similar percentages over the next 528-ft interval. Since the rut depths were determined at 1-ft intervals, the analysis proceeded in 1-ft steps until the last 528-ft section along the test lane was processed.

Figure 4 and Figure 5 illustrate the results from the continuous interval analysis using the rutting data collected along both travel lanes of FM 166. Researchers used this information to identify test sections where reference rut measurements were collected to assess the accuracy of the computed rut depths from automated rut measurement systems. For this purpose, researchers established ten 550-ft reference test sections along FM 166. Table 33 identifies these test sections where researchers collected reference measurements with the straightedge in accordance with ASTM E 1703. The following sections present the evaluation of test data from automated rut measurement systems.

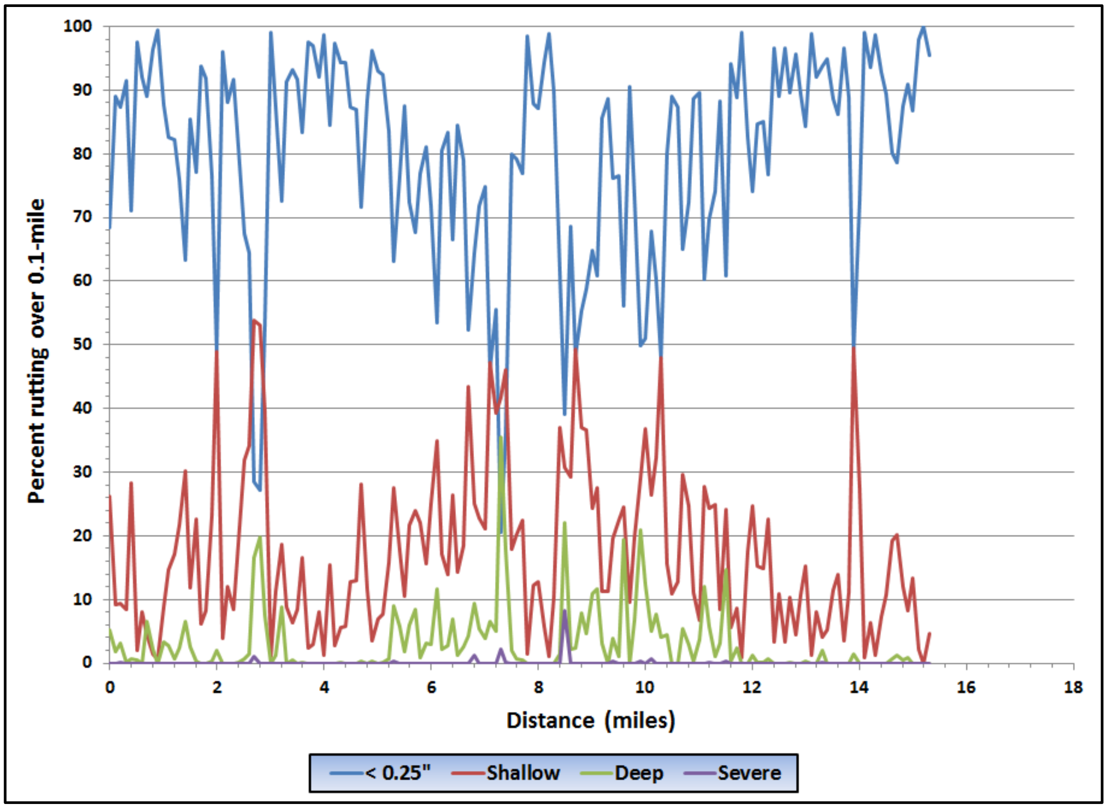

Figure 4. Percentages of Different Levels of Rutting on Eastbound Lane of FM 166.

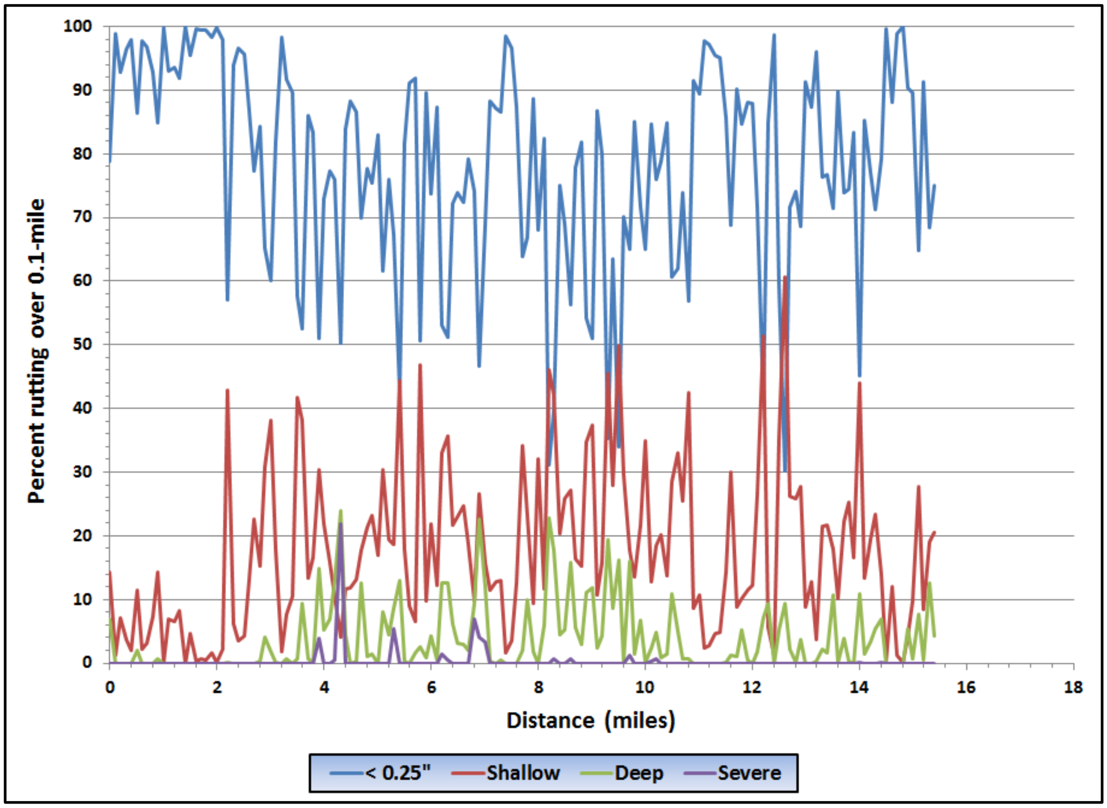

Figure 5. Percentages of Different Levels of Rutting on Westbound Lane of FM 166.

Table 33. FM 166 Reference Test Sections.

Test Lane	Section ID	Latitude at start	Longitude at start	Comment
Westbound	WB1	30.544779	-96.510839	
Westbound	WB2	30.541942	-96.516465	Starts at sign for FM 2039 junction.
Westbound	WB3	30.543077	-96.552287	Starts at sign for Brazos Way on eastbound shoulder
Westbound	WB4	30.543946	-96.553725	Contiguous with WB3
Westbound	WB5	30.544869	-96.555083	Contiguous with WB4
Westbound	WB6	30.547693	-96.566996	Starts where edge stripe begins
Westbound	WB7	30.533265	-96.631806	Starts at sign for CR 233 on eastbound shoulder
Eastbound	EB1	30.543807	-96.574008	Starts at sign for FM 1362 junction
Eastbound	EB2	30.545267	-96.555603	Starts beside a driveway
Eastbound	EB3	30.544383	-96.533455	Starts beside mailboxes for house # 10540

REPEATABILITY OF TEST STATISTICS FROM AUTOMATED RUT MEASUREMENTS

The service providers who participated in this study collected rut measurements over the test segments identified in Table 32. For each specified segment, three repeat runs were made on each test lane for the required distance. Table 32 shows that the segments are 4 and 12 miles in length. Since this project is concerned with investigating the application of automated distress measurement systems to assess pavement condition over the state maintained road network, researchers evaluated measurement repeatability based on rutting statistics that TxDOT uses to manage the Texas road network. Researchers determined these statistics using the rut depths computed from automated distress measurements by the participating service providers. The test statistics include the shallow and deep levels of rutting identified previously, as well as the average rut depths computed over a 528-ft base length. Researchers then compared corresponding test statistics using a continuous 528-ft interval analysis as described earlier in this chapter.

Two participating service providers submitted detailed rutting data on the FM 166 and FM 908 test segments with rut depths reported between 0.3- and 1-ft intervals. Researchers used the data from these service providers to evaluate the repeatability of rut measurements as presented herein. For the purpose of reporting this evaluation the test results are identified generically as Vendor I and Vendor II.

Figure 6 to Figure 10 illustrate the repeatability of various rut depth statistics computed from the rutting data submitted by Vendor I, while Figure 11 to Figure 15 show the corresponding charts based on data submitted by Vendor II. These figures plot the following rut depth statistics computed from the 528-ft continuous interval analysis of data collected along the northbound (NB) lane of FM 908 from TRM 604 to TRM 600:

1. Average left wheel path (LWP) rut depth.
2. Average right wheel path (RWP) rut depth.
3. Average rut depth over the given 528-ft section.
4. Percentage of shallow rutting.
5. Percentage of deep rutting.

TxDOT uses the last two statistics to determine the distress score for a given PMIS segment based on utility values computed for different distress types that include rutting for flexible pavements. Specifically, the distress score (DS) is computed using the following equation:

$$DS = 100 \prod_{i=1}^{n} U_i \tag{1}$$

where,

U_i = utility value.

i = a PMIS distress type.

n = total number of distress types to compute DS for the given pavement type.

The utility value ranges from 0 to 1 with the value diminishing as the level of distress increases. The following general equation is used to compute the utility value:

$$U_i = 1 - \alpha e^{-\left[\left(\frac{\rho}{l_i}\right)^{\beta}\right]} \tag{2}$$

l_i = level of distress type *i*.

e = base of the natural logarithm.

α = a horizontal asymptote coefficient that controls the maximum amount of utility loss.

β = a slope coefficient that controls how steeply the utility is lost in the middle of the curve.

ρ = a prolongation coefficient factor that controls how long the utility curve will last above a certain value.

The coefficients α, β, and ρ, vary depending on distress type and pavement type. Note that the utility value from equation (2) approaches 1 as the level of distress approaches zero. In current practice, the utility value is set to 1 when the distress level is reported as zero.

To quantify the repeatability of the rutting statistics shown in Figure 6 to Figure 15, researchers determined the point-to-point variance of corresponding 528-ft continuous rutting statistics computed using the data from repeat runs. The square root of the average variance was then calculated and normalized with respect to the overall mean value of the given rutting statistic. Researchers refer to this measure of repeatability as the normalized root-mean-square error ($RMSE_{norm}$) in Figure 6 to Figure 15. Lower values of $RMSE_{norm}$ indicate better repeatability of the test data.

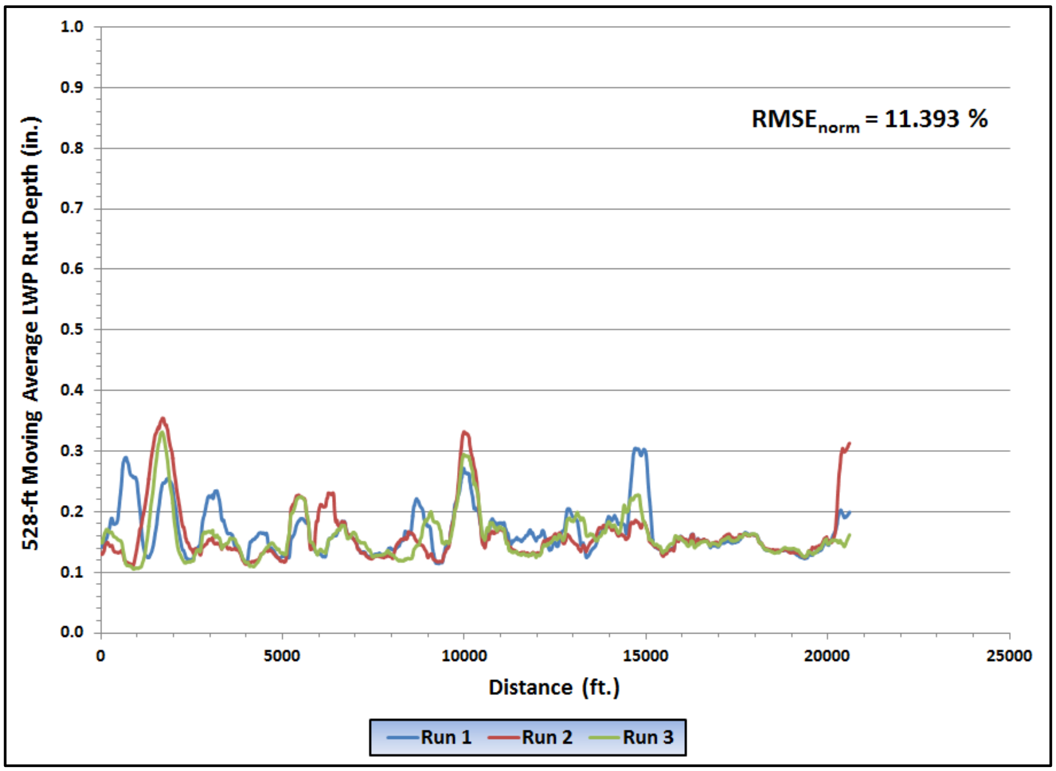

Figure 6. Continuous 528-ft Average LWP Rut Depths from Repeat Runs of Vendor I on FM 908 NB Test Segment (TRM 604 to TRM 600).

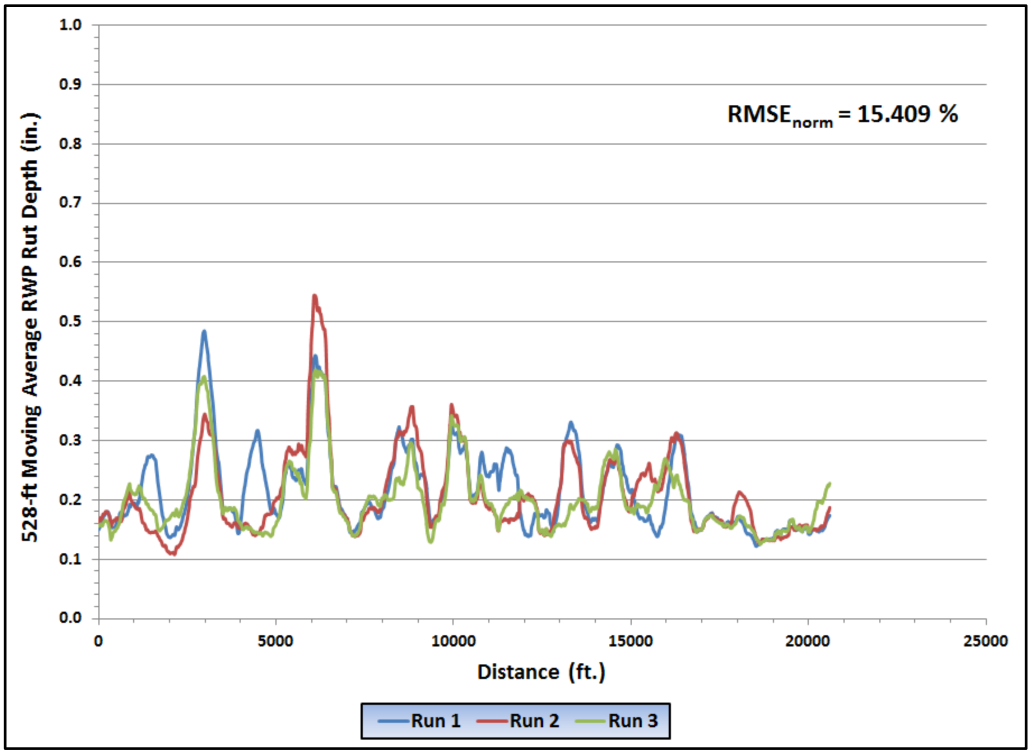

Figure 7. Continuous 528-ft Average RWP Rut Depths from Repeat Runs of Vendor I on FM 908 NB Test Segment (TRM 604 to TRM 600).

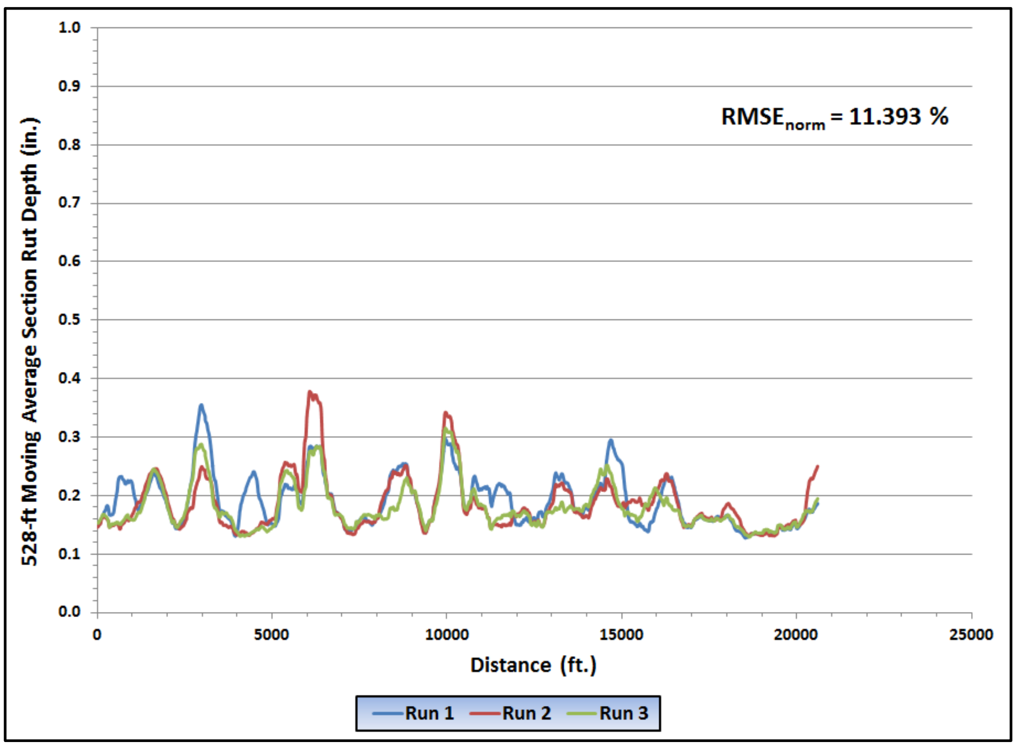

Figure 8. Continuous 528-ft Average Section Rut Depths from Repeat Runs of Vendor I on FM 908 NB Test Segment (TRM 604 to TRM 600).

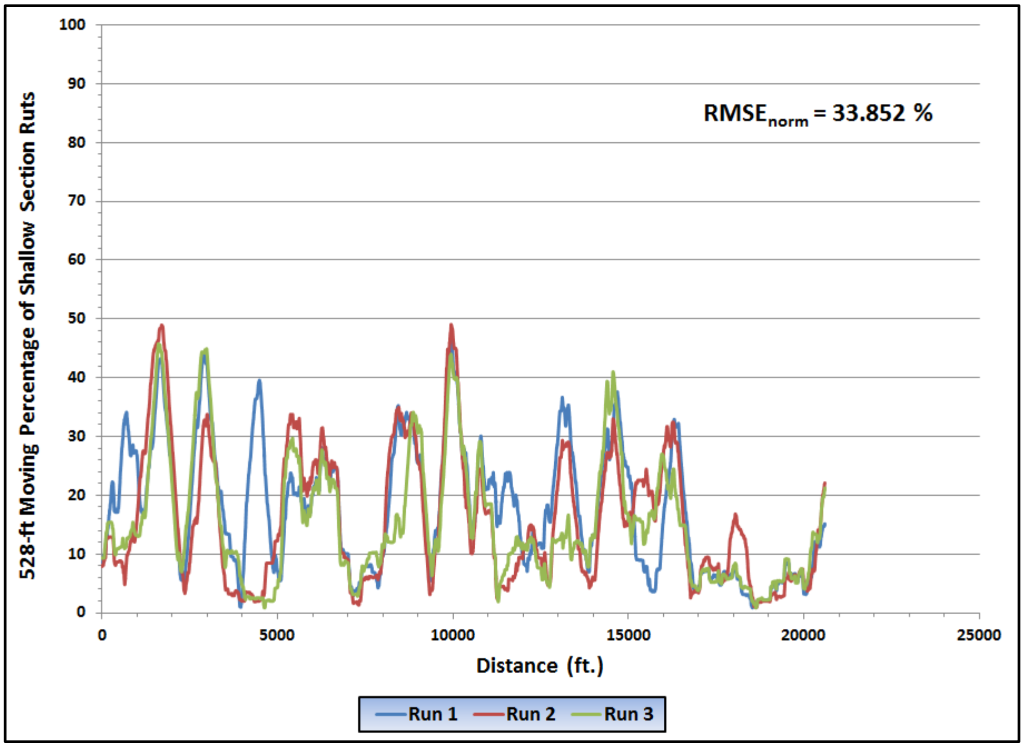

Figure 9. Continuous 528-ft Percentages of Shallow Rutting from Repeat Runs of Vendor I on FM 908 NB Test Segment (TRM 604 to TRM 600).

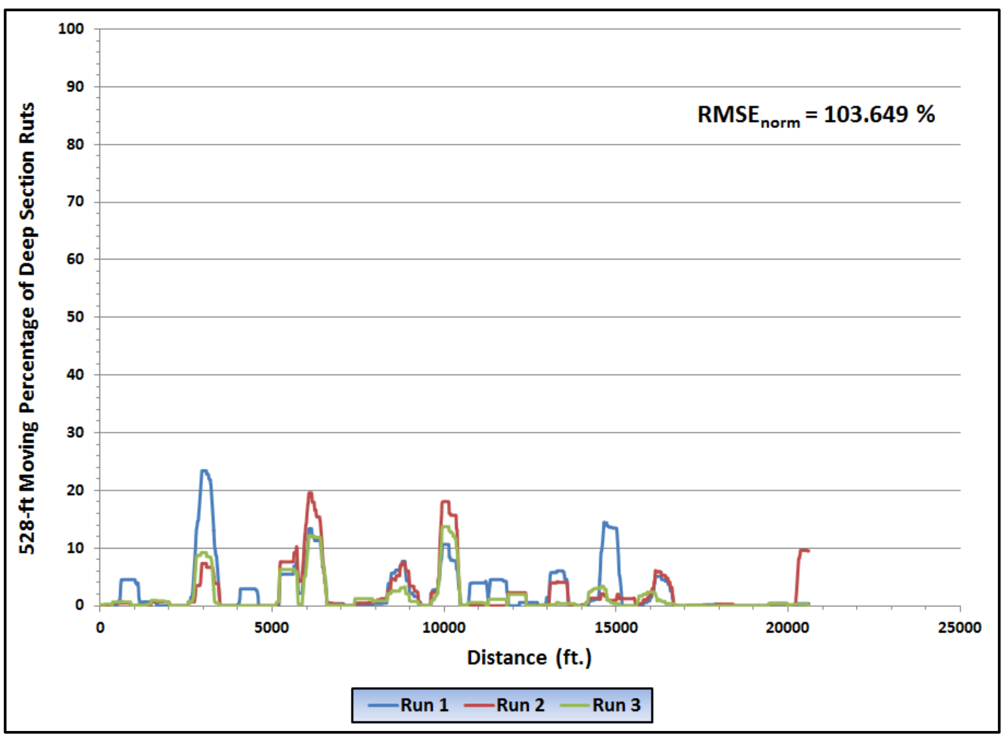

Figure 10. Continuous 528-ft Percentages of Deep Rutting from Repeat Runs of Vendor I on FM 908 NB Test Segment (TRM 604 to TRM 600).

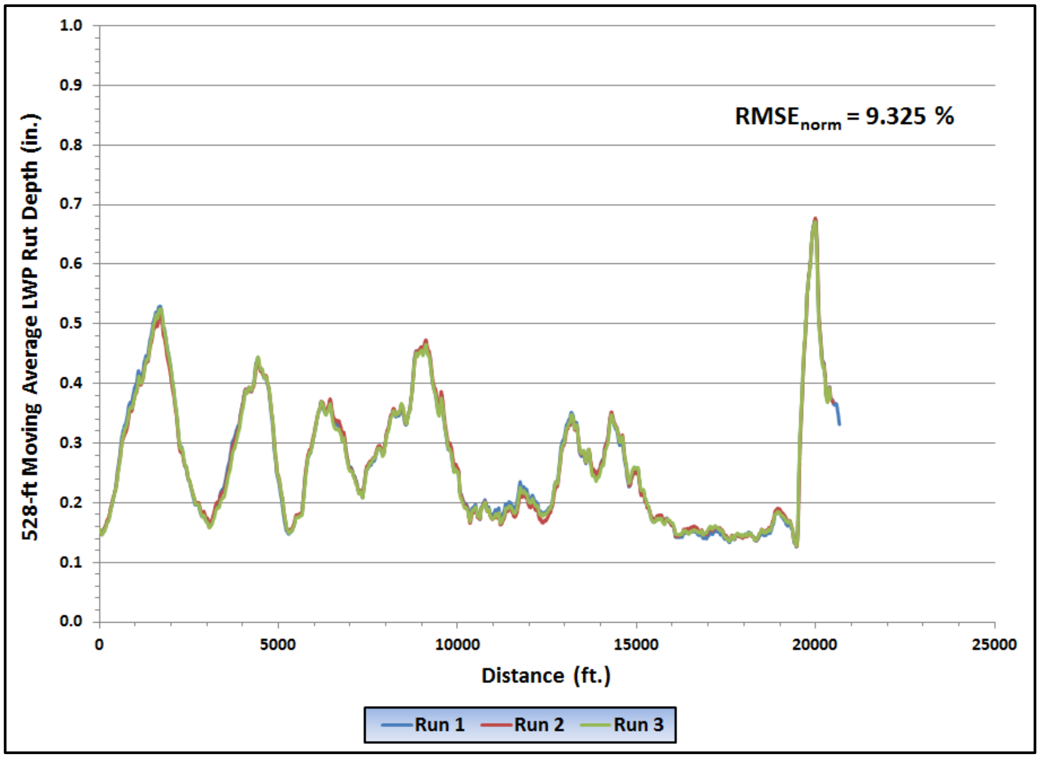

Figure 11. Continuous 528-ft Average LWP Rut Depths from Repeat Runs of Vendor II on FM 908 NB Test Segment (TRM 604 to TRM 600).

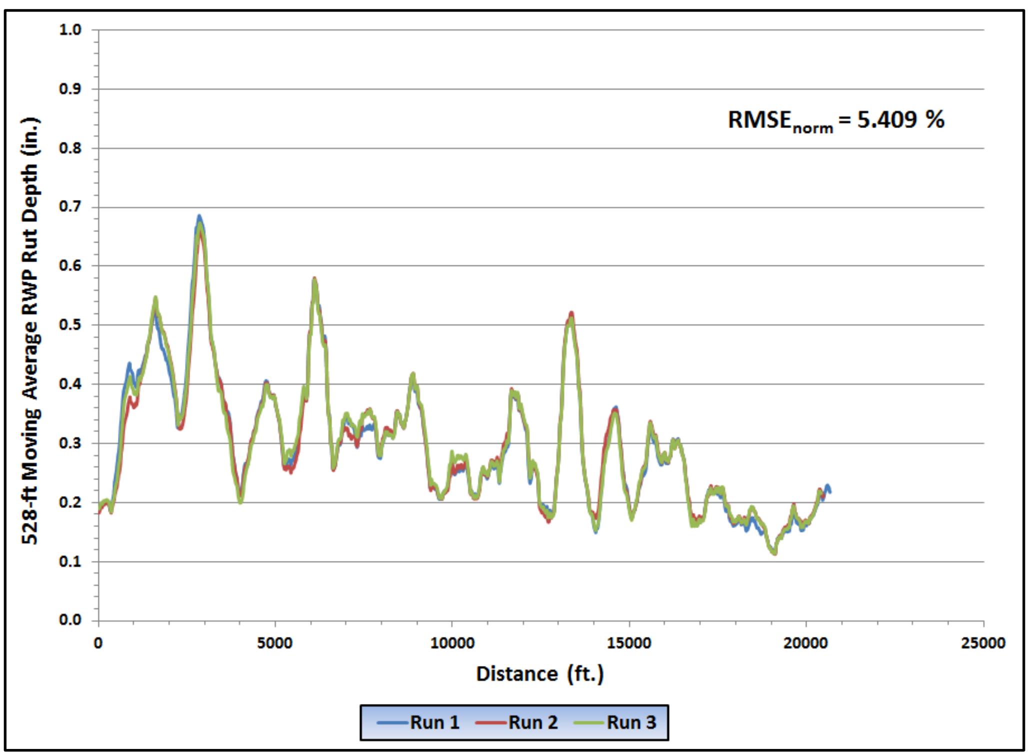

Figure 12. Continuous 528-ft Average RWP Rut Depths from Repeat Runs of Vendor II on FM 908 NB Test Segment (TRM 604 to TRM 600).

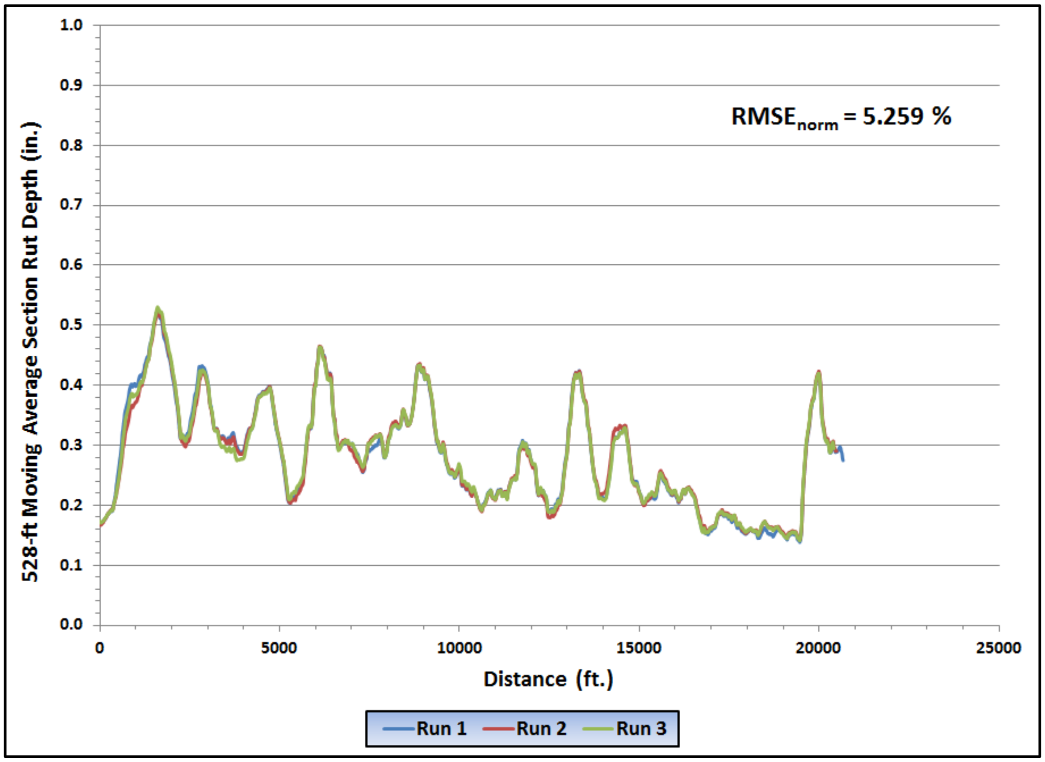

Figure 13. Continuous 528-ft Average Section Rut Depths from Repeat Runs of Vendor II on FM 908 NB Test Segment (TRM 604 to TRM 600).

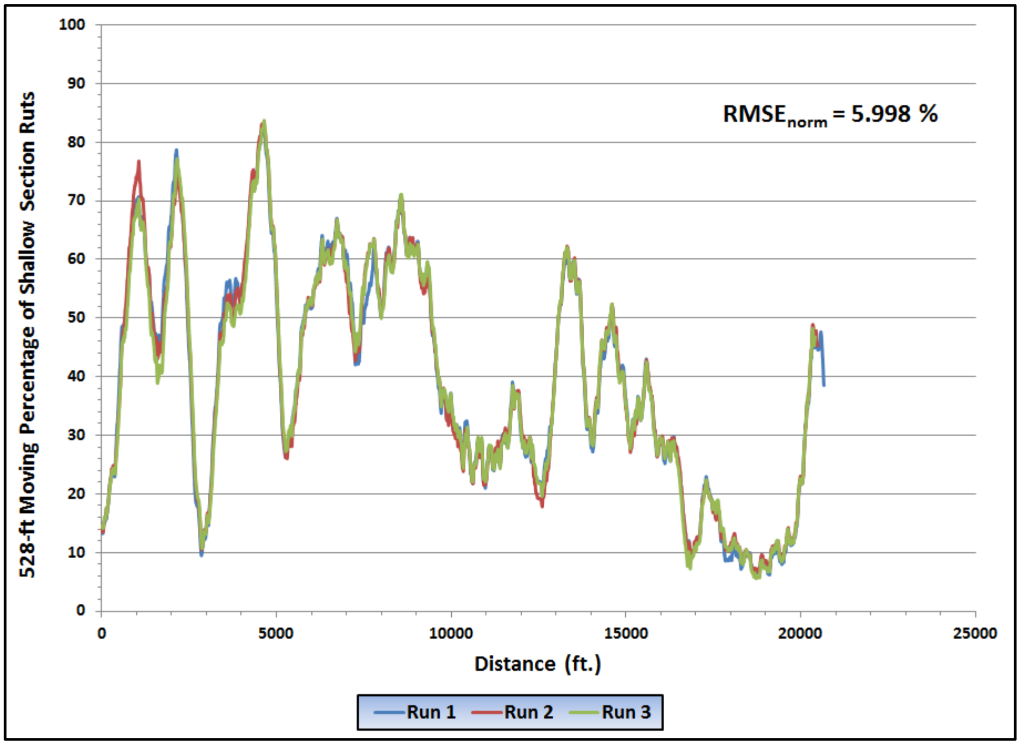

Figure 14. Continuous 528-ft Percentages of Shallow Rutting from Repeat Runs of Vendor II on FM 908 NB Test Segment (TRM 604 to TRM 600).

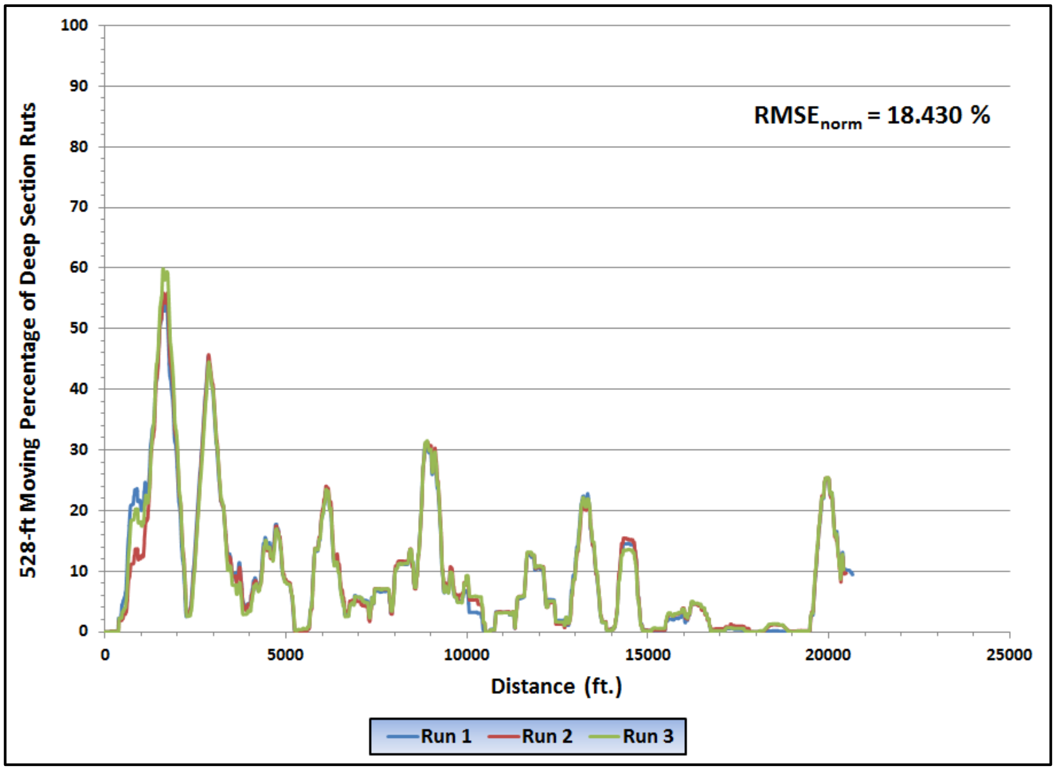

Figure 15. Continuous 528-ft Percentages of Deep Rutting from Repeat Runs of Vendor II on FM 908 NB Test Segment (TRM 604 to TRM 600).

Comparing Figure 6 to Figure 10 with Figure 11 to Figure 15 shows a noticeable difference in the repeatability of the computed rut depths between Vendor I and Vendor II. Specifically, the test data from Vendor II shows better repeatability based on the rutting statistics used in this evaluation. A similar observation was made using the data collected on the other FM 908 test segments. This finding is apparent from Table 34, which summarizes the normalized root-mean-square error statistics determined from the rutting data collected along FM 908.

ACCURACY OF AUTOMATED RUT MEASUREMENTS

To evaluate the accuracy of automated rut measurements, researchers also conducted a continuous interval analysis where rutting statistics computed from the test data were compared with corresponding statistics determined from reference straightedge measurements of wheel path rutting collected in accordance with ASTM E 1703 by TxDOT certified PMIS raters. On each reference test section identified in Table 33, the raters collected rut measurements at 10-ft intervals beginning at station 0+00 to station 5+50.

Table 34. Summary Results from Comparison of Repeatability of Automated Rut Measurements on FM 908 Test Segments.

Test Lane	TRM Limits		Rutting Statistic	Normalized Root-Mean-Square Error Statistic (%)	
	Start	End		Vendor I	Vendor II
Northbound	604	600	Mean LWP rut depth	11.393	9.325
			Mean RWP rut depth	15.409	5.409
			Mean section rut depth	11.393	5.259
			Percent shallow rutting	33.852	5.998
			Percent deep rutting	103.649	18.430
Southbound	600	604	Mean LWP rut depth	10.289	8.616
			Mean RWP rut depth	16.794	9.430
			Mean section rut depth	11.588	7.167
			Percent shallow rutting	23.634	9.184
			Percent deep rutting	77.335	25.158
Northbound	594	590	Mean LWP rut depth	12.856	9.586
			Mean RWP rut depth	13.366	15.331
			Mean section rut depth	8.717	11.199
			Percent shallow rutting	14.279	13.106
			Percent deep rutting	69.487	29.393
Southbound	590	594	Mean LWP rut depth	12.139	5.185
			Mean RWP rut depth	11.197	6.583
			Mean section rut depth	6.911	5.003
			Percent shallow rutting	19.055	6.155
			Percent deep rutting	59.607	18.018

A total of 56 rut measurements with the straightedge were made on each wheel path of each reference test section. From these measurements, researchers computed the same rutting statistics used to evaluate the repeatability of test data presented previously, except that the statistics were determined over the 550-ft section length.

Since test runs were made in a manner that more closely resemble how data are collected on automated network level visual distress surveys, the service providers did not run each reference test section individually. Instead, researchers instructed the service providers to collect automated rut measurements on both lanes of FM 166 and submit test data over the 12-mile distance interval specified in Table 33 for this route. Indeed, the locations of the reference test sections were not known to the service providers.

Thus, to evaluate the accuracy of the test data from Vendor I and Vendor II, researchers first had to locate each reference test section in the data submitted by these vendors on each FM 166 travel lane. This step was made using the reference section GPS coordinates given in Table 33 along with the GPS coordinates included in the records of the data files the vendors submitted. It is noted that GPS coordinates were included in the rut data file format researchers provided and discussed at length with the vendors who participated in this study.

After finding the starting location of a reference section in the test data, researchers extracted the data records over the 550-ft length of the reference section, plus the records over the contiguous 20-ft interval upstream of the section, and the contiguous 20-ft interval downstream of the section. Thus, researchers extracted test data file over a 590-ft interval that bracketed the reference section based on GPS coordinates. Researchers included 20-ft of additional data before and after the section to account for possible GPS location errors.

Once test data corresponding to the reference section were extracted, researchers performed a continuous interval analysis over a base length of 550 ft. Researchers then compared the rutting statistics computed from the test data over each 550-ft interval with the corresponding statistics determined from the reference rut measurements to assess the level of agreement (or lack thereof) between the test and reference data. Table 35 and Table 36 illustrate the results from this analysis using rut depth data collected on section WB1 from Vendor I and Vendor II, respectively. The first column in each table identifies the rutting statistics determined from the test and reference data. The next 5 columns show the minimum, maximum, average, standard deviation, and coefficient of variation (CV) of the rutting statistic for each 550-ft segment included in the continuous interval analysis of test data that bracket the reference section. These values can be compared with the corresponding reference value shown in the last column to assess the level of agreement or disagreement of the test data relative to the reference.

Note that the utility factors for shallow rutting and deep rutting were also included in assessing the accuracy of the test measurements relative to the reference. Given these utility values, researchers determined the percent change in the distress score due to differences between test and reference rut depths. Specifically, the percent change in distress score is defined as follows:

$$\% \Delta DS = 100 \times \left(\frac{DS^T - DS^R}{DS^R} \right) \quad (3)$$

where,

$\%\Delta DS$ = percent change in distress score.

DS^T = distress score corresponding to the rut depths determined the test vehicle.

DS^R = distress score computed from the reference rut measurements.

Table 35. Summary Results from Evaluation of Accuracy of Automated Rut Measurements from Vendor I against Reference Values on FM 166 Section WB1.

Rutting Statistic	Minimum	Maximum	Average	Std. dev.	CV %	Reference data
Mean LWP rut depth (in.)	0.23	0.23	0.23	0.0017	0.72	0.30
Mean RWP rut depth (in.)	0.38	0.41	0.39	0.01	2.44	0.50
Mean section rut depth (in.)	0.30	0.32	0.31	0.01	1.80	0.40
Percent shallow rutting	35.56	37.37	36.38	0.49	1.35	45.54
Percent deep rutting	12.37	15.25	14.05	0.99	7.08	22.32
$U_{shallow}$	0.8171	0.8220	0.8197	0.0013	0.16	0.7990
U_{deep}	0.7625	0.8149	0.7838	0.0179	2.28	0.6671
$U_{shallow} \times U_{deep}$	0.6231	0.6698	0.6425	0.0157	2.44	0.5330
% change in DS	16.90	25.67	20.55	2.94	14.30	

Table 36. Summary Results from Evaluation of Accuracy of Automated Rut Measurements from Vendor II against Reference Values on FM 166 Section WB1.

Rutting Statistic	Minimum	Maximum	Average	Std. dev.	CV %	Reference data
Mean LWP rut depth (in.)	0.29	0.30	0.29	0.0036	1.23	0.30
Mean RWP rut depth (in.)	0.54	0.55	0.54	0.00	0.72	0.50
Mean section rut depth (in.)	0.41	0.43	0.42	0.00	0.87	0.40
Percent shallow rutting	40.21	42.45	41.60	0.80	1.92	45.54
Percent deep rutting	21.53	23.00	21.86	0.41	1.89	22.32
$U_{shallow}$	0.8052	0.8102	0.8071	0.0018	0.2189	0.7990
U_{deep}	0.6599	0.6759	0.6723	0.0046	0.6784	0.6671
$U_{shallow} \times U_{deep}$	0.5314	0.5469	0.5426	0.0042	0.7763	0.5330
% change in DS	-0.29	2.62	1.80	0.79	44.02	

The distress scores corresponding to the test and reference data may be computed from equation (1). Assuming that the utility factors for other distress types are constant, DS^T and DS^R may be computed as follows:

$$DS^T = 100\,K\,(U^T_{shallow} \times U^T_{deep}) \tag{4}$$

$$DS^R = 100\,K\,(U^R_{shallow} \times U^R_{deep}) \tag{5}$$

where,

K = a constant representing the product of the utility values for the other distress types.

$U^{T}_{shallow}$ = utility value for shallow rutting based on rut depths determined from the test vehicle.

U^{T}_{deep} = utility value for deep rutting based on rut depths determined from the test vehicle.

$U^{R}_{shallow}$ = utility value for shallow rutting based on reference rut measurements.

U^{R}_{deep} = utility value for deep rutting based on reference rut measurements.

Substituting equations (4) and (5) into equation (3), the percent change in the distress score due to inaccuracies in test measurements relative to the reference values is determined as follows:

$$\% \Delta DS = 100 \left(\frac{U^{T}_{shallow} \times U^{T}_{deep}}{U^{R}_{shallow} \times U^{R}_{deep}} - 1 \right) \tag{6}$$

Researchers used equation (6) to compute the percent change in distress score on each 550-ft segment included in the continuous interval analysis of test data. In this regard, Table 35 and Table 36 also include the range of $\%\Delta DS$ along with the average, standard deviation, and the coefficient of variation of the percent change in DS. Given its significance in determining maintenance and rehabilitation strategies based on pavement condition measurements, researchers selected this parameter as the overall indicator of the level of agreement between the rut depth test data and the corresponding reference values. In this regard, $\%\Delta DS = 0$ indicates that the same utility values for shallow and deep rutting were determined from the test and reference data. The percent change in DS can be positive or negative depending on whether the rut measurements underestimate or overestimate, respectively, the corresponding reference values. The closer this parameter is to zero, the better the agreement between the test and reference measurements. Figure 16 compares the $\%\Delta DS$ between Vendor I and Vendor II. For the majority of the reference sections, the percent change in DS is closer to zero for Vendor II indicating better accuracy of the rut measurements from this vendor based on the reference data collected on the same sections. Table 37 to Table 54 compare the rutting statistics determined from automated rut measurements with the corresponding values determined from the reference data on the other nine sections included in this evaluation.

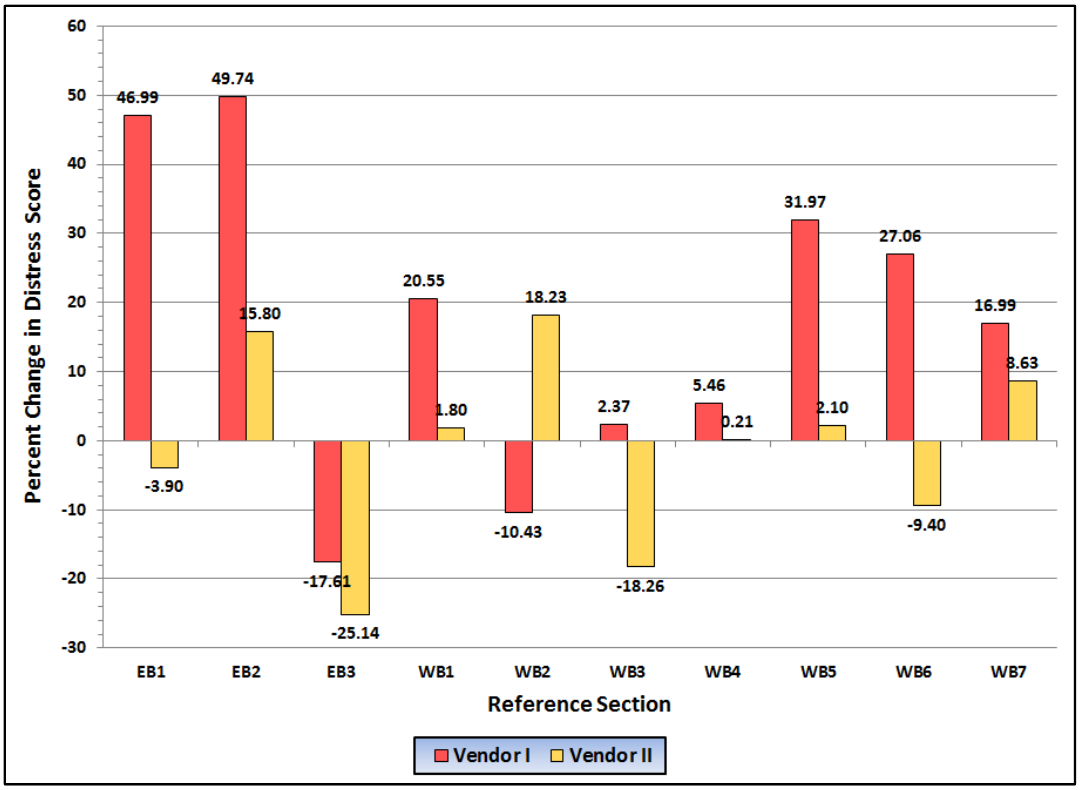

Figure 16. Comparison of the Percent Change in Distress Score Calculated from Vendor I and Vendor II Rutting Data.

Table 37. Summary Results from Evaluation of Accuracy of Automated Rut Measurements from Vendor I against Reference Values on FM 166 Section WB2.

Rutting Statistic	Minimum	Maximum	Average	Std. dev.	CV %	Reference data
Mean LWP rut depth (in.)	0.13	0.13	0.13	0.0004	0.30	0.10
Mean RWP rut depth (in.)	0.65	0.67	0.66	0.01	1.27	0.89
Mean section rut depth (in.)	0.39	0.40	0.39	0.00	1.06	0.49
Percent shallow rutting	11.91	13.63	13.01	0.60	4.59	4.46
Percent deep rutting	29.51	32.67	31.21	1.01	3.23	27.68
$U_{shallow}$	0.9271	0.9408	0.9320	0.0048	0.5109	0.9963
U_{deep}	0.5807	0.6024	0.5904	0.0069	1.1733	0.6167
$U_{shallow} \times U_{deep}$	0.5463	0.5589	0.5503	0.0040	0.7327	0.6144
% change in DS	-11.09	-9.03	-10.43	0.66	-6.29	

Table 38. Summary Results from Evaluation of Accuracy of Automated Rut Measurements from Vendor II against Reference Values on FM 166 Section WB2.

Rutting Statistic	Minimum	Maximum	Average	Std. dev.	CV %	Reference data
Mean LWP rut depth (in.)	0.11	0.12	0.12	0.0005	0.44	0.10
Mean RWP rut depth (in.)	1.07	1.09	1.09	0.01	0.76	0.89
Mean section rut depth (in.)	0.59	0.61	0.60	0.00	0.67	0.49
Percent shallow rutting	4.22	4.89	4.48	0.29	6.48	4.46
Percent deep rutting	16.88	17.88	17.40	0.36	2.09	27.68
$U_{shallow}$	0.9945	0.9971	0.9961	0.0011	0.1125	0.9963
U_{deep}	0.7223	0.7369	0.7292	0.0053	0.7290	0.6167
$U_{shallow} \times U_{deep}$	0.7202	0.7329	0.7264	0.0045	0.6178	0.6144
% change in DS	17.23	19.29	18.23	0.73	4.01	

Table 39. Summary Results from Evaluation of Accuracy of Automated Rut Measurements from Vendor I against Reference Values on FM 166 Section WB3.

Rutting Statistic	Minimum	Maximum	Average	Std. dev.	CV %	Reference data
Mean LWP rut depth (in.)	0.13	0.14	0.13	0.0016	1.22	0.08
Mean RWP rut depth (in.)	0.29	0.33	0.32	0.01	4.45	0.38
Mean section rut depth (in.)	0.21	0.23	0.23	0.01	2.77	0.23
Percent shallow rutting	18.68	20.58	19.36	0.55	2.83	31.25
Percent deep rutting	4.69	7.40	5.87	0.97	16.51	3.57
$U_{shallow}$	0.8811	0.8921	0.8881	0.0032	0.3606	0.8351
U_{deep}	0.9234	0.9785	0.9556	0.0197	2.0585	0.9927
$U_{shallow} \times U_{deep}$	0.8238	0.8648	0.8487	0.0147	1.7315	0.8290
% change in DS	-0.63	4.32	2.37	1.77	74.83	

Table 40. Summary Results from Evaluation of Accuracy of Automated Rut Measurements from Vendor II against Reference Values on FM 166 Section WB3.

Rutting Statistic	Minimum	Maximum	Average	Std. dev.	CV %	Reference data
Mean LWP rut depth (in.)	0.16	0.16	0.16	0.0007	0.43	0.08
Mean RWP rut depth (in.)	0.54	0.67	0.60	0.04	6.56	0.38
Mean section rut depth (in.)	0.34	0.40	0.37	0.02	5.15	0.23
Percent shallow rutting	18.69	22.82	20.10	1.27	6.31	31.25
Percent deep rutting	14.67	15.15	15.01	0.15	1.02	3.57
$U_{shallow}$	0.8694	0.8921	0.8840	0.0071	0.8009	0.8351
U_{deep}	0.7643	0.7725	0.7666	0.0026	0.3394	0.9927
$U_{shallow} \times U_{deep}$	0.6656	0.6891	0.6777	0.0073	1.0710	0.8290
% change in DS	-19.71	-16.88	-18.26	0.88	-4.80	

Table 41. Summary Results from Evaluation of Accuracy of Automated Rut Measurements from Vendor I against Reference Values on FM 166 Section WB4.

Rutting Statistic	Minimum	Maximum	Average	Std. dev.	CV %	Reference data
Mean LWP rut depth (in.)	0.13	0.13	0.13	0.0002	0.13	0.09
Mean RWP rut depth (in.)	0.37	0.39	0.37	0.01	1.59	0.47
Mean section rut depth (in.)	0.25	0.26	0.25	0.00	1.15	0.28
Percent shallow rutting	24.73	24.91	24.79	0.08	0.33	20.54
Percent deep rutting	10.47	11.19	11.04	0.19	1.72	14.29
$U_{shallow}$	0.8595	0.8604	0.8601	0.0004	0.0432	0.8813
U_{deep}	0.8388	0.8541	0.8419	0.0040	0.4787	0.7791
$U_{shallow} \times U_{deep}$	0.7216	0.7349	0.7241	0.0035	0.4771	0.6866
% change in DS	5.10	7.03	5.46	0.50	9.21	

Table 42. Summary Results from Evaluation of Accuracy of Automated Rut Measurements from Vendor II against Reference Values on FM 166 Section WB4.

Rutting Statistic	Minimum	Maximum	Average	Std. dev.	CV %	Reference data
Mean LWP rut depth (in.)	0.16	0.16	0.16	0.0008	0.48	0.09
Mean RWP rut depth (in.)	0.70	0.77	0.74	0.02	3.19	0.47
Mean section rut depth (in.)	0.42	0.45	0.44	0.01	2.66	0.28
Percent shallow rutting	17.81	18.04	17.96	0.07	0.37	20.54
Percent deep rutting	13.51	15.85	14.98	0.78	5.20	14.29
$U_{shallow}$	0.8961	0.8976	0.8966	0.0004	0.0469	0.8813
U_{deep}	0.7528	0.7931	0.7674	0.0132	1.7259	0.7791
$U_{shallow} \times U_{deep}$	0.6747	0.7115	0.6881	0.0121	1.7623	0.6866
% change in DS	-1.74	3.62	0.21	1.77	841.36	

Table 43. Summary Results from Evaluation of Accuracy of Automated Rut Measurements from Vendor I against Reference Values on FM 166 Section WB5.

Rutting Statistic	Minimum	Maximum	Average	Std. dev.	CV %	Reference data
Mean LWP rut depth (in.)	0.13	0.13	0.13	0.00	0.18	0.15
Mean RWP rut depth (in.)	0.34	0.38	0.36	0.01	3.00	0.51
Mean section rut depth (in.)	0.23	0.25	0.24	0.01	2.25	0.33
Percent shallow rutting	18.23	18.41	18.33	0.08	0.43	28.57
Percent deep rutting	10.47	13.90	12.15	1.01	8.33	23.21
$U_{shallow}$	0.8938	0.8949	0.8943	0.0005	0.0548	0.8445
U_{deep}	0.7860	0.8541	0.8196	0.0201	2.4555	0.6577
$U_{shallow} \times U_{deep}$	0.7029	0.7644	0.7330	0.0182	2.4851	0.5554
% change in DS	26.56	37.62	31.97	3.28	10.26	

Table 44. Summary Results from Evaluation of Accuracy of Automated Rut Measurements from Vendor II against Reference Values on FM 166 Section WB5.

Rutting Statistic	Minimum	Maximum	Average	Std. dev.	CV %	Reference data
Mean LWP rut depth (in.)	0.23	0.23	0.23	0.0003	0.12	0.15
Mean RWP rut depth (in.)	0.88	0.90	0.89	0.0048	0.54	0.51
Mean section rut depth (in.)	0.55	0.56	0.56	0.0024	0.42	0.33
Percent shallow rutting	19.42	20.07	19.81	0.28	1.43	28.57
Percent deep rutting	24.56	25.60	24.96	0.29	1.15	23.21
$U_{shallow}$	0.8839	0.8877	0.8854	0.0016	0.1844	0.8445
U_{deep}	0.6345	0.6442	0.6405	0.0027	0.4203	0.6577
$U_{shallow} \times U_{deep}$	0.5609	0.5695	0.5671	0.0021	0.3679	0.5554
% change in DS	0.99	2.54	2.10	0.38	17.85	

Table 45. Summary Results from Evaluation of Accuracy of Automated Rut Measurements from Vendor I against Reference Values on FM 166 Section WB6.

Rutting Statistic	Minimum	Maximum	Average	Std. dev.	CV %	Reference data
Mean LWP rut depth (in.)	0.19	0.20	0.19	0.0014	0.71	0.20
Mean RWP rut depth (in.)	0.19	0.20	0.20	0.0027	1.39	0.34
Mean section rut depth (in.)	0.19	0.20	0.19	0.0020	1.04	0.27
Percent shallow rutting	20.04	23.65	22.18	1.05	4.75	50.00
Percent deep rutting	0.36	0.45	0.45	0.01	3.09	9.82
$U_{shallow}$	0.8654	0.8841	0.8727	0.0055	0.6282	0.7910
U_{deep}	1.0000	1.0000	1.0000	0.0000	0.0000	0.8684
$U_{shallow} \times U_{deep}$	0.8654	0.8841	0.8727	0.0055	0.6282	0.6869
% change in DS	25.98	28.72	27.06	0.80	2.95	

Table 46. Summary Results from Evaluation of Accuracy of Automated Rut Measurements from Vendor II against Reference Values on FM 166 Section WB6.

Rutting Statistic	Minimum	Maximum	Average	Std. dev.	CV %	Reference data
Mean LWP rut depth (in.)	0.25	0.26	0.26	0.0012	0.45	0.20
Mean RWP rut depth (in.)	0.38	0.40	0.39	0.0061	1.55	0.34
Mean section rut depth (in.)	0.32	0.33	0.32	0.0032	0.98	0.27
Percent shallow rutting	46.81	49.65	48.21	0.87	1.81	50.00
Percent deep rutting	13.31	14.14	14.03	0.20	1.44	9.82
$U_{shallow}$	0.7916	0.7966	0.7941	0.0015	0.1924	0.7910
U_{deep}	0.7816	0.7968	0.7837	0.0037	0.4692	0.8684
$U_{shallow} \times U_{deep}$	0.6188	0.6336	0.6223	0.0036	0.5769	0.6869
% change in DS	-9.92	-7.76	-9.40	0.52	-5.56	

Table 47. Summary Results from Evaluation of Accuracy of Automated Rut Measurements from Vendor I against Reference Values on FM 166 Section WB7.

Rutting Statistic	Minimum	Maximum	Average	Std. dev.	CV %	Reference data
Mean LWP rut depth (in.)	0.14	0.15	0.14	0.0010	0.72	0.12
Mean RWP rut depth (in.)	0.15	0.15	0.15	0.0014	0.96	0.20
Mean section rut depth (in.)	0.15	0.15	0.15	0.0012	0.84	0.16
Percent shallow rutting	4.33	4.87	4.48	0.13	2.98	26.79
Percent deep rutting	0.00	0.00	0.00	0.00	N/A	0.00
$U_{shallow}$	0.9946	0.9967	0.9962	0.0005	0.0525	0.8515
U_{deep}	1.0000	1.0000	1.0000	0.0000	0.0000	1.0000
$U_{shallow} \times U_{deep}$	0.9946	0.9967	0.9962	0.0005	0.0525	0.8515
% change in DS	16.80	17.05	16.99	0.06	0.36	

Table 48. Summary Results from Evaluation of Accuracy of Automated Rut Measurements from Vendor II against Reference Values on FM 166 Section WB7.

Rutting Statistic	Minimum	Maximum	Average	Std. dev.	CV %	Reference data
Mean LWP rut depth (in.)	0.12	0.13	0.13	0.0020	1.59	0.12
Mean RWP rut depth (in.)	0.18	0.21	0.19	0.0076	3.96	0.20
Mean section rut depth (in.)	0.16	0.17	0.16	0.0028	1.76	0.16
Percent shallow rutting	13.12	14.26	13.89	0.34	2.46	26.79
Percent deep rutting	0.00	1.97	0.66	0.65	97.91	0.00
$U_{shallow}$	0.9222	0.9311	0.9251	0.0026	0.2854	0.8515
U_{deep}	0.9998	1.0000	1.0000	0.0000	0.0041	1.0000
$U_{shallow} \times U_{deep}$	0.9222	0.9311	0.9251	0.0026	0.2863	0.8515
% change in DS	8.30	9.34	8.63	0.31	3.60	

Table 49. Summary Results from Evaluation of Accuracy of Automated Rut Measurements from Vendor I against Reference Values on FM 166 Section EB1.

Rutting Statistic	Minimum	Maximum	Average	Std. dev.	CV %	Reference data
Mean LWP rut depth (in.)	0.17	0.18	0.17	0.0010	0.59	0.23
Mean RWP rut depth (in.)	0.31	0.32	0.32	0.0026	0.83	0.46
Mean section rut depth (in.)	0.24	0.25	0.25	0.0018	0.73	0.34
Percent shallow rutting	26.72	29.96	27.86	0.95	3.42	42.86
Percent deep rutting	6.95	7.13	7.07	0.05	0.69	22.32
$U_{shallow}$	0.8395	0.8518	0.8473	0.0036	0.4302	0.8043
U_{deep}	0.9296	0.9336	0.9308	0.0011	0.1173	0.6671
$U_{shallow} \times U_{deep}$	0.7803	0.7953	0.7887	0.0042	0.5354	0.5366
% change in DS	45.43	48.21	46.99	0.79	1.67	

Table 50. Summary Results from Evaluation of Accuracy of Automated Rut Measurements from Vendor II against Reference Values on FM 166 Section EB1.

Rutting Statistic	Minimum	Maximum	Average	Std. dev.	CV %	Reference data
Mean LWP rut depth (in.)	0.28	0.29	0.29	0.0025	0.87	0.23
Mean RWP rut depth (in.)	0.52	0.53	0.53	0.0033	0.62	0.46
Mean section rut depth (in.)	0.40	0.41	0.41	0.0029	0.71	0.34
Percent shallow rutting	47.76	48.89	48.51	0.29	0.60	42.86
Percent deep rutting	23.58	24.49	24.00	0.39	1.62	22.32
$U_{shallow}$	0.7929	0.7949	0.7935	0.0005	0.0638	0.8043
U_{deep}	0.6449	0.6540	0.6498	0.0038	0.5911	0.6671
$U_{shallow} \times U_{deep}$	0.5115	0.5198	0.5156	0.0033	0.6371	0.5366
% change in DS	-4.68	-3.13	-3.90	0.61	-15.69	

Table 51. Summary Results from Evaluation of Accuracy of Automated Rut Measurements from Vendor I against Reference Values on FM 166 Section EB2.

Rutting Statistic	Minimum	Maximum	Average	Std. dev.	CV %	Reference data
Mean LWP rut depth (in.)	0.20	0.20	0.20	0.0007	0.34	0.23
Mean RWP rut depth (in.)	0.29	0.30	0.30	0.0012	0.42	0.55
Mean section rut depth (in.)	0.24	0.25	0.25	0.0008	0.33	0.39
Percent shallow rutting	30.87	32.04	31.48	0.33	1.05	42.86
Percent deep rutting	6.23	6.23	6.23	0.00	0.00	23.21
$U_{shallow}$	0.8325	0.8364	0.8343	0.0011	0.1308	0.8043
U_{deep}	0.9494	0.9494	0.9494	0.0000	0.0000	0.6577
$U_{shallow} \times U_{deep}$	0.7904	0.7940	0.7921	0.0010	0.1308	0.5290
% change in DS	49.42	50.11	49.74	0.20	0.39	

Table 52. Summary Results from Evaluation of Accuracy of Automated Rut Measurements from Vendor II against Reference Values on FM 166 Section EB2.

Rutting Statistic	Minimum	Maximum	Average	Std. dev.	CV %	Reference data
Mean LWP rut depth (in.)	0.27	0.31	0.29	0.0111	3.85	0.23
Mean RWP rut depth (in.)	0.63	0.70	0.66	0.0206	3.11	0.55
Mean section rut depth (in.)	0.45	0.50	0.47	0.0158	3.34	0.39
Percent shallow rutting	34.52	36.92	35.45	0.85	2.39	42.86
Percent deep rutting	13.97	18.38	16.40	1.36	8.30	23.21
$U_{shallow}$	0.8183	0.8249	0.8223	0.0023	0.2852	0.8043
U_{deep}	0.7153	0.7847	0.7450	0.0214	2.8686	0.6577
$U_{shallow} \times U_{deep}$	0.5893	0.6421	0.6126	0.0159	2.5952	0.5290
% change in DS	11.40	21.39	15.80	3.01	19.02	

Table 53. Summary Results from Evaluation of Accuracy of Automated Rut Measurements from Vendor I against Reference Values on FM 166 Section EB3.

Rutting Statistic	Minimum	Maximum	Average	Std. dev.	CV %	Reference data
Mean LWP rut depth (in.)	0.20	0.20	0.20	0.0017	0.87	0.25
Mean RWP rut depth (in.)	0.49	0.51	0.50	0.0084	1.68	0.35
Mean section rut depth (in.)	0.34	0.35	0.35	0.0050	1.44	0.30
Percent shallow rutting	32.04	33.39	32.76	0.47	1.43	58.93
Percent deep rutting	19.22	21.75	20.68	0.90	4.35	8.93
$U_{shallow}$	0.8282	0.8325	0.8302	0.0015	0.1768	0.7782
U_{deep}	0.6734	0.7040	0.6861	0.0108	1.5811	0.8885
$U_{shallow} \times U_{deep}$	0.5602	0.5831	0.5696	0.0080	1.4127	0.6914
% change in DS	-18.97	-15.66	-17.61	1.16	-6.61	

Table 54. Summary Results from Evaluation of Accuracy of Automated Rut Measurements from Vendor II against Reference Values on FM 166 Section EB3.

Rutting Statistic	Minimum	Maximum	Average	Std. dev.	CV %	Reference data
Mean LWP rut depth (in.)	0.31	0.31	0.31	0.0011	0.37	0.25
Mean RWP rut depth (in.)	0.48	0.50	0.49	0.0043	0.89	0.35
Mean section rut depth (in.)	0.39	0.40	0.40	0.0021	0.54	0.30
Percent shallow rutting	55.56	56.96	56.50	0.44	0.78	58.93
Percent deep rutting	22.10	23.93	22.77	0.57	2.49	8.93
$U_{shallow}$	0.7807	0.7826	0.7813	0.0006	0.0762	0.7782
U_{deep}	0.6504	0.6695	0.6624	0.0059	0.8972	0.8885
$U_{shallow} \times U_{deep}$	0.5090	0.5227	0.5175	0.0043	0.8225	0.6914
% change in DS	-26.38	-24.39	-25.14	0.62	-2.45	

CHAPTER 5. SUMMARY AND IMPLEMENTATION RECOMMENDATIONS

SUMMARY

Based on the analysis of data provided by TxDOT, information from Chapter 2, and discussions with three experienced providers of distress data collection, the research team submitted a draft performance specification to TxDOT as Product P1 in March 2015. After conducting the work and analysis described in Chapters 3 and 4, the team submitted the final performance specification as Product P3 in July 2016.

In addition, research team members met regularly with Magdy Mikhail at TxDOT to develop written contact language for a request for proposals (RFP) to procure pavement condition data collection services. The team used information from Chapter 2 and the results from Chapter 4 when developing the contract language. The team also gave TxDOT a list of test sections to include in the RFP.

TxDOT personnel decided that the performance specification would not be included in the RFP. TxDOT personnel will instead use the specification when auditing data submitted by the service provider that is selected by TxDOT.

IMPLEMENTATION RECOMMENDATIONS

TxDOT personnel plan to use the performance specification when auditing data from the service provider. The research team recommends that this specification be revisited when the Department obtains RFP responses from the vendors. The vendors will be submitting the results of their data analysis on the test sections.

REFERENCES

American Association of State and Highway Transportation (AASHTO) (2014). Standard Practice for Quantifying Cracks in Asphalt Pavement Surfaces from Collected Images Utilizing Automated Methods. AASHTO PP067-14-UL. Washington, DC.

American Society for Testing and Materials (ASTM) (2011). Standard Practice for Roads and Parking Lots Pavement Condition Index Surveys. ASTM D6433. West Conshohocken, PA.

American Society for Testing and Materials (ASTM) (2013). Standard Practice for Preparing Precision and Bias Statements for Test Methods for Construction Materials. ASTM C670 - 13. West Conshohocken, PA.

British Columbia Ministry of Transportation and Infrastructure (BCMoTI) (2012). Pavement Surface Condition Rating Manual. Fourth Edition. http://www.th.gov.bc.ca/publications/const_maint/2012_pavement.pdf

Caltrans – California Department of Transportation (2014). Request For Proposal - RFP Number 56A0410: Automated Pavement Condition Survey (APCS).

Caltrans – California Department of Transportation (2009). An Assessment of Automated Pavement Distress Identification Technologies in California – Technical Memorandum UCPRC-TM-2008-13. http://www.ucprc.ucdavis.edu/pdf/stage_4_3.3_eval_autom_pav_assess_techn_tm.pdf

Drosg, M. (2009). Dealing with Uncertainties: A Guide to Error Analysis. Springer: Wien Austria.

Federal Highway Administration (FHWA) (2003). Distress Identification Manual for the Long-Term Pavement Performance Project – Fourth Revised Edition. SHRP Report 338. McLean, VA. http://www.fhwa.dot.gov/publications/research/infrastructure/pavements/ltpp/reports/03031/03031.pdf

Fu, P., Lea, J. D., Lee, J. N. and Harvey, J. T. (2013). Comprehensive evaluation of automated pavement condition survey service providers' technical competence. *International Journal of Pavement Engineering*. Vol. 14, No. 1, 36–49.

Georgia Department of Transportation (GDOT) (2012). Request for Proposals: Videolog Digital Imaging Services. Atlanta, Georgia.

Iowa Department of Transportation (IOWA DOT) (2010). Request For Proposal: Automated Pavement Distress Data Collection Services.

Louisiana Department of Transportation and Development (LADOTD) (2010). Request for Qualification Statements (RFQ): Pavement Distress Data Collection. Baton Rouge, LA.

Mraz, A., Gunaratne, M., Nazef, A. and Choubane, B. (2006). Experimental evaluation of a pavement imaging system: Florida Department of Transportation's multipurpose survey vehicle. Transportation Research Record, 1974, pp. 97–106.

McGHEE, K. H. (2004). NCHRP Synthesis 334: automated pavement distress collections techniques. Transportation Research Board.

North Carolina Department of Transportation (NCDOT) (2011). Digital imagery distress evaluation handbook. https://connect.ncdot.gov/resources/Asset-Management/AssetManagementDocs/NCDOT%20High%20Speed%20Distress%20Manual%20V1.0%2011-15-2011.pdf

New Mexico department of Transportation (NMDOT) (2014). Request for Proposal: No. 14-50 Pavement Condition Data Collection.

Oklahoma Department of Transportation (ODOT) (2013). Request For Proposals (RFP): Pavement Management System Data Collection. Oklahoma City, Oklahoma.

Ong, G. P., Noureldin, S. and Sinha, K. C. (2010). Final report: automated pavement condition data collection quality control, quality assurance and reliability. Federal Highway Administration (FHWA) Technical Report FHWA/IN/JTRP-2009/17.

Pierce, L. M., McGovern, G. and Zimmerman, K. A. (2013). Practical Guide for Quality Management of Pavement Condition Data Collection. Federal Highway Administration (FHWA) Technical Report. http://www.fhwa.dot.gov/pavement/management/qm/data_qm_guide.pdf

Vavrik, W., Evans, L., Sargand, S. and Stefanski, J. (2013). PCR Evaluation – Considering transition from manual to semi-automated pavement distress collection and analysis. Ohio Department of Transportation - Final report.

Virginia Department of Transportation (VDOT) (2012). A guide to evaluating pavement distress through the use of digital images. http://www.virginiadot.org/business/resources/local_assistance/A_Guide_to_Evaluating_Pavement_Distress_Through_the_Use_of_Digital_Images_v2.6_1.pdf

Wang, K.C.P. and Smadi, O. (2011). *Automated Imaging Technologies for Pavement Distress Surveys*. Transportation Research Board, Transportation Research E-Circular E-C156.