Implementation of a 3-D Sensing Technology for Automated Pavement Data Collection in Connecticut

Prepared By:
Iliya Yut,
James Mahoney,
and Donald Larsen

Report Number: CT-2297-F-18-2

Final Report

June 25, 2018

SPR-2297

Connecticut Transportation Institute
School of Engineering
University of Connecticut

Submitted to:
Connecticut Department of Transportation
Bureau of Policy and Planning
Research and Performance Measures Unit

Edgardo D. Block
Transportation Supervising Engineer

Disclaimer

The contents of this report reflect the views of the author(s), who is/are responsible for the facts and accuracy of the data presented herein. The contents do not reflect the official views or policies of the State or the Federal Highway Administration. This report does not constitute a standard, specification or regulation.

Acknowledgements

This report was prepared by the University of Connecticut in cooperation with the Connecticut Department of Transportation and the United States Department of Transportation, Federal Highway Administration. The opinions, findings and conclusions expressed in the publication are those of the authors and not necessarily those of the Connecticut Department of Transportation or the Federal Highway Administration.

Standard Conversions

SI* (MODERN METRIC) CONVERSION FACTORS

APPROXIMATE CONVERSIONS TO SI UNITS

Symbol	When You Know	Multiply By	To Find	Symbol
		LENGTH		
in	inches	25.4	millimeters	mm
ft	feet	0.305	meters	m
yd	yards	0.914	meters	m
mi	miles	1.61	kilometers	km
		AREA		
in^2	square inches	645.2	square millimeters	mm^2
ft^2	square feet	0.093	square meters	m^2
yd^2	square yard	0.836	square meters	m^2
ac	acres	0.405	hectares	ha
mi^2	square miles	2.59	square kilometers	km^2
		VOLUME		
fl oz	fluid ounces	29.57	milliliters	mL
gal	gallons	3.785	liters	L
ft^3	cubic feet	0.028	cubic meters	m^3
yd^3	cubic yards	0.765	cubic meters	m^3
		NOTE: volumes greater than 1000 L shall be shown in m^3		
		MASS		
oz	ounces	28.35	grams	g
lb	pounds	0.454	kilograms	kg
T	short tons (2000 lb)	0.907	megagrams (or "metric ton")	Mg (or "t")
		TEMPERATURE (exact degrees)		
°F	Fahrenheit	5 (F-32)/9 or (F-32)/1.8	Celsius	°C
		ILLUMINATION		
fc	foot-candles	10.76	lux	lx
fl	foot-Lamberts	3.426	$candela/m^2$	cd/m^2
		FORCE and PRESSURE or STRESS		
lbf	poundforce	4.45	newtons	N
lbf/in^2	poundforce per square inch	6.89	kilopascals	kPa

APPROXIMATE CONVERSIONS FROM SI UNITS

Symbol	When You Know	Multiply By	To Find	Symbol
		LENGTH		
mm	millimeters	0.039	inches	in
m	meters	3.28	feet	ft
m	meters	1.09	yards	yd
km	kilometers	0.621	miles	mi
		AREA		
mm^2	square millimeters	0.0016	square inches	in^2
m^2	square meters	10.764	square feet	ft^2
m^2	square meters	1.195	square yards	yd^2
ha	hectares	2.47	acres	ac
km^2	square kilometers	0.386	square miles	mi^2
		VOLUME		
mL	milliliters	0.034	fluid ounces	fl oz
L	liters	0.264	gallons	gal
m^3	cubic meters	35.314	cubic feet	ft^3
m^3	cubic meters	1.307	cubic yards	yd^3
		MASS		
g	grams	0.035	ounces	oz
kg	kilograms	2.202	pounds	lb
Mg (or "t")	megagrams (or "metric ton")	1.103	short tons (2000 lb)	T
		TEMPERATURE (exact degrees)		
°C	Celsius	1.8C+32	Fahrenheit	°F
		ILLUMINATION		
lx	lux	0.0929	foot-candles	fc
cd/m^2	$candela/m^2$	0.2919	foot-Lamberts	fl
		FORCE and PRESSURE or STRESS		
N	newtons	0.225	poundforce	lbf
kPa	kilopascals	0.145	poundforce per square inch	lbf/in^2

*SI is the symbol for the International System of Units. Appropriate rounding should be made to comply with Section 4 of ASTM E380.
(Revised March 2003)

Foreword

The reader will note that this report contains mixed units of measurement, both SI and US Customary. The CTDOT ARAN vehicles that are used to collect pavement data and information have traditionally, since approximately 2000, utilized increments of measurement in kilometers and related metric values. However, the pavement management unit (PMU) and others who utilize the data generally report the output in US Customary units of measure. For example, the PMS reports give chainage in meters and the data in US customary (IRI in in/mi, cracking in ft, rutting in inches etc.). Due to familiarity within the CTDOT of the data collection methods, this report contains both systems of measurement.

Technical Report Documentation Page

1. Report No. CT-2297-F-18-2	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle Implementation of a 3-D Sensing Technology for Automated Pavement Data Collection in Connecticut		5. Report Date June 25, 2018 6. Performing Organization Code SPR-2297
7. Author(s) Iliya Yut, James Mahoney, and Donald A. Larsen		8. Performing Organization Report No. CAPLAB 02-2018
9. Performing Organization Name and Address University of Connecticut Connecticut Transportation Institute 270 Middle Turnpike, U-5202 Storrs, Connecticut 06269-5202		10 Work Unit No. (TRIS) N/A 11. Contract or Grant No. SPR-2297 13. Type of Report and Period Covered August 2015 – June 2018 - Final Report
12. Sponsoring Agency Name and Address Connecticut Department of Transportation 2800 Berlin Turnpike Newington, CT 06131-7546		14. Sponsoring Agency Code SPR-2297

15. Supplementary Notes
Prepared in cooperation with the U.S. Department of Transportation and Federal Highway Administration

16. Abstract

This report summarizes research on the impact of replacing older ARAN equipment (2-D Strobe cameras, ultrasonic rut bar, and inertial profiler) by the newest 3-D laser technology (Pavemetric LCMS and RoLine sensors) on the accuracy and precision of pavement performance data collected by the CTDOT. The main goal of the project was to develop transfer functions and/or correction factors relating cracking, rutting, and roughness values reported by the older and newer systems. In addition, precision and accuracy of the new technology was evaluated by comparison of data collected by two identical 3-D systems. This was to investigate comparability of historical and newly collected data. Therefore, one older and two newer ARAN vans were run simultaneously on about 127 km of preselected pavements to collect spatial location data and pavement surface images. The data was processed by the Fugro's Roadware Vision software to obtain cracking length, rut depth, and roughness (IRI) values. Next, statistical significance of differences between 2-D and 3-D pavement performance datasets was evaluated.

Overall, this project revealed the new 3-D laser scanning technology produces a much larger dataset and ultimately a much greater degree of resolution on the profiles due to an increase in the number of data points measured in both longitudinal and transverse directions. This substantially larger dataset created a smoothing effect (averaging of hundreds of values in the old system versus thousands of values in the new system) and under-estimating of the rut depths by the new system, for example. The new height sensors and accelerometers were also found to be extremely sensitive to undesirable vehicle movements that occur in response to obstacles, such as manhole covers, railroad crossings, and bridge joints. Speed gradient, including low speed and high acceleration or deceleration also adversely affected the output. The same undesirable movements apparently affected 3-D profile measurements that were used to generate 3-D images of pavement surfaces, thus, leading to differences in brightness, color, and range (depth) of images generated by the two supposedly identical ARAN systems (Van 8 and Van 9). Nevertheless, the research team believes that the precision and accuracy of the ARAN measurements can be positively improved by the adjustment of the settings and configurations of both hardware and software, the development of repeatable/reproducible system/vehicle operation routines, and the stricter use of quality assurance principles during data collection and data processing.

17. Key Words ARAN,Laser, 3-D, Pavemetric, LCMS, RoLine, Pavement, IRI	18.Distribution Statement No restrictions. This document is available to the public through the National Technical Information Service,

	Springfield, Virginia 22161. The report is available on-line from National Transportation Library at http://ntl.bts.gov.		
19. Security Classif. (of report) Unclassified	20. Security Classif. (of this page) Unclassified	21. No. of Pages 111	21. Price
Form DOT F 1700.7 (8-72)	Reproduction of completed page authorized		

Table of Contents

List of Tables

List of Figures

Executive Summary

The Connecticut Department of Transportation has recently upgraded the automated pavement data collection systems on their two Automatic Road Analyzer (ARAN) vehicles to include 3-D sensing technology which is a significant upgrade of their existing measurement capabilities. This upgrade of equipment and its sensitivity caused concern that there would be significant differences between the data collected using the older, less sensitive equipment, and the newer, high resolution collection systems. An older ARAN vehicle equipped with the older data collections and not upgraded was available for part of this research study until the equipment ceased functioning. This allowed an analysis of the old data collection system and the new data system using data collected on the same day. There were five (5) different loops run by the equipment to collect data for comparative analysis. These loops were run several times throughout the year to collect additional data as issues with the new equipment were identified and corrected. After analyzing the data, if the differences between the old and new systems was found to be too dramatically different, the research team would attempt to develop transfer functions to help prevent discontinuities caused by the new data collection system.

An analysis was performed between the different systems to determine if there was a significant difference between the different collection systems. Differences in the data when comparing the old data collection system and the new system were observed. The most significant differences were observed in the cracking data as the newer system had a higher degree of sensitivity.

During the analysis it was found that there were some substantial differences in the data being produced between Van 8 and Van 9. This difference was brought forward to the equipment manufacturer and was ultimately corrected. After the equipment was corrected, the Vans 8 and 9 were run through several of the loops and discrepancy between the data had resolved itself.

It was discovered that the higher resolution data collection systems, collected a much more robust dataset which contained many more data points. The effect of the much larger datasets was to help smooth out some of the data through the software post processing, reducing the magnitude of the highs and lows as seen with the previous data collection systems. While there was found to be a difference between the higher resolution data and the lower resolution data, it was determined that the development of transfer functions to avoid discontinuities at the transition point was not needed.

Chapter 1. Introduction

This report summarizes background information, the research methodology, and major findings from the SPR-2297 project for implementation of 3-D sensing technology for automated pavement data collection in Connecticut. The research team from the Connecticut Transportation Institute (CTI) at the University of Connecticut (UConn) performed this project for the Connecticut Department of Transportation (CTDOT). This study compares the automated pavement survey reports generated from the data collected by older and newer Automatic Road Analyzer (ARAN) vehicles to identify differences in reported road profiles, rut depths, and crack length values. In addition, the research team evaluates the precision and accuracy of the 3-D system on the newer ARANs for both project (5-m level) and network (0.1 mile level) data collection. This report also provides recommendations for further research on potential modification to CTDOT's existing Pavement Management Information System (PMIS) in order to address the effects of changes in technology of automated pavement condition surveys.

Problem Statement

In order to keep up with advances in Fugro's® Roadware suite of software, the CTDOT has been continuously upgrading pavement data collection systems. CTDOT's most advanced 3-D laser system introduces new data in a different format, which may not necessarily be available from the existing 2-D technology. For example, the 3-D laser scans the pavement surface at 4096-pixel resolution (1-mm interval) as compared with 1024 pixels (3-mm interval) by the 2-D Linescan system. This difference in resolution is anticipated to influence identification and reporting of crack severities. Furthermore, the new laser system collects more than 800 points for a transverse profile whereas the existing ultrasound system collects 25 points, which may cause differences in rut depth outputs between the two systems. In addition, the new 3-D system collects a 4-inch wide longitudinal profile as opposed to single laser line measurements collected with the older system. In summary, the increased amount of data collected by the new 3-D system may affect the reported averages of IRI and rutting, as well as the pavement condition index, which is an aggregated parameter used for making pavement management decisions.

In order to preserve year-to-year continuity in reporting the type, severity, and extent of pavement distresses, there is an identified need to match outputs from the newer and older CTDOT ARAN systems. The continuity of reported cracking, rutting, and roughness is imperative for making consistent pavement management decisions over the long term. Furthermore, collection of these data sets is required for the National Highway Performance Program (NHPP) created by the FHWA MAP-21 legislation and continued under the Fixing America's Surface Transportation Act ("FAST Act"). Therefore, while there is a need to optimize the technology transition process, there is also a competing need not to impair the annual data collection effort due to, for example, differences between the 2-D and 3-D imaging systems.

Objectives

1. Identify differences in cracking, roughness, and rutting outputs between existing 2-D and new 3-D pavement imaging systems.
2. Analyze the impact of new data-rich technology on the reported cracking, rutting, and roughness.

3. Develop transfer functions/correction factors to ensure comparability of historical and newly collected data.

Organization of Report

Chapter 2 of this report provides a background on progress and current state-of-practice in automated pavement condition surveys. Chapter 3 explains the design of the experiment and methodology of the analysis for the ARAN data and PMIS reports. Chapter 4 summarizes technical issues found with data collection in the initial phase of this project. The analysis of longitudinal profiles and ride quality reported by the 2-D and 3-D systems is covered in Chapter 5. The evaluation of transverse profiles and rut depth data are discussed in Chapter 6. Chapter 7 compares the cracking data generated from surface images collected by 2-D Strobe cameras with cracking data reported by the 3-D Linear Cracking Measurement System (LCMS). Chapter 8 presents a summary, conclusions, and recommendations for implementation of the new 3-D systems.

Chapter 2. Literature Review and Summary

This chapter summarizes the literature search that was performed for pavement condition indicators or metrics collected by State Transportation Agencies, advances in technology of pavement condition surveys, and the advantages and challenges of automated pavement data collection. The literature search also targeted discussion of the precision and accuracy of automated pavement survey systems.

Pavement Condition Indicators

With the advent of the Long-Term Pavement Performance (LTPP) program by the Federal Highway Administration (FHWA), multiple types of pavement surface cracking (e.g. longitudinal, transverse, block, and alligator) as well as patching, depressions, and rutting were standardized in terms of measurement units and severity level, and used to create a pavement condition index (PCI) system (*1*). The PCI system is sophisticated and therefore, many transportation agencies use customized pavement performance measures such as, for example, Distress Index (DI) in Michigan, and Pavement Condition Rating (PCR) in Indiana, Missouri, Maine, and Ohio (*2*). The vast majority of transportation agencies in the U.S. utilize longitudinal profile measurements, specifically, the International Roughness Index (IRI), to assess pavement serviceability (*3*). The advantage of such an approach is that profile measurements are fully automated and relatively simple to obtain and utilize. However, the IRI is only a functional indicator of pavement ride quality, whereas cracking and rutting measurements provide a better indication of structural integrity of the pavement. Therefore, many agencies combine distress and ride quality indicators into an aggregated index (*2, 3*). CTDOT had been using a custom-developed Pavement Serviceability Rating (PSR) since the 1980's which in later years was modified to the Pavement Condition Index (PCI) (*4*). Besides cracking, rutting, and ride quality, CTDOT's PCI index incorporates weighting factors for surface age and drainage condition (*4*).

Progress in Pavement Condition Surveys

Profile Surveys (IRI and rutting)

The geometric characteristics of a pavement surface, such as grade and cross-slope, are typically included in PMIS reports. The grade and cross-slope are computed from the longitudinal and transverse profiles, respectively. The longitudinal profile is also used to determine surface roughness, most often in terms of IRI. Manual measurements of roughness can be performed using a rod and level survey, or with a semi-automated device such as the DipStick™, which measures differences in elevation along a line using a 10-in (25-cm) interval (*5*). While these two methods ultimately deliver higher quality profile measurements that can be used as reference values or ground truth, they are slow and labor intensive, and thus not suitable for effective network-level determinations of IRI.

The first traffic-speed inertial profiling method was introduced in the 1960s and it became an ASTM standard in the late 1990s (*6*). This method uses a measurement of the distance between an inertial plane of reference and the traveled surface along with the acceleration of the inertial platform. The software then extracts average rectified slope, which is reported in m/km or in/mi as IRI (*7*). Other high-speed profilers use acoustic and/or infrared sensors. The acoustic sensors

are more suitable for rut depth measurements than for roughness because of their low reliability when coarse textured surfaces are encountered (*3*). In the last decade, laser sensors have reduced transverse profile measurement intervals to 10 mm at very high speeds (up to 60 mi/hr) (*3*). Transverse profile measurements are used by many agencies to report cross-slope and rut depth, using multiple (5 to 37) ultrasonic sensors installed on a rut bar, which provide transverse profile resolution intervals as large as 4 inches (100 mm) (*3*). The laser technology has improved the resolution of transverse profiles to 10-mm, thus allowing for more precise rut depth measurement (*8*).

Cracking Surveys

The technology for surveying pavement surface distresses in the U.S. has undergone significant change over the last three decades. Until the late 1980s, pavement performance ratings were based on subjective scores obtained via windshield observations or from walking surveys at representative sample locations. The progress in image technology led to development of automated photolog, and videolog systems mounted on vehicles (typically a van), which are capable of simultaneously obtaining front, back, and side views of the roadway right of way. Once collected, the two-dimensional (2-D) road images were used in an office environment to extract the amount of distresses and to rate pavement performance (*2, 11*). In the digital era, a few commercial software packages are capable of extracting quantitative distress data (mostly, cracking and rutting) from 2-D digital pavement images. The interpretation of digital images is usually complicated by factors such as camera resolution, illumination level, and cleanness of the pavement surface, as well as by speed and wandering of the measuring vehicle. The aforementioned factors not only affect the amount of distress reported by the system but also the distress severity, both of which are important for rigorous pavement rating (*8, 9, 10*). While the 2-D system can identify crack patterns very well, it often fails to report other pavement defects such as distortion, patching, rutting, weathering, and raveling (*5*).

The CTDOT pioneered incorporating technology into pavement cracking surveys. The CTDOT Research and Pavement Management groups developed and implemented a computer-based pavement rating system during the 1980s, where cracks were measured from zoom images of the pavement surface (*11*). Those measurements were recorded into a database and processed to arrive at a pavement condition index. However, beginning in 1997, CTDOT acquired vans capable of automated crack measurement through pattern-recognition and image-processing software. In 2003, the acquisition process evolved to be entirely digital. As of 2014, two Automatic Road Analyzer (ARAN) vans with 2-D imaging systems were in use: a digital technology called "Strobe" (for images collected with digital cameras and synchronized with lighting strobes) and a newer laser technology called "Linescan" (for collecting images with a laser-illuminated Linescan camera).

Challenges of the Automated Distress Survey

During the most recent five years, a new 3-D laser imaging technology was evaluated in a few states (e.g., California, Ontario, and Ohio) for the possibility of replacing the 2-D technology for cracking detection (*9, 10, 12*). The main advantages of the 3-D system are higher resolution and definition, which not only allow for capturing crack depth and rut images but also for identification of patches, depressions, and faulting. One drawback of the new 3-D measurement technology is

that it is "blind" when surface defects are covered with sand, mud, water, etc. (*12*). In addition, some manual processing and correction is still required (*9, 10, 12, 14*). The 3-D LCMS is able to report cracks as narrow as 1 mm, which naturally increases the amount of detected and reported cracks as compared with 2-D systems. This may require changes in existing PMIS ratings and pavement performance models (*12*).

While automated pavement imaging systems ultimately accelerate the process of data collection, there remains a concern about reliability and accuracy of reported distress data, particularly in terms of matching with traditional manual surveys. A recent study in Ohio reported that while about 73 percent of manually collected distresses were identified successfully by the 3-D laser system run at highway speeds, only 32 to 35 percent of severity level assignments were matched. Ultimately, the combined distress/severity/extent correlation between manually collected and vendor-equipment collected data ranged between 13 and 19 percent (*10*). A comparison of manual and automated rutting, cracking, and IRI datasets in Alabama produced even worse results indicating only 17%, 3%, and 65% of average agreement, respectively, between manual and automated surveys (*13*). The exponential increase in amount of data collected by a 3-D laser system (compared to 2-D system) requires a significant increase in computation capacity (*14*). The "black-box" character of commercially available processing algorithms adds to the complexity of the automated crack detection (3). Nevertheless, some studies report a very high level of precision and accuracy (85-95%) for some customized algorithms used for crack detection (*15*). While those higher-accuracy algorithms are not yet commercially available, an upcoming NCHRP 01-60 project has as an objective to determine traceable, objective, practical, repeatable, and transparent methods and approaches to assess the accuracy of image-based crack detection systems (*16*). This NCHRP study is not anticipated to be completed until at least 2018.

It has been reported by previous studies that despite significant deviation between automatically measured profiles and ground truth, the mean rut depth values reported by the automated systems stayed well within pavement management rutting categories determined on the ground (e.g., no rutting, low, medium, and high severities) (*10, 17*). Other factors may influence measurement, such as texture and cross slope.

The new RoLine laser system collects a 3-D longitudinal profile covering a 4-in width as opposed to the currently used single point 2-D system. The main advantage of the new system is that it is capable of smoothing out short-wavelength anomalies such as wide cracks and texture. This, however, results in the RoLine system reporting IRI values between 2 and 15 in/mi lower as compared with single-point measurements (*18*). Another potential challenge of using the RoLine system for measuring profiles is its high sensitivity to speed gradient and vehicle stops, which results in higher IRI values when greater variability of speed exists. This is particularly prevalent at speeds under 30 mi/hr and in urban environments (7, *19, 20*). A current NCHRP 10-93 project is targeting the development of means for measuring, characterizing, and reporting pavement roughness on low-speed and urban roads (*21*). Concurrently, a benchmark slow-speed profiler is being developed under Pooled Fund Study TPF-5(063) “Improving the Quality of Pavement Profile Measurement,” with 22 states participating, including Connecticut (*22*). In understanding the limitations of high-speed profilers in a slower-speed urban environment, the new Federal Rule 23 CFR §490 allows agencies to report Pavement Serviceability Rating (PSR) instead of IRI on roads with posted speeds lower than 40 mi/hr (*23*).

Conclusion

While reliable IRI and rutting measurements are enabled by collecting longitudinal and transverse profile with laser systems, full automation of cracking identification is still a challenge. The exponential increase in the amount of data collected by automated systems does not necessarily improve their precision and accuracy. In addition, there is a concern about the ability of the sophisticated laser systems to reflect ground truth measurements, especially with regard to actual crack length and severity. Lastly, changes in technology can lead to gaps in historical deterioration trends due to variations in reported pavement condition. Therefore, a successful implementation of the new 3-D laser system in Connecticut warrants an investigation of (1) system precision and accuracy, (2) differences between 3-D and existing 2-D systems in terms of reported distress and roughness values, and (3) the ability of both automated systems to reflect ground truth.

Chapter 3. Methodology

This chapter discusses the design of experiment and the analysis approach for pavement performance data collected for this project. In particular, it describes data collection and processing procedures, quality assurance of the collected data, and analytical methods of evaluation for differences in reported distress values and road profile parameters, between the older and newer ARAN vehicles.

Description of Automated Pavement Condition Survey Vehicles Utilized by the CTDOT

Table 3.1 provides a comparison of the equipment installed on and parameters applied for the three ARAN vehicles used in the experiment. Note that the older ARAN system (Van 7) has been in use since 2000, (installed on a 17-year old Ford Econoline Chassis). The two newer ARAN 9000 models, while having identical equipment, are installed on Mercedes Benz Sprinter© vans, one of which (Van 8)a 2008 model year and the other (Van 9) a 2015 model year. The three major differences in equipment between Van 7 and Vans 8 and 9 are (1) pavement surface capture method, (2) collecting longitudinal profile data, and (3) measuring transverse profile. Those differences are discussed in the same order below.

1. Van 7 is equipped with two progressive-scan digital cameras and strobe illumination that produce 2-D gray-scale images of the left and right half-lane. Vans 8 and 9 feature 3-D Laser Crack Measurement System (LCMS), which first produces a three-dimensional map of pavement surface and, then, reports collected data as a 3-D RGB color image, with the option of reporting 2-D map and depth data separately as intensity and range images. As shown in Table 3.2, the 3-D LCMS system produces 4.12x10 m surface images with three-fold high bit depth as compared with 3.8x5 m images by the 2-D Strobe system. Interestingly, the area of pavement image for one square pixel of 2-D image provides information for 9 sq. mm of surface, whereas the 3-D system compresses 64 sq. mm into one square pixel, thus providing lower resolution. This is derived by dividing the area of captured pavement by the area of the image in pixels.

2. Vans 8 and 9 are equipped with RoLine 4" Footprint Line Lasers installed on the bumper of the vehicle to measure height above the surface, which is later used to compensate vehicle body movements in IRI calculations. The RoLine measures a 4-in wide longitudinal surface in each wheelpath, from which an optimal line profile is computed. The Selcom sensors on Van 7, on the other hand, produce a single line profile for both left and right wheelpaths. As noted in Table 3.2, both Selcom and RoLine sensors are able to report data with a 100-mm interval.

3. With respect to transverse profile measurements, Van 7 and Vans 8 and 9 employ significantly different technology. The rut bar installed on the front bumper of Van 7 has 37 ultrasonic sensors installed at 100-mm intervals, which allow capturing 3.6 m of pavement width. Due to safety concerns, however, only 25 sensors are used by the CTDOT, which reduces the captured width of transverse profile to 2.4 meters. Vans 8 and 9 employ infrared lasers, which are capable of scanning up to 4.2 m of pavement width with a 10-

mm transverse interval. This drastic increase in the number of measurements across the lane results in a smoother profile, which affects rutting depth calculations.

Table 3.1 Summary of Data Collection Equipment on CTDOT ARAN vehicles

System equipment	ARAN Vehicle ID (System, Chassis, Make Year)		
Chassis	Van 7 (ARAN, Ford Econoline, 2000)	Van 8 (ARAN 9000, Mercedes Benz Sprinter, 2008)	Van 9 (ARAN 9000, Mercedes Benz Sprinter 2015)
Chassis Characteristics	Wheelbase: 138 in Width: 79.3 in Curb weight: 5140 lbs	Wheelbase: 144.3 in Width: 79.7 in Curb weight: 5050 lbs	Wheelbase: 144.3 in Width: 79.7 in Curb weight: 5050 lbs
Geographic Coordinates	Leica GPS receiver	Real-Time Differential GPS +POS LV Inertial Positioning System	Real-Time Differential GPS +POS LV Inertial Positioning System
Distance	Wheel-mounted Distance Measurement Instrument	Wheel-mounted Distance Measurement Instrument	Wheel-mounted Distance Measurement Instrument
Longitudinal Profile	Two Selcom Laser Height Sensors	Two RoLine 4" Footprint Line Lasers	Two RoLine 4" Footprint Line Lasers
Roughness	Axle-mounted accelerometers	Axle-mounted accelerometers	Axle-mounted accelerometers
Transverse Profile	Ultrasonic Rut Bar	3-D Pavemetrics Laser Crack Measurement System	3-D Pavemetrics Laser Crack Measurement System
Pavement Surface Image	Two Progressive-Scan Digital Cameras with Strobe Illumination	3-D Pavemetrics Laser Crack Measurement System	3-D Pavemetrics Laser Crack Measurement System
Front View Image	HDTV Camera	HDTV Camera	HDTV Camera

Table 3.2 Summary of Differences in Data Collection Settings

	ARAN Vehicle ID (System, Make Year)		
Measured Parameters	Van 7 (ARAN, 2000)	Van 8 (ARAN 9000, 2015)	Van 9 (ARAN 9000, 2015)
Geographic Coordinates Units	Longitude: Radians Latitude: Radians Elevation: meters	Longitude: Radians Latitude: Radians Elevation: meters	Longitude: Radians Latitude: Radians Elevation: meters
Distance Units	meters	meters	meters
Longitudinal Profile Interval	100.2 mm	300 mm (Jun 2016 data) 100 mm (Nov 2016 data)	100 mm
Accelerometer Data Interval	12.5 mm	25 mm (Jun 2016 data) 12.5 mm (Nov 2016 data)	12.5±0.01 mm
Transverse Profile Interval and Resolution	5 meter/ 100 mm	100 mm/ 10 mm	100 mm/ 10 mm
Pavement Surface Image Characteristics	Pavement WxL:3.82x5m Image WxH: 1204x1525px Bit Depth: 8	Pavement WxL: 4.05x10m Image WxH: 511x1250px Bit Depth: 24	Pavement WxL: 4.12x10m Image WxH: 520x1250px Bit Depth: 24
Front View Image Characteristics	Image WxH: 1920x1080px Bit Depth: 24	Image WxH: 1920x1080px Bit Depth: 24	Image WxH: 1920x1080px Bit Depth: 24

Data Collection Plan

In order to evaluate differences between older (Van 7) and newer (Van 8 and 9) ARAN vehicles in terms of reported pavement performance data, the three vans were run simultaneously on a pre-selected continuous 127-km loop of consecutive pavement sections, hereafter called the 'Big Loop.'. The sections contained in Big Loop were chosen to have different pavement types, surface ages, and pavement condition as measured in PSR, IRI, and rut depth. Other variables included extent of rutting and functional classification. A simultaneous run by all vehicles was designed to eliminate such factors as variable speed, differences in trajectories (wandering) and differences in brightness and contrast of the Photolog images that might occur with variable ambient light due to weather and time of day. Figures 3.1 through 3.3 illustrate the breakdown of the Big Loop by roadway class, surface age, and surface condition, respectively.

As shown in Figure 3.1, a 4-lane divided Interstate and a 4-lane divided Connecticut State Route comprised 43% of the surveyed lengths, whereas 51% of surveyed sections were undivided local arterials and collectors. The rest of the loop consisted of access ramps for which no PMIS information was available. Figure 3.2 indicates that before the project started (2014), new, intermediate, and old surfaces comprised 10, 49, and 28 percent of the surveyed loop length, respectively, with no age information available for 12 percent of the loop. In 2015, about 25 percent of the loop was overlaid, resulting in a relatively balanced distribution of ages (37, 35, and 28 % for new, intermediate, and old surfaces, respectively).

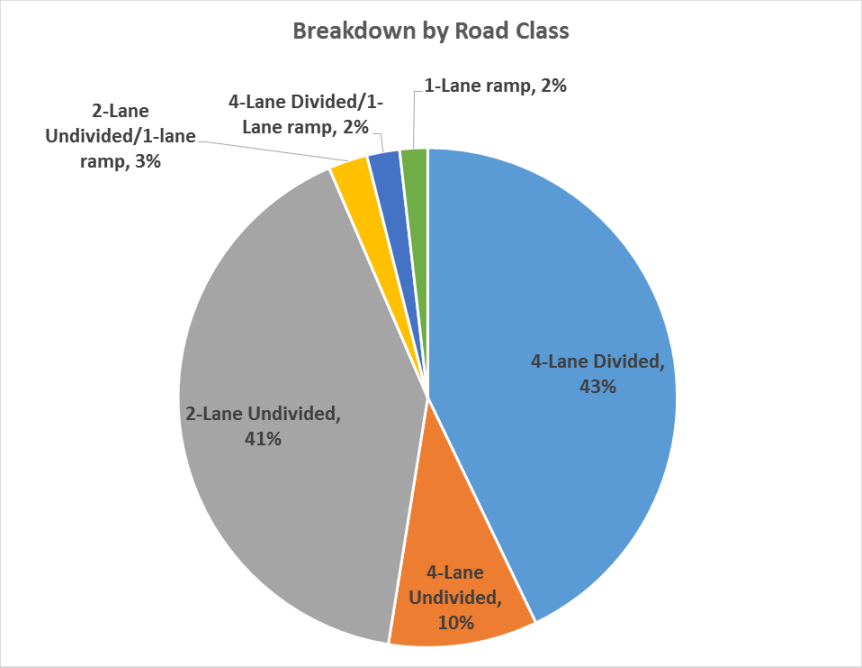


Figure 3.1 Breakdown of "Big Loop" by road classification.

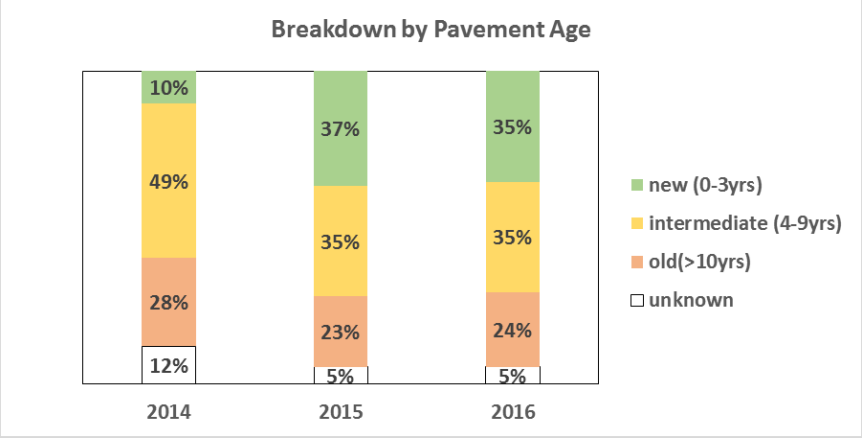


Figure 3.2 Breakdown of 'Big Loop' experimental pavement set by age group.

Figure 3.3 indicates that the surface condition of the surveyed Big Loop kept changing through the two years of survey. One can notice that before the survey (2014), 85% of the loop was in fair to

poor condition, whereas after two years of maintenance operations, 50% of the loop yielded a good rating in terms of cracking index as calculated per CTDOT guidelines. The reason for repeated surveys is explained below.

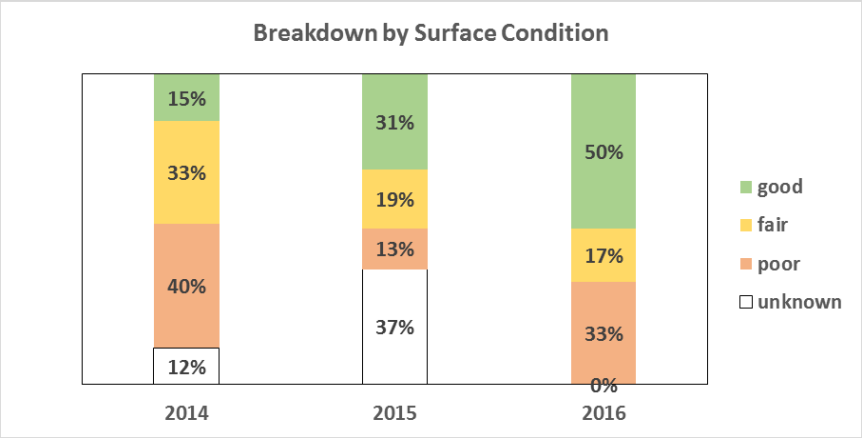


Figure 3.3 Breakdown of "Big Loop' experimental pavement set by road surface condition.

Implementation of Data Collection

First Run in September 2015

The first simultaneous run of Vans 7, 8, and 9 was completed in September 2015 on the 127-km Big Loop, as well as a 10-km roadway section of I-84 between West Hartford and Farmington, CT. However, a preliminary analysis of PMIS reports generated from the ARAN data revealed some rather unexpected differences in reported crack extent, rut depths, and IRI values both between Van7 and Vans 8 and 9 as well as between the two new vans (8 and 9). The significant differences between Vans 8 and 9 were not anticipated due to known apparent equivalency of onboard hardware and software, and were attributed to technical issues and/or variations in ARAN sensor configurations, as well as vehicle speed and wandering. In addition, problems with image processing software precluded adequate comparison of cracking values reported by the 2-D Strobe and 3-D LCMS systems. Other issues included missing data due to occasional malfunctions of some onboard systems on both Vans 8 and 9 during the Big Loop run. Therefore, after the vendor performed preventive maintenance on the ARAN data collection systems, the Big Loop run was repeated in June 2016.

Second Run in June 2016

In June 2016, all three vans were simultaneously run through the 127-km Big Loop and the data generated by their ARAN systems was analyzed for differences. At that time, the sensor data was successfully retrieved for all vehicles for the whole length of the Loop, which allowed for a thorough analysis of longitudinal and transverse profiles. However, a malfunction of the 2-D Strobe cameras on Van 7 reduced the comparable cracking dataset to only 40 km.

Repeatability Runs in November 2016

In November 2016, in order to assess precision and accuracy of the new 3-D LCMS system and RoLine lasers, Van 8 and 9 were simultaneously run 5 times on 5 lane-km of Thornbush Rd. in Wethersfield, CT, and Brook St. in Rocky Hill, CT. The 127-km Big Loop was surveyed by Vans 8 and 9 again to assess reproducibility of the survey by those two vehicles.

ARAN Data Processing

Originally, The UConn team was to receive all pavement condition data in the form of a SQL database, populated by processing the ARAN data through Roadware Vision software at the CTDOT. Due to technical issues during the survey in September 2015, a significant portion of the ARAN data was not processed. The 2016 survey data from Vans 8 and 9 (3-D LCMS + RoLine) was processed by the UConn team using Roadware Vision, whereas the Van 7 cracking data was processed by the CTDOT using 2D WiseCrax software. Van 7 rutting data was also retrieved by the CTDOT and transferred to the UConn team.

Roughness Data Processing

The roughness (IRI) was computed by using the Roadware Vision Roughness Processor. This processor uses raw roughness data including acceleration values from body-mounted sensors and elevation data from RoLine sensors to compute IRI values in SI Units (m/km). Due to a high sensitivity to survey speed that was revealed in the 2015 survey, a post-processing algorithm was employed for the 2016 survey data to remove unusual IRI values caused by stops and/or large speed gradients. Lastly, the 5-m and 0.1-mi average IRI values in U.S. customary units (in/mi) for each wheelpath from the post-processed data were reported.

Rut Depth Data Processing

The rut depth computation and reporting methods for Van 7 (ultrasonic rut bar) and Vans 8 and 9 (3-D laser system) differ significantly. The Van 7 system reports rut values directly to a user of ARAN data. Those values are computed from a transverse profile every 5 meters of travelled survey based on a straight edge method. The new ARAN systems compute the transverse profiles first, while extracting vertical measurements from the 3-D spatial data. Note that transverse profiles are reported for each 0.1 meter of travelled survey. Next, a Rutting Processor is employed where the method of rut depth calculation is defined. Thus, the 3-D system reports an average of 50 rut depth values for each wheelpath every 5 meters, as compared with Van 7's single value. To ensure comparability with the new system, Van 7 transverse profile data was processed through an Excel routine mimicking the half-lane straight edge algorithm for rutting used by vans 8 and 9 and compared with rut values output directly from Van 7 ARAN data. The 3-D system's rutting data from the PMIS reports was used for the analysis.

Grade and Cross-Slope Data Processing

The Van 7 system uses gyroscope pitch and roll values to compute grade and crossfall as angular values reported in radians. Both grade and crossfall are adjusted to vehicle movements relative to the pavement surface. While grade values can be found in Attitude data tables, the crossfall is reported in conjunction with transverse profile readings from the rut bar in the Rut Report data table. All aforementioned tables are generated directly by the ARAN on-board system. Interestingly, the heading, pitch, and grade measured by Van 7 is reported with 4-meter intervals,

whereas the roll and crossfall are collected with a 5-meter interval. The new ARAN system (Vans 8 and 9) use the same approach as Van7 to compute grade and crossfall. However, due to increased frequency of measurements (20 Hz by Vans 8 and 9 as compared with 3Hz by Van 7), the average values may differ significantly.

The CTDOT PMIS uses U.S. customary units, whereas SI (metric) units are used in the SQL database tables generated from the ARAN data. Therefore, the PMIS report generation processor in Vision software converts angular values from radians to degrees. In order to ensure comparability, the Van 8 and 9 cross-slope in degrees had to be converted to units of slope as reported by Van 7.

Cracking Data Processing

The new 3-D laser system generates 3-D spatial data that is processed through Roadware Vision software to generate intensity, range, and 3-D gray-scale images of the pavement surface. The LCMS processor uses those images to classify and rate cracks. The crack data for Vans 8 and 9 was generated by the UConn team. Three classes of cracks (longitudinal, transverse, and area) were identified in each of five zones across the lane and rated by severity level, as summarized in Table 3.3. The severity of cracks in this project was determined based on crack width.

The Strobe 2-D images from Van 7 could not be processed through the Vision software and, therefore, were processed through stand-alone WiseCrax software by the CTDOT. The main difference between WiseCrax and Vision is that WiseCrax does not identify area cracking. In addition, the severity thresholds in WiseCrax for Low/Moderate and Moderate/High are defined by crack width of 3 and 6 mm, respectively, as compared with 6 and 12 mm in Vision (Table 3.3).

Table 3.3 Summary of Cracking Classification

Processing System	Crack Type Class	Crack Severity Levels (width range)	Widths of Lane Zones	Total number of zone-crack types
3-D LCMS	Longitudinal Transverse Area	Low (0-6 mm) Moderate (6-12 mm) High (>12 mm)	Left Edge (~0.625 m) Left Wheelpath (0.75 m) Center (0.75 m) Right Wheelpath (0.75 m) Right Edge (~0.625 m) Total:~3.5 m	3*3*5 =45
2-D WiseCrax	Longitudinal Transverse	Low (0-3 mm) Moderate (3-6 mm) High (>6 mm)	Left Edge (0.625 m) Left Wheelpath (0.75 m) Center (0.75 m) Right Wheelpath (0.75 m) Right Edge (0.625 m) Total:~3.5 m	2*3*5 = 30

After the processing as described above, the cracking data for the analysis in this project was summarized in PMIS reports with 5-m (project level) and 0.1-mi (network level) intervals.

Segmentation and Matching

While all three ARAN vehicles travelled in tandem over the whole length of survey, due to operator factors and preset frequency of recording, for example, they recorded different lengths of travel, and reported different amounts of data, Therefore, matching the beginning and the end of each segment was mandatory to ensure comparability of the datasets. For this study, the Big Loop was subdivided into five segments (A through E) as summarized in Table 3.4 and illustrated in Figure 3.4. The edges (boundary) of each segment were matched by GPS latitude and longitude within a 5-m interval. The Van 8 and 9 segments were matched automatically through the Vision Global Processor with great precision, while Van 7 segment edges were matched manually to those of Vans 8 and 9.

Table 3.4 Summary of Sections and Segments

Section ID	Segment ID	Start	End	Matched Length by 5-m Stations [lane-meter]		
				Van7	Van8	Van 9
Big Loop	A	Rt. 15 NB (CTDOT) Newington	Rt. 17 &Rt. 160 Glastonbury	19905	19859	19859
	B	Rt. 17 &Rt. 160 Glastonbury	Rt. 66 & Main St. Marlborough	27359	27325	27325
	C	Rt. 66 & Main St. Marlborough	Rt. 02 Exit to Rt. 03 SB Glastonbury	18482	18491	18491
	D	Rt. 02 Exit to Rt. 03 SB Glastonbury	Country Club Rd & Middle St., Cromwell	21726	22009	22009
	E	Country Club Rd & Middle St., Cromwell	Rt. 15 SB (CTDOT), Newington	38922	38898	38898
Repeat ability Runs	Thornbush Rd. SB	Thornbush Rd. & Rt. 287, Wethersfield	Thornbush Rd. & Highland St., Wethersfield	No data	930	930
	Brook St. EB	Brook St. &Rt. 3, Rocky Hill	Brook St. & Rt. 99, Rocky Hill	No data	2680	2680
	Brook St. WB	Brook St. & Rt. 99, Rocky Hill	Brook St. &Rt. 3, Rocky Hill	No data	2680	2680

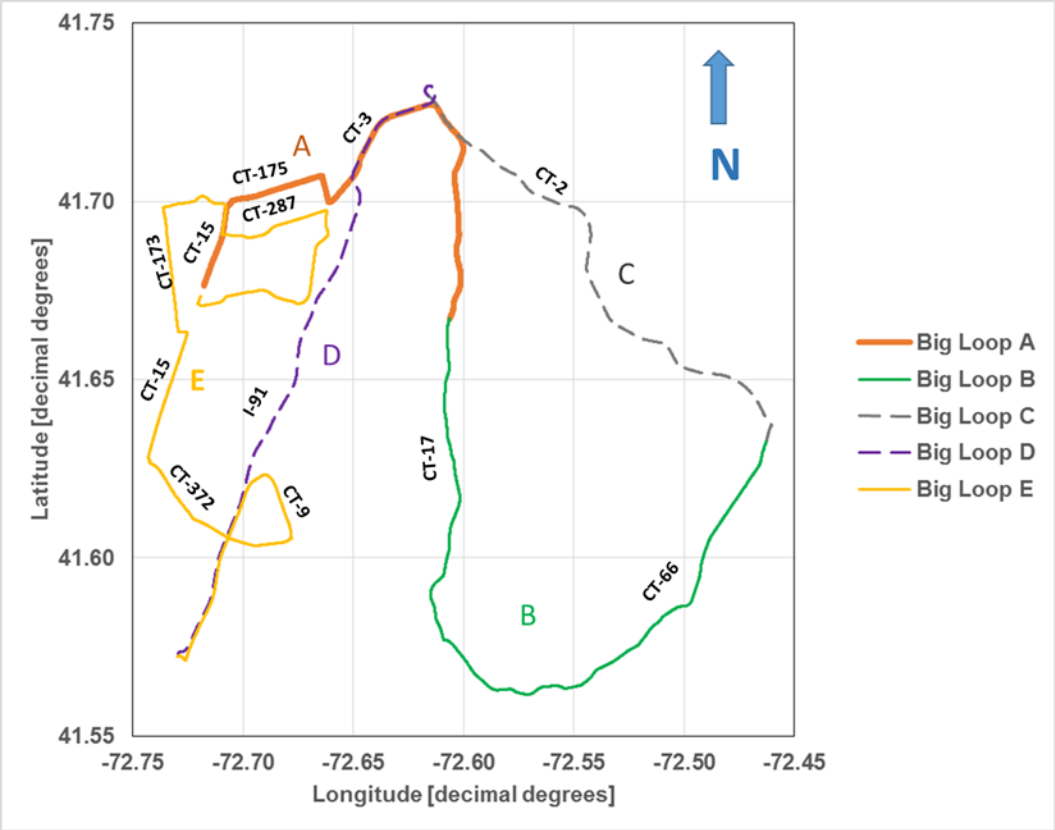


Figure 3.4 Spatial location of Big Loop segments.

Analysis of PMIS Reports

The new 3-D laser survey system produces a great deal of data with very fine resolution that differs by an order of magnitude from the 2-D system. To ensure comparability of the 3-D and 2-D datasets, the UConn team used the 5-m and 0.1-mi average values to create IRI, rutting, grade, and cross-slope datasets for each ARAN vehicle involved in the experiment (Van 7, 8, and 9). With respect to cracking data, sum of cracks by type class, severity, and location (lane zone) per 5 m and 0.1 mile were computed separately to create cracking datasets for each ARAN vehicle. The steps and routines of the survey data analysis are illustrated in Figure 3.5.

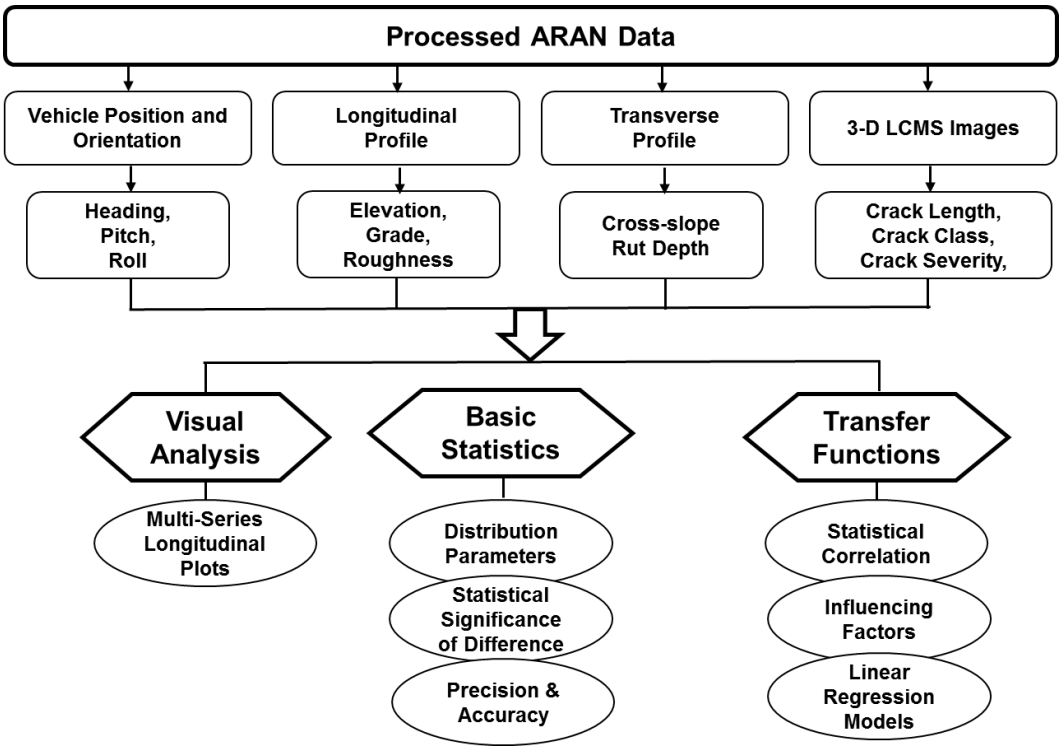
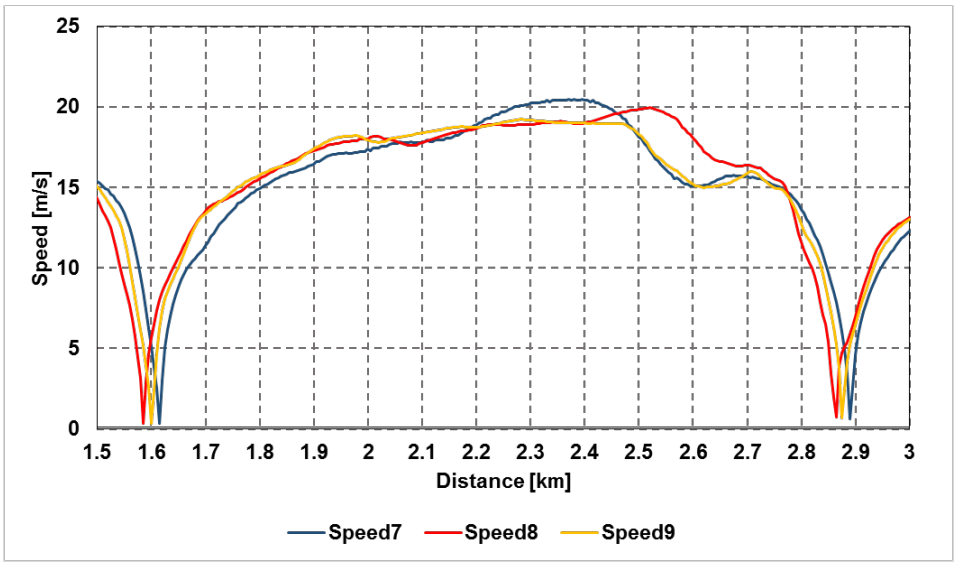


Figure 3.5 Flow chart of ARAN data processing and analysis.

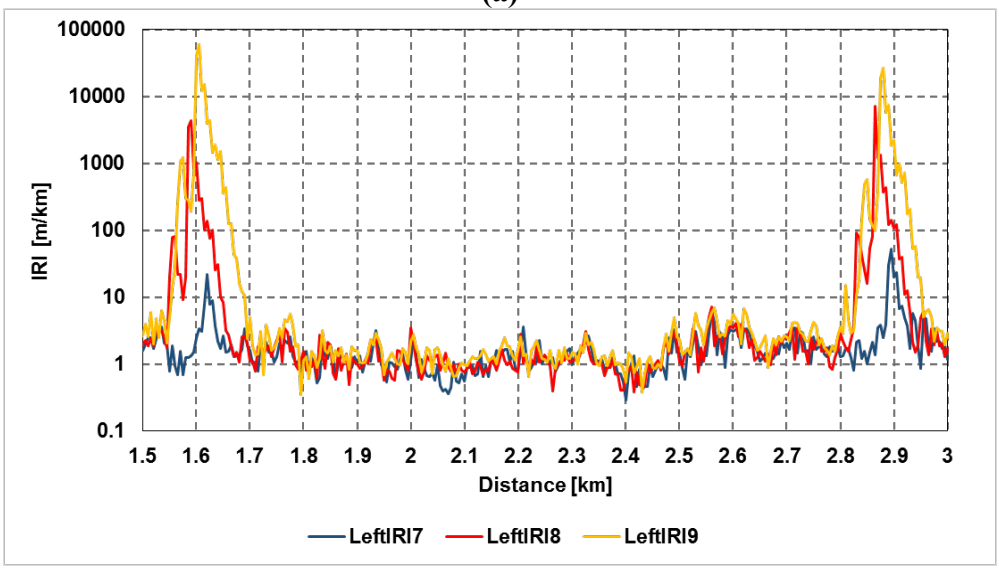
Visual Analysis of Plots

The main variables of interest in this project were IRI, grade and cross-slope, rut depths, and crack lengths. For each of those variables, three data sets were available from the three ARAN vans. The IRI and rut depths were reported for left and right wheelpath, separately. Plotting three series on the Y- axis along the length of segment/section on the X-axis allowed identifying unusual values and trends as well as comparing patterns. For example, initial analysis of IRI plots for 5-meter stations shown in Figure 3.6, revealed super sensitivity of the new accelerometers for Vans 8 and 9 to significant changes in velocity, including stops. It is clear that slowing followed by accelerating of Vans 8 and 9 (Figure 3.6a) results in 2 to 3 orders of magnitude (i.e., 100 to 1000 times) higher reported IRI as compared with Van 7 (Figure 3.6b). One can also notice that, while Van 7 exhibits some sensitivity to large speed gradient, its IRI peak width (~50m) is much shorter than the IRI peak width reported by Vans 8 and 9 (~200 m), resulting in fewer "bad data" values.

In another example, Figure 3.7 superimposes longitudinal, transverse, and area cracking for Big Loop segment D. One observation is that crack lengths vary by at least three orders of magnitude, which suggests that cracking data has a logarithmic distribution. Secondly, the patterns of the three classes of cracking are similar, which suggests strong correlation between longitudinal, transverse, and area crack lengths reported by the LCMS. However, such a trend may not necessarily reflect the ground truth because the LCMS may misclassify a longitudinal or transverse crack as an area crack, for instance.



(a)



(b)

Figure 3.6 Longitudinal plot of average 5-m speed (a) and IRI (b) for Big Loop segment A with two stops at around km 1.6 and km 2.875.

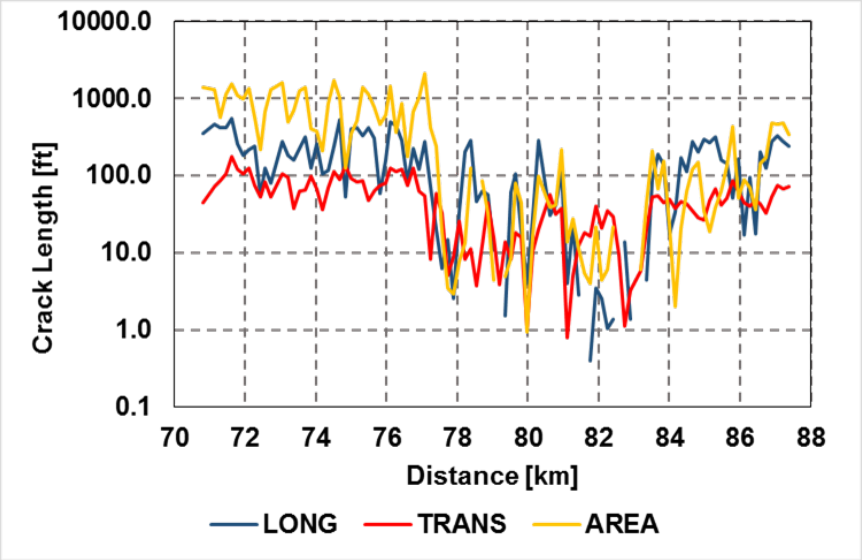
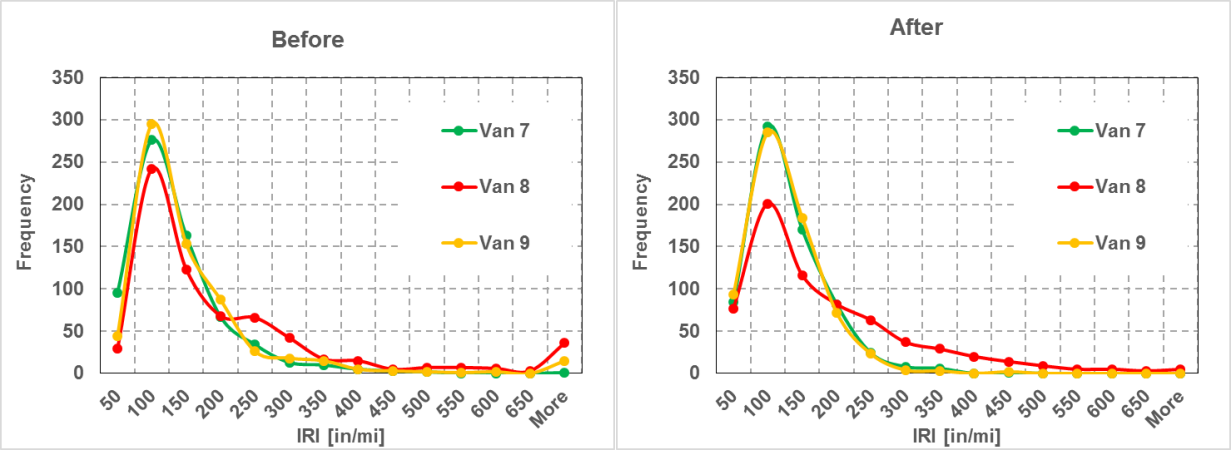


Figure 3.7 Crack length trends along Big Loop segment D.

Basic Statistics (Univariate Analysis)

Analysis of Distributions. The survey of pavement surfaces by three ARAN vehicles generated massive amounts of data, which definitely fell under the definition of a large dataset. Most of the investigated metrics, such as IRI, rutting, or cross-slope, for instance, were averaged over 5-m and 160-m for the analysis. This yielded more than 25,000 and 800 measurements for project and network level analyses, respectively, for each of the investigated parameters in a given ARAN dataset. While the ultimate goal of this project was to develop transfer functions from historical ARAN (Van 7) data to new data from Vans 8 and 9, the first step was to compare the distributions of said parameters measured by different ARAN systems. The distributions were characterized by basic statistics, such as mean, median, and standard deviation. The distribution shapes were looked upon to estimate their normality and identify unusual values. Ultimately, the analysis of distributions helped to identify potential forms of relationships between parameters of interest.

An example of the distribution of IRI values before and after running a post-processing algorithm is shown in Figure 3.7, whereas Table 3.5 compares basic statistics for the 160-m average IRI between Vans 7, 8, and 9. Figure 3.8 (a) shows that Vans 8 and 9 recorded a significant amount of illogical, higher than 600 in/mi, IRI values. Furthermore, Van 8 stood apart from Vans 7 and 9, while reporting more IRI measurements within a "poor" range of 250 to 400 in/mile. Removing IRI values obtained during low speeds, by a post-processing routine, helped eliminate most illogical values for both Vans 8 and 9. However, the distribution of Van 8 remained dispersed for reasons unknown at the time of analysis (see Figure 3.8 (b)), however, postulated to be likely due to problems with hardware. The data in Table 3.5 indicate that on average, IRI values reported by Van7 changed very little, as compared with Vans 8 and 9, which indicates that Van 7 motion sensors are much less sensitive to acceleration/deceleration than those installed on the newer ARAN vehicles.



(a) (b)

Figure 3.8 Distributions of 0.1-mi average IRI values from ARAN datasets (a) before and (b) after post-processing.

Table 3.5 Summary of distribution statistics for 0.1-mi average IRI from ARAN datasets before and after post-processing.

Statistics	Van 7		Van 8		Van 9	
	Before	After	Before	After	Before	After
Minimum	32	32	38	34	39	34
Maximum	780	416	22495	1092	14781	450
Median	92	92	127	126	98	92
Average	112	104	285	164	180	102
Standard Deviation	74	55	1075	134	677	53

Significance Tests for Difference. The *Student t-test* of difference of means computes the probability of mean values for two variables of being the same. If such probability is less than the level of significance (equal 100% - level of confidence), than the null hypothesis about zero difference between mean is accepted. A *paired t-test* can also be performed on the mean difference between two measurements by different instruments for the same section. As an example, Table 3.6 summarizes the two-sided t-test results for vans 7, 8, and 9 on segments B, C, and D of the Big Loop. The level of significance is established as 0.05, and zero difference is assumed. The results indicate that no statistically significant difference (or bias) exists between paired network-level IRI reported results for the analyzed segment. On the other side, the paired t-test detected highly significant, yet of low magnitude, bias between Vans 8 and 9. One could argue, however that the average difference of 3 in/mi is only about 3% of the 90 in/mi, which is actually acceptable bias and the reproducibility of the survey is therefore not a concern.

Table 3.6 Summary of paired t-test for significance of difference between 0.1-mi average post-processed IRI.

Parameter	Van 7		Van 8	Van 9
Mean	93.2		93.8	90.8
Standard Deviation	50.3		48.9	45.6
Number of observations	423		423	423
t-test Parameter	Van 7 vs. Van 8	Van 7 vs. Van 9	Van 8 vs. Van 9	
t Critical one-tail	1.966	1.966	1.966	
t Stat	-0.452	1.952	5.755	
P(T<=t) one-tail	0.325	0.052	0.000	
Conclusion	No difference between Vans 7 and 8	Marginal difference between Vans 7 and 9	Significant Difference between Vans 8 and 9	

Precision and Accuracy

With the introduction of two identical 3-D systems installed on vans 8 and 9, evaluation of random and systematic errors was conducted to determine precision and accuracy of the new pavement survey technology. Random errors result in dispersion, or low precision, around a reference value, while systematic errors shift the observed mean of a series of measurements away from the actual value, resulting in bias or low accuracy (See Figure 3.9). In this project, the precision was measured as a coefficient of variation (C.o.V.), i.e. ratio of standard deviation over the mean of five repeated measurements. Since both vans were run simultaneously over a number of 0.1-mile sections, the average C.o.V. value reflected its precision. The ground truth-value was not available for this analysis. Therefore, the mean of two means reported by Van 8 and 9, respectively, served as a reference value for accuracy analysis. The accuracy of measurements was determined then as mean C.o.V. from reference value over a segment (e.g., Thornbush Road or Brook Street).

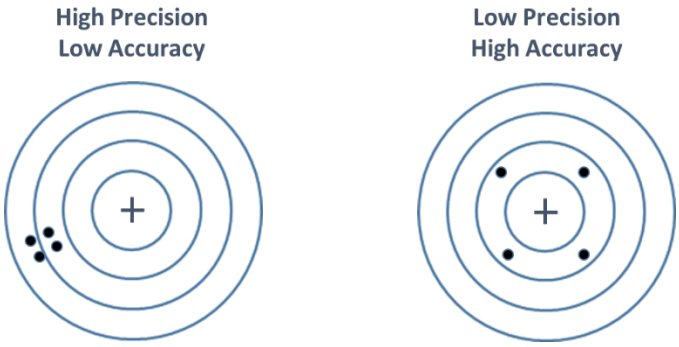


Figure 3.9 Illustration of concepts for precision and accuracy.

An example of precision and accuracy analysis is given in Table 3.7, which shows the mean and their corresponding 95%-confident C.o.V. values (in parentheses) for longitudinal, transverse, and area crack lengths per 0.1 mile. One can notice that both Van 8 and Van 9 produce good repeatability (precision) for all type of cracks with longitudinal cracks reported with highest precision of 5 percent (10 percent with 95- percent confidence), while lower precision of transverse and area cracking. The very high precision of total crack length measurement, however, indicates that the processing software may classify the same crack as a transverse or area randomly, while not affecting the total length report. The lower values of C.o.V. between Van 8 and Van 9 indicate that accuracy of the new 3-D system is slightly higher than its precision. This means that random misclassification of cracking due to the processing algorithm settings is more likely than bias between the two pavement survey vehicles.

Table 3.7 Comparison of precision of cracking measurement by Van 8 and Van 9.

Crack Type (ft/0.1-mi)	Precision (Within) Mean C.o.V. (95% Confident Mean C.o.V.)		Accuracy (Between) Mean C.o.V.
	Van 8	Van 9	
Longitudinal	5% (10%)	5% (10%)	4% (11%)
Transverse	11% (21%)	10% (13%)	7% (18%)
Area	8% (13%)	10% (16%)	6% (17%)
Total All Types	4% (8%)	5% (9%)	3% (11%)

Development of Transfer Functions.

The ultimate goal of this project was to develop correction factors and/or transfer functions from the outputs reported by the historical ARAN system to the outputs produced by the new 3-D Laser system. The survey outputs were represented by pavement performance metrics such as IRI, rut depth, and crack length. In addition, the differences between the historical and new ARAN systems in reported grade and cross slope were of interest. In engineering, a transfer function is defined as a mathematical representation of fit, or relationship between input and output of the so-called "black-box" models. The WiseCrax crack detection algorithm is one example of such a model where there is no obvious relationship between collected surface image and the reported crack

length and severity. The same goes for determining the rut depth from the fitted transverse profile or computing IRI based on accelerometer data when using Roadware Vision processors. Thus, the relationship between cracking outputs reported by Van 7 (2-D WiseCrax) and Van 8 (3-D LCMS), for instance, is also a transfer function from one "black-box" function to another.

In this project, a simple linear correlation between older and newer system's outputs was first evaluated by fitting a straight line. Ideally, a slope within 0.95 to 1.05 with no intercept and coefficient of determination, or R-squared, of greater than 95%, would be considered a good fit. If no good fit could be achieved by simple regression, the influencing factors of the difference between ARAN systems, such as road classification, surface condition rating, and survey speed were introduced into the relationship as categorical variables in an attempt to improve the correlation. An example of a bivariate plot of total crack lengths per 0.1 mi as reported by Van 7 and Van8 for a 16-km long segment of the Big Loop is shown in Figure 3.10. One can notice that, while the intercept of the trend line to predict Van 8 cracking from Van 7 is very small, its slope (1.424) indicates that, on average, Van 8 reports 40 percent more cracking per 0.1 mile than Van 7.

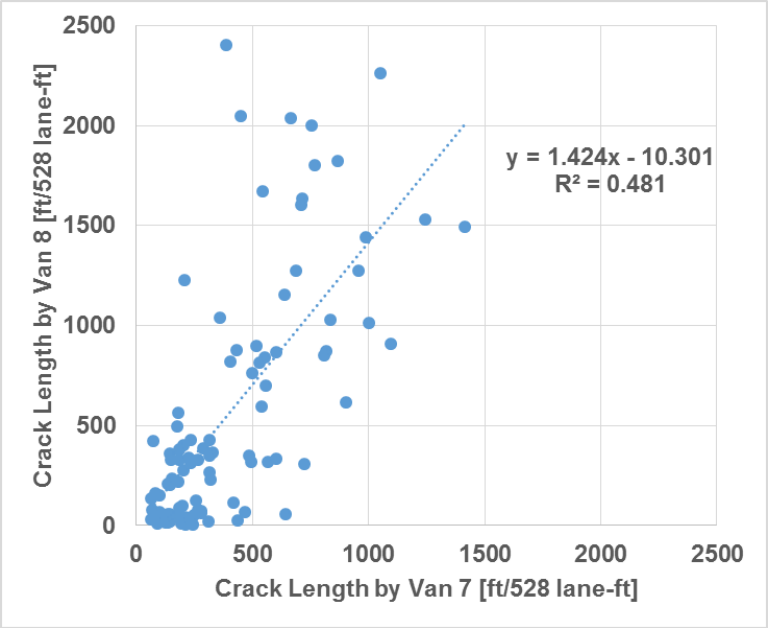


Figure 3.10 Bivariate plot of cracking lengths per 0.1mile reported by Vans 7 and 8.

In an attempt to improve the correlation between the Van7 and Van 8 datasets, the factor of condition rating is introduced into the regression model, as shown in Equation 3.1:

$$Cracking\_Van7 = \beta_0 + \beta_1 * Cracking\_Van8 + \beta_2 * RATING \quad (3.1)$$

where β_0, β_1, *and* β_2 are regression coefficients.

Table 3.8 summarizes regression coefficients for the above model, with corresponding p-values indicating the statistical significance of the variables. The introduction of RATING factor into model (3.1) results in slightly improved R-sq. of 0.57 and a slope of one, while indicating clear bias between Van 8 and Van 7, depending on pavement condition. Note that the difference between the crack lengths reported on the poor and good surfaces are significant (p-values smaller than level of confidence α=0.05) as opposed to fair surfaces.

Table 3.8. Summary of Regression Analysis for Predicting Van 8 Cracking from Van 7 Cracking

Model (3.1)		
Variable Name	**Estimate**	**p-value**
Cracking_Van7 (Slope)	1.000	0.0000
RATING[POOR] (Intercept)	496	0.0000
RATING[GOOD] (Intercept)	-121	0.0006
RATING[FAIR] (Intercept)	274	0.1268
Model Parameters	Number of cases	94
	R-Squared	0.57
	Standard Error	411

Summary of Analysis Approach

As described in this Chapter, a multi-faceted approach to the analysis of ARAN data was implemented to derive conclusions on the effect of change in ARAN technology on the reported pavement performance indicators. As shown in the flowchart in Figure 3.4, the visual analysis of changes in those indicators over the length of surveyed segments was followed by the statistical analysis of the relationships between the corresponding measurements reported by the three ARAN systems (Vans 7, 8 and 9).

Chapter 4. Analysis of Results and Technical Issues Found with 2015 Data Collection

The first round of pavement surveys with ARAN vans 7, 8, and 9 was conducted in the summer of 2015. In accordance with the original objectives of this project, pavement images and profile data were collected on the 127-km long 'Big Loop' in order to generate roughness, rutting, and cracking data. In addition, roughness and rutting data were collected on I-84 between exits 46 and 39 (West Hartford to Farmington). The ARAN data was processed by the CTDOT Photolog Unit and the resultant database in SQL format was transferred to the UConn research team for the analysis. Due to constraints of the original schedule (9 weeks between data collection and submission of the preliminary results) and some missing data, only data from a 27-km section of Big Loop (Segment B) and I-84 was examined. The primary goal of the analysis was to identify differences in cracking reported by the 2-D Strobe (van 7) and the 3-D Laser (Vans 8 and 9) systems, and differences in IRI and rutting values between all three ARAN vehicles.

Missing Data

The analysis of 2015 data began with examining the SQL databases for completeness. It was found that the transverse profile data from Vans 8 and 9 was missing. In addition, multiple rut depth data gaps were identified, which might indicate malfunctioning or incorrect on-board settings.

Findings from First Set of Data Collected by All Three Vans

Roughness

The comparison of IRI reports generated by the three ARAN systems led to some unexpected findings. For example, the Van 8 system reported consistently higher IRI values in the left wheelpath as compared with Vans 7 and 9 (Figure 4.1 and Table 4.1), which might indicate malfunction of a RoLine sensor on Van 8.

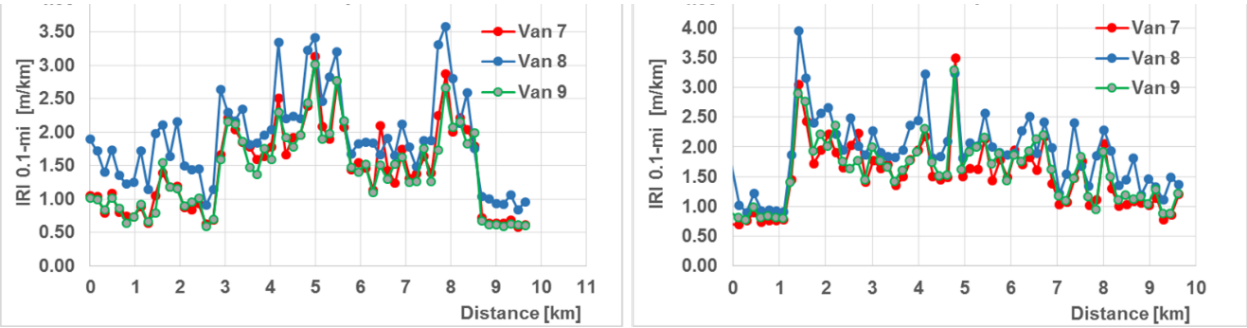


Figure 4.1 Left wheelpath IRI on I-84 Westbound (left) and East bound (Right)

Table 4.1 Summary of IRI on I-84

Direction	I-84 Westbound				I-84 Eastbound			
ARAN ID	Left IRI Mean [m/km]	Left IRI St. Dev. [m/km]	Right IRI Mean [m/km]	Right IRI St. Dev. [m/km]	Left IRI Mean [m/km]	Left IRI St. Dev. [m/km]	Right IRI Mean [m/km]	Right IRI St. Dev. [m/km]
Van 7	1.45	0.64	1.46	0.59	1.51	0.56	1.47	0.61
Van 8	1.90	0.68	1.46	0.54	1.91	0.62	1.42	0.52
Van 9	1.41	0.61	1.41	0.58	1.59	0.54	1.46	0.58

Another finding showed that the RoLine sensors on Vans 8 and 9 were extremely sensitive to change in speed gradient (acceleration and deceleration) as opposed to the Selcom sensors installed on Van 7. One can observe on Figure 4.2 that at speeds lower than 13.4 m/s (30 mi/hr) the IRI reported by Vans 8 and 9 increased three to four orders of magnitude, with Van 9 showing even greater sensitivity than Van 8.

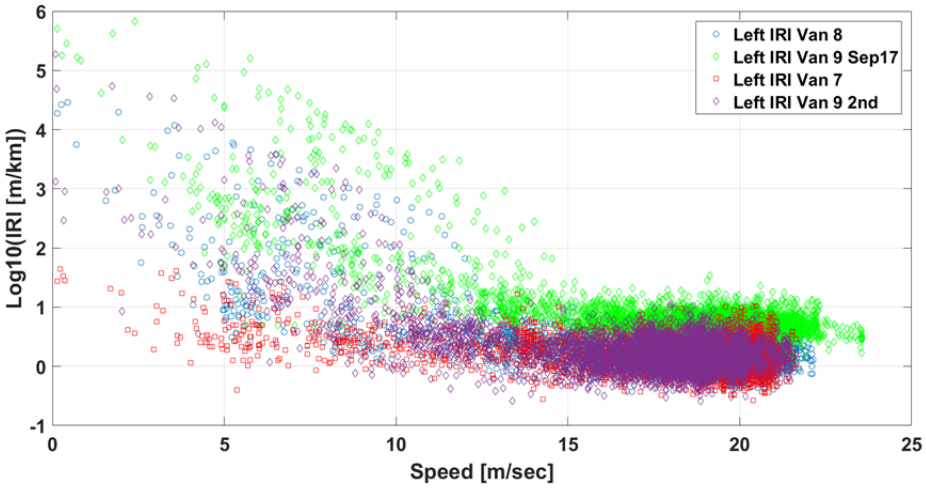


Figure 4.2 Effect of speed on reported IRI on Big Loop Segment B.

Rutting Data

The preliminary analysis of network-level rutting data (mean rut depth per 0.1 mile) showed Van 7 reporting significantly larger rut depths in the left wheelpath (LWP) with no discernable trend in the right wheelpath (RWP) (Figure 4.3). It should be noted that 0.1-mile mean rut depth reported by Vans 8 and 9 did not exceed 6 mm (0.25 in), which, by definition was low to no rutting.

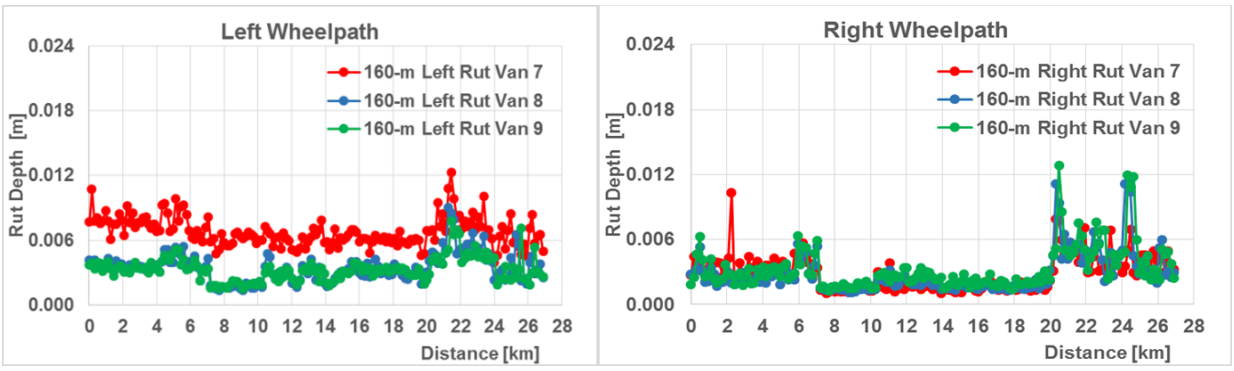


Figure 4.3 Rut depth profile on Big Loop B.

The 5-m average rut depths, which were found to be more representative of the severity level, rated 10 to 15 percent of values as medium severity (12 to 25 mm), as shown in Figure 4.4. Interestingly, Van 7 reported less medium rut depth values than Van 9 did. This effect might be due to differences in rut data collection and calculation methodologies. Note that Van 7 uses ultrasound sensors and collects transverse profiles along a 10-cm transverse interval and at a 5-m longitudinal (vehicle travel) interval. The new 3-D laser system, however, collects transverse profiles with 1-cm resolution in both the transverse and longitudinal directions. Therefore, the Van 7 system might "miss" some deeper ruts. Further investigation of the rut depth database revealed that all three ARAN systems reported values calculated by their respective algorithms, which were quite different (Table 4.2). Lastly, there was a significant amount of wandering (right and left) noted within the travel lane, which likely affected reported ruts for Vans 8 and 9. A combination of the aforementioned issues might explain the very low correlation (less than 60 percent on average) between rutting reported by all three vans.

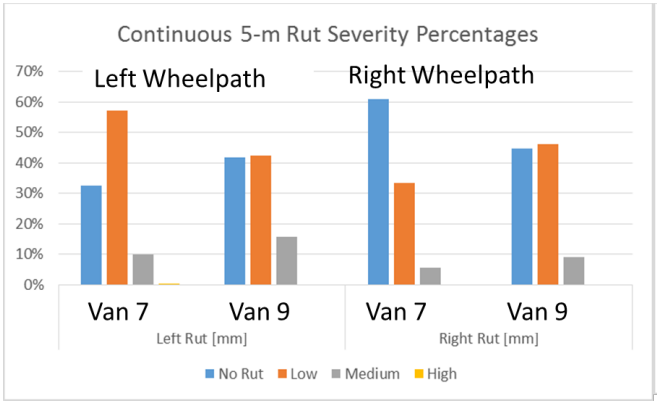


Figure 4.4 Summary of rut depth severities reported by Van 7 and 9.

Table 4.2 Summary of algorithms for rut depth calculation used by ARAN systems.

ARAN ID	Rut Calculation Methods
Van 7	NinePoint_(every ten cm) StraightEdge_FullLane
Van 8	Wire (every 1 cm)
Van 9	StraightEdge_HalfLane (every 1 cm)

Cracking Data

The pavement surface images were collected simultaneously by all three ARAN vehicles. All the images were processed through the newly developed Roadware Vision software by the CTDOT Pavement Management Unit. Unfortunately, Van 7 cracking data was not found suitable for analysis because 99.7 percent of all cracks were reported to be at high severity (Table 4.3). Furthermore, Van 7 reported seven times more cracking as compared with Vans 8 and 9, most likely due to incorrect settings. Vans 8 and 9, however, yielded a very similar distribution of crack lengths by severity. A longitudinal plot of 0.1-mile total crack lengths reported by Vans 8 and 9 (Figure 4.5) indicates potentially good reproducibility between the two newer ARAN vehicles, which contains identical 3-D LCMS systems.

Table 4.3 Summary of crack lengths by severity on Big Loop Segment B.

ARAN ID	Indicator	Severity			Total Sum
		Low	Medium	High	
Van 7	Total Crack Length	240,850	669	209	241,728
	Proportion	99.6%	0.3%	0.1%	100.0%
Van 8	Total Crack Length	19,028	15,504	2,155	36,687
	Proportion	52%	42%	6%	100%
Van 9	Total Crack Length	19,274	14,377	2,355	36,006
	Proportion	53%	40%	7%	100%

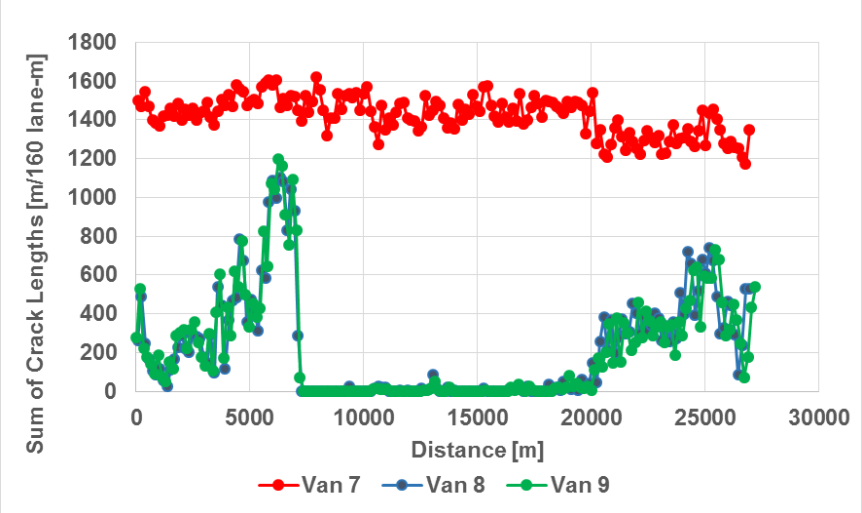


Figure 4.5 Plot of total crack lengths per 0.1 mile on Big Loop Segment B.

In summary, the first round of pavement data collection in 2015 was not very successful due to difficulties experienced with the new hardware and software as well as technical issues with some of the sensors. Therefore, it was decided to repeat the experiment in the summer of 2016 after the vendor completed all necessary repairs and maintenance.

Chapter 5 Analysis of Longitudinal Profiles and Ride Quality Data

The older (Van 7) and newer (Vans 8 and 9) systems collect and report relative elevation for left and right wheelpaths (LWP and RWP, respectively) as computed from height sensors installed on the front bumper of each vehicle. The longitudinal profile of relative elevations is later used to compute continuous IRI values, which are ultimately averaged over 5 m and 160.93 m (0.1 mile) intervals for inclusion in PMS reports that are used for project- and/or network-level analysis. The quality of the longitudinal profile measurement in the end affects the quality of IRI computations. In this project, it was found during the initial analyses that IRI reported by the three vans produced significant discrepancies. This led to further scrutiny of longitudinal profile data for the study.

Longitudinal Profiles

Identifying the differences between ARAN systems

This Chapter focuses on the analysis of the two sets of data collected in 2016. The first run on the Big Loop was performed in June 2016 by all three ARAN vehicles with the objective of direct comparison between data from Van 7 and Vans 8 and 9. It was found, however, that each ARAN vehicle collected longitudinal profile data using different intervals and reported in different units (100.2 mm (SI), 0.984 ft (US customary), and 0.328 ft (US customary)) for Van 7, 8, and 9, respectively. Those differences were attributed to default settings used by the Selcom (Van7) and RoLine sensors (Vans 8 and 9).

Figure 5.1 (top) shows superimposed examples of relative LWP longitudinal profiles collected at posted or prevailing traffic speeds in June 2016. One can easily notice that Van 7 and 8 profiles are very similar, while the Van 9 profile exhibits long wavelength fluctuations (about 90-m or approximately 300-ft), which are generally caused by features in the road profile and are generally filtered out of the final data analysis. Figure 5.1 (bottom) shows the corrected longitudinal profile for Vans 7 and 9 for this section of Big Loop Segment A after removing both short (250-mm) and long (300-ft) wavelengths using a so-called "Roughness Processor" embedded in the Roadware Vision software. The locations of the peaks on both Van 7 and 9 profiles match, albeit some differences in amplitude can still be seen.

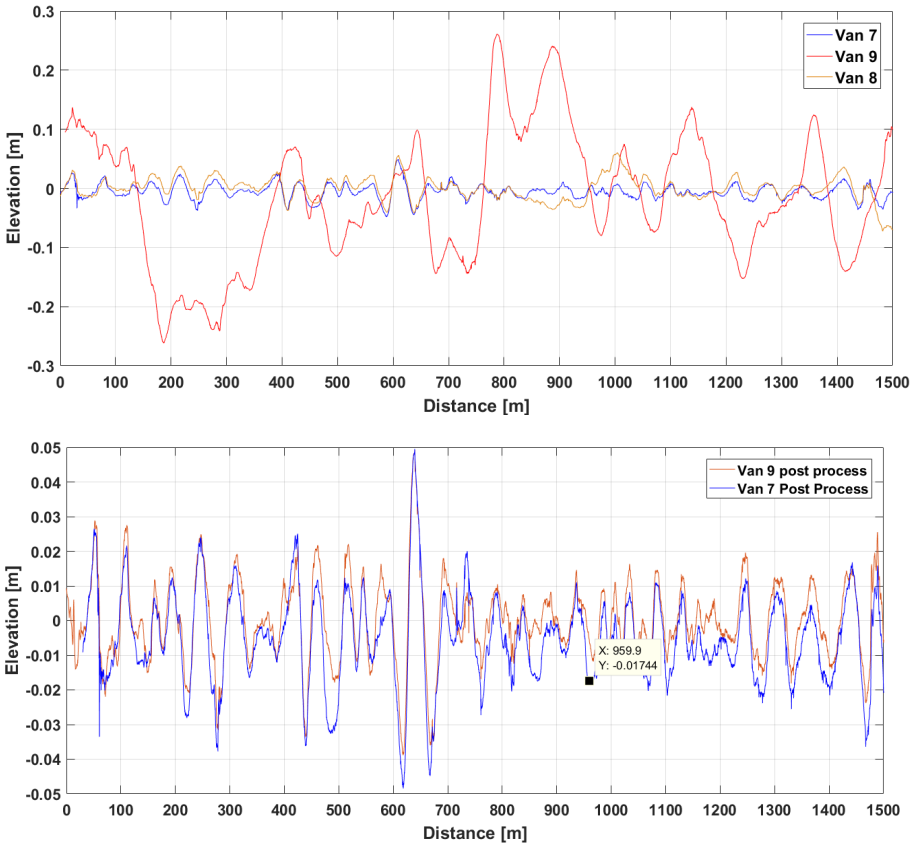


Figure 5.1 Longitudinal LWP profile on a section of Big Loop Segment A before (top) and after (bottom) processing through "Roughness Processor."

Figure 5.2 focuses on longitudinal profiles for Van 8 and 9 at the locations where vehicle slowdowns and accelerations occurred due to obstacles, turns, and stops at intersections. A stop followed by a sharp left turn occurs between 7150 and 7250 m. It is evident from Figure 5.2 that both Van 9 and Van 8 sensors have a much higher sensitivity to speed gradient (~3000- and 200-m range of elevations, respectively) than does Van 7 (~ 1-m range of elevations). Another challenging trend is the Van 8 elevation trend is reversed. Blow-ups on the right side of Figure 5.2 show that even at constant speed, Van 7 maintains about 5 to 10 times lower sensitivity to localized roughness (manhole and railroad crossing at around 7400-m mark) as compared with Vans 8 and 9.

In an attempt to mitigate the issues of stop and go traffic situations, Fugro developed the Roughness post-processor. The laser measuring system paired with the accelerometer on the vehicles work well when the vehicle is traveling at a constant speed. However, it is not possible to always be collecting data at a constant speed in a straight line. The post-processor filters data into two categories. The first category is when the ARAN vehicle is moving at a constant speed and little to no correction is required to account for movement of vehicle due to acceleration/deceleration or turning. The second category is when the vehicle is changing speed or direction and does require the data to be corrected to account for this. The data associated with changes in the vehicle speed or direction is filtered to nullify the large gaps in the data set caused by the change in speed or direction. The post-processing software for these types of situations,

attempts to fill in the data gaps utilizing an integration mechanism that uses more than just one sample to cover the data gaps.

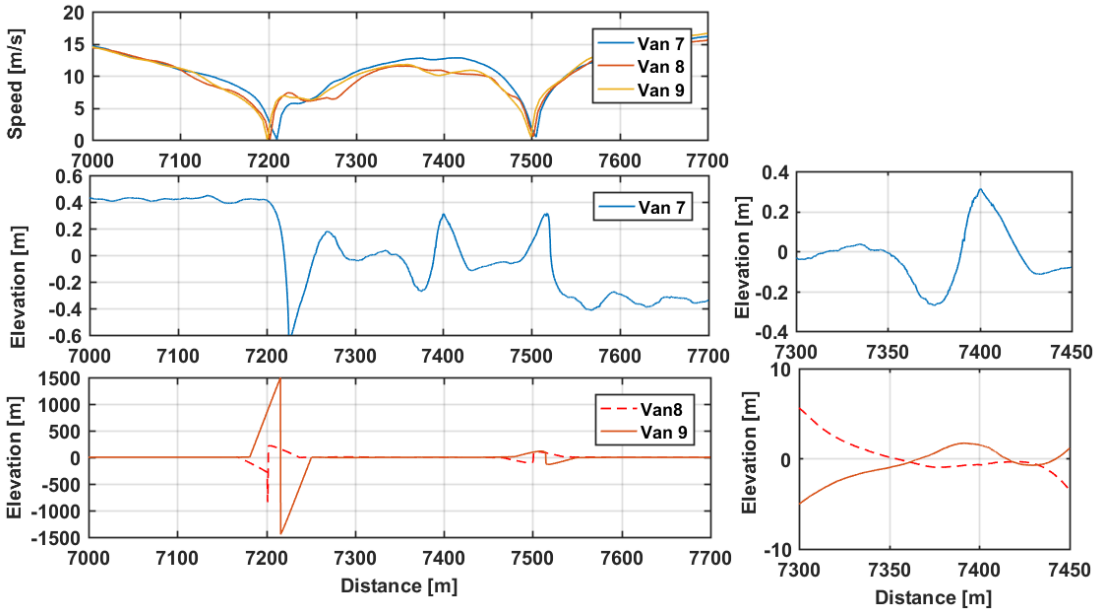


Figure 5.2 Sample of longitudinal LWP profile including slowdowns, obstacles, and accelerations before post-processing.

After processing the raw roughness data for all three vans through the "Roughness Processor," the Vision software produces “improved” longitudinal profiles. The result is reasonable comparability for data from all three vans (Figure 5.3). The exact algorithm behind the "Roughness Processor" is proprietary to Fugro Roadware and it cannot be controlled by the user. Interestingly, this post-processing does not noticeably change the Van 7 original values of elevation, but mostly affects Vans 8 and 9 unreasonably high accelerometer readings.

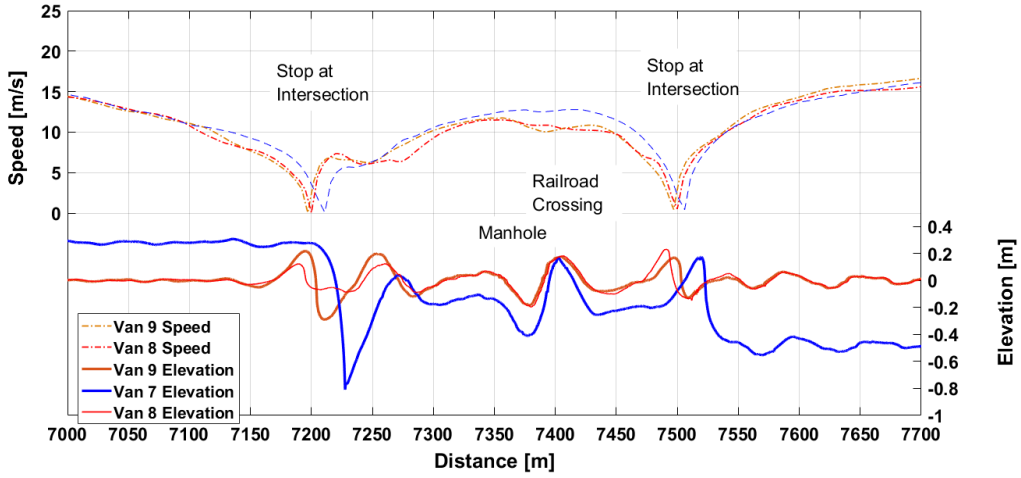


Figure 5.3 Sample of longitudinal LWP profile including slowdowns, obstacles, and accelerations after post-processing.

Analysis of correlations

Due to differences in elevation recording methods/sensor settings, and also specifically due to localized roughness recorded by Van 7 after each stop (Figure 5.3), a direct comparison between longitudinal profiles of Van 7 (Selcom sensors) and Vans 8 and 9 (RoLine) is not possible. Therefore, the "improved" longitudinal profiles (post-processed by Roadware Vision Roughness Processor) were processed through ProVal (TransTec, Inc.) software, which allows for automated synchronization and filtering of the profiles, as well as analysis of correlation. Due to the tremendous amount of profile data collected over the 127 km of the Big Loop (tens of millions of rows), only a 20-km long segment (Big Loop Segment A) is discussed in this report, with assurance that similar trends were exhibited by the ARAN vans on the other segments.

Table 5.1 summarizes parameters of the correlation analysis including overall mean IRI (MRI), Pearson's correlation coefficient *r*, and, slope, and intercept for a pairwise linear relationship between relative profiles in LWP and RWP, separately. The slope represents the relative difference between corresponding sensors measuring the profile, whereas the intercept indicates the bias. Note that for this analysis, elevation is expressed in centimeters. The first attempt at correlation is made over the full 20-km length of Segment A. As expected, due to an exponential increase in elevations during slowdowns and accelerations, a very low correlation exists overall between the three ARAN vehicles. Nevertheless, one can notice higher correlation between the right wheelpath elevations for all three inter-relationships. A significantly higher correlation between Vans 7 and 9 as compared with both Vans' correlation with Van 8 indicated a potential issue with Van 8 profile measurements, most likely at low speeds.

When sections with frequent stops and/or obstacles are excluded, both correlation and the slope values moved closer to a value of one, indicating smaller differences between the three vans. The bias is considered to be negligible since it does not exceed 3 mm. The correlation between all three profiles is very high in the beginning of the segment (0.83 to 0.94 in the first 1.6 km). It decreases for longer sections with variable levels of roughness.

To conclude the discussion of elevation profiles, it was found that post-processing through the Roughness processor improves the longitudinal profile significantly. However, the differences in sensitivity of the ARAN sensors to speed gradient yields an overall lower correlation than a desired 90 percent. While it is possible to correlate short profile sections on an individual basis, it is deemed undesirable on the network level due to time constraints. Note that this project does not target classification of the Fugro ARAN profilers as per ASTM nor the calibration of ARAN distance measuring instrument (DMI), accelerometers, and height sensors. Therefore, the further analysis of profiles is focused on the final product of the PMIS, which contains project- and network-level IRI reporting.

Table 5.1 Summary of correlation analysis for longitudinal profiles on Big Loop Segment A.

Section	Overall MRI (in/mi)	Transfer Function	LWP Elevation			RWP Elevation		
		From /To	r	slope	intercept	r	slope	intercept
Full	136 (Van 7)	Van 7 to 8	0.02	0.018	0.001	0.15	0.175	0.001
	160 (Van 9)	Van 7 to 9	0.41	0.622	0.002	0.56	0.712	0.000
	141 (Van 8)	Van 8 to 9	0.12	0.219	0.002	0.21	0.236	0.001
km 0-1.6	121 (Van 7)	Van 7 to 8	0.94	0.873	0.005	0.92	0.869	0.022
	123 (Van 9)	Van 7 to 9	0.83	0.977	0.031	0.85	0.650	0.017
	128 (Van 8)	Van 8 to 9	0.90	1.135	0.025	0.88	0.711	0.001
km 8-14.5	118 (Van 7)	Van 7 to 8	0.74	0.647	-0.001	0.72	0.642	-0.001
	130 (Van 9)	Van 7 to 9	0.53	0.482	-0.001	0.51	0.446	-0.001
	119 (Van 8)	Van 8 to 9	0.80	0.834	-0.001	0.81	0.794	0.000
km 15.5-19.2	135 (Van 7)	Van 7 to 8	0.76	0.679	-0.001	0.69	0.616	-0.001
	131 (Van 9)	Van 7 to 9	0.43	0.360	-0.002	0.57	0.460	-0.002
	134 (Van 8)	Van 8 to 9	0.56	0.514	-0.001	0.75	0.678	-0.001

Surface Roughness

Surface roughness is a measure of ride quality and, as such, it is one of the most popular pavement performance measures among the U.S. Departments of Transportation (DOTs). Therefore, the analysis of differences in reported ride quality in terms of IRI became the most important and comprehensive part of this project. This section summarizes results of the simultaneous runs for the three ARAN vehicles (vans 7, 8, and 9) in June 2016. A discussion on precision and accuracy of IRI measurements by the new ARANs (Vans 8 and 9) is provided in a separate section named "Precision and Accuracy of Longitudinal Profiles and IRI Report."

Comparison of Post-Processed 5-m IRI Reports (Project Level)

After matching the beginning and the end of each of the five Big Loop Segments (A through E), the 5-m average post-processed IRI values were examined. The post-processing procedure appears to eliminate extremely high reported IRI values that occur within slowdowns and accelerations adjacent to stops. It also eliminates the extreme IRI values reported for areas of localized roughness, such as unadjusted manhole covers, high-severity cracks, and railroad tracks. For Vans 8 and 9, the post-processing mostly resulted in empty data rows at speeds under 30-mi/hr and at the localized roughness locations. As the summary of post-processed data in Table 5.2 shows, Van 7 did not report any empty rows. It was also found that the raw roughness data, hence post-processed IRI data, for Van 7 was only available on Big Loop Segments C and D. Only matched non-empty rows of IRI data could be compared for the vans.

Table 5.2 Summary of post-processed 5-m average IRI data on Big Loop

	ARAN ID		
	Van 7	Van 8	Van 9
Total # of Stations Matched (Length. km)	25,257 (126.3 km)	25,257 (126.3 km)	25,257 (126.3 km)
Total # of Stations Processed	8,152 (40.8 km)	25,257 (126.3 km)	25,257 (126.3 km)
# of Missing Values after Processing (percent missing)	0 (0%)	1,165 (4.6%)	1,109 (4.4%)

Analysis of Distributions of 5-m Average IRI

Figure 5.4 compares distributions of 5-m average IRI values reported by the three ARAN systems using a logarithmic scale. All three systems yield almost perfect shapes of so-called lognorm distributions (normality of logarithmic values) centered around 100 in/mi (logIRI=2). While the distributions for Van 7 and 9 are virtually identical, Van 8 produces skewness toward an IRI range of 300 to 1000 in/mi (logIRI of 2.5 to 3). Further investigation of this trend shows that on the last 40 out 127 km (Big Loop Segment E), Van 8 reported consistently higher IRI values than the other two vehicles. In Figure 5.5, note the offset between centers of distributions (e.g., between means) of about 0.5 logIRI (190 in/mi IRI) on Figure 5.5 (left) and the identity of the three distributions on Figure 5.5 (right) where Segment E is excluded from the analysis. The shown trends were found to apply to both the left and right wheelpaths.

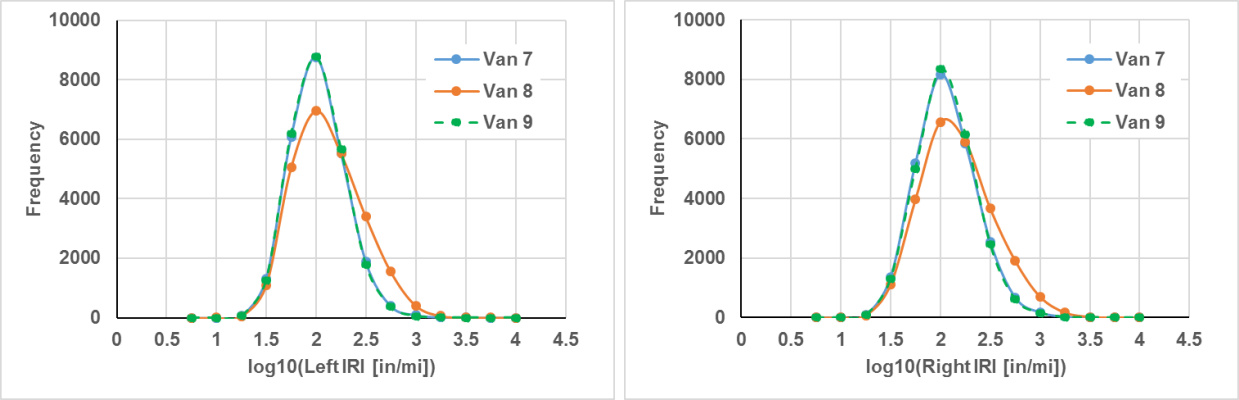


Figure 5.4 Comparison of lognorm distributions of 5-m average IRI values in left (left) and right (right) wheelpaths on the Big Loop

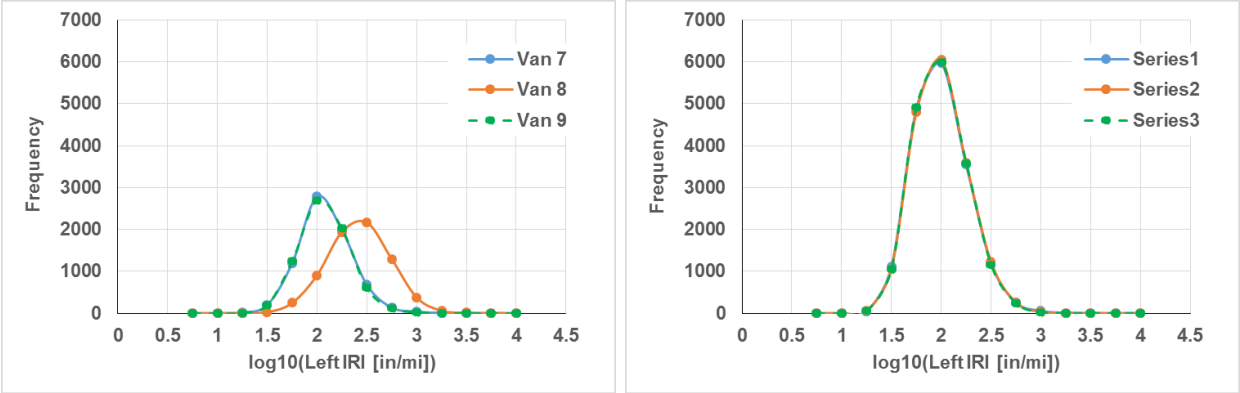


Figure 5.5 Lognorm distributions of 5-m average left wheelpath IRI values on Big Loop E (left) and Big Loop A, B, C, and D combined (right)

Based on the analysis of distributions for 5-m average IRI values, the following conclusions are made:

- For 87 out 127 km of the surveyed sections, the post-processing of the IRI resulted in identical log-norm distributions of measurements for all three ARAN vans. This outcome indicates that measurements with the newer systems (Vans 8 and 9) and older system (Van 7) belong to the same population, and they can be predicted from each other.
- A clear difference in post-processed IRI between Van 8 and the other two vehicles warrants further investigation of the potential factors of influence, such as speed, road condition, and sensor malfunction.

Bivariate Regression Analysis of 5-m Average IRI

Once the normality of logIRI distributions was established, a pairwise linear relationship between IRI datasets for Vans 7, 8, and 9 was investigated to find slope, bias (intercept), and goodness of fit of the pairwise linear models. The first attempt at the regression included all values from Big Loop Segments A, B, C, and D. Figure 5.6 depicts bivariate relationships between the corresponding 5-m average IRI values reported by the three ARAN systems. The linear trend models are interpreted as $y=slope*x+intercept$ where x and y are decimal logarithms of the corresponding IRI values. The slope and intercept values shown in Figure 5.6 indicate that, on average, the relationship between IRI values in the left and right wheelpaths are very similar. With respect to the older ARAN system (Van 7), Vans 8 and 9 yield very similar slopes and intercepts of around 0.6 to 0.7 with significant yet poor correlation. The R-squared of 0.4 indicates that only 40 percent of variability in ARAN measurements can be explained by a log-log relationship. With respect to the equality line, only 29 percent of the data support the direct relationship between the newer and older ARAN systems. On the other hand, Van 9 and Van 8 yield highly correlated IRI values on a logarithmic scale (R-Squared of 0.73 to 0.75) both with and without intercept. Therefore, it can be inferred that on average, Van 8 and 9 measurements are not significantly different.

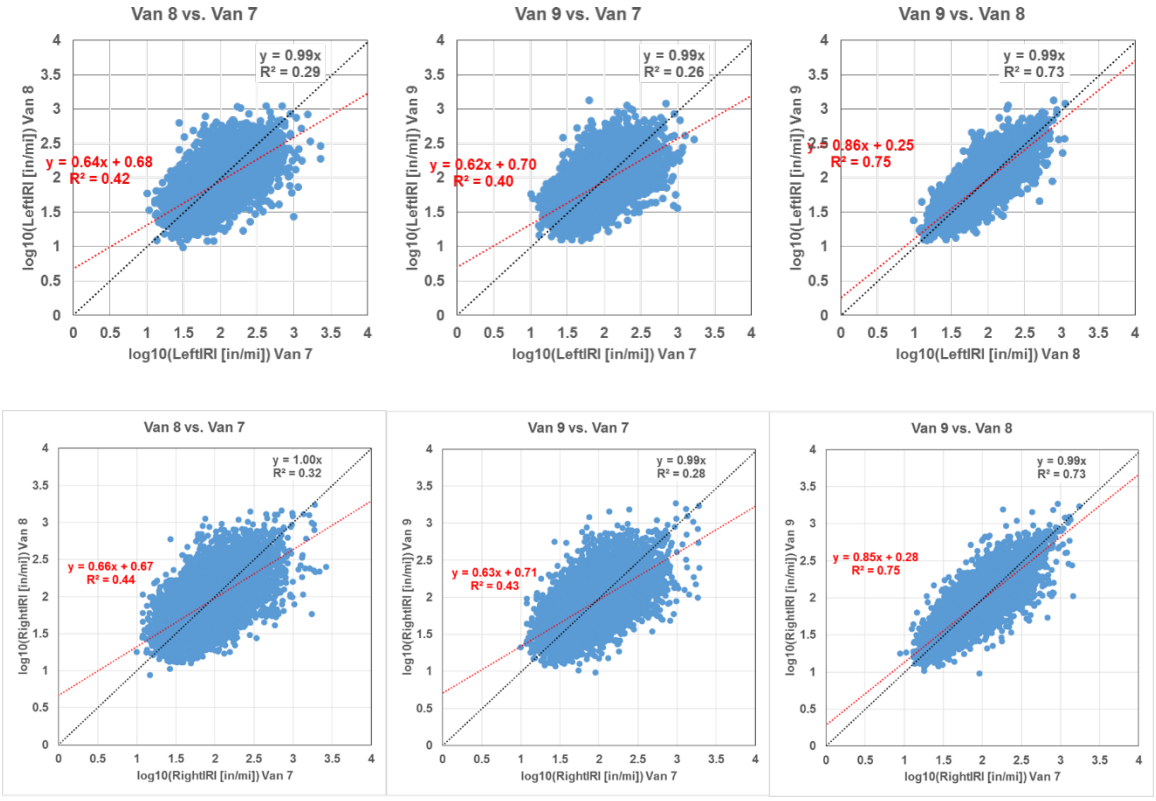


Figure 5.6 Bivariate plots of 5-m average IRI values reported by the three CTDOT ARAN systems.

Table 5.3 summarizes average statistical parameters for the bivariate log-log relationships shown in Figure 5.6. The log-log equations are converted to non-linear equations (power model with multiplicative error (standard log(error)), which in turn, is converted into relative percent error between pairs of ARAN measurements.

Table 5.3 Summary of statistical models and standard errors for bivariate analysis of 5-m average IRI.

From/ to	log-log Equation	Non-linear Equation	R-Sq.	Standard log(Error)	Relative Linear Error
Van 7 to Van 8	Log(IRI8)=0.65log*(IRI7)+0.68	IRI8=4.79*(IRI7)$^{0.65}$	0.43	0.21	63%
Van 7 to Van 9	Log(IRI9)=0.63log*(IRI7)+0.71	IRI9=5.13*(IRI7)$^{0.63}$	0.42	0.21	62%
Van 8 to Van 9	Log(IRI9)=0.86log*(IRI8)+0.27	IRI9=1.86*(IRI7)$^{0.86}$	0.75	0.14	37%

Next, based on earlier reported findings, all corresponding IRI values collected at speeds lower than 30 mi/hr (13.3 m/s) and all IRI values greater than 630 in/mi (10 m/km) were considered to

be potential outliers and were excluded from the analysis. This action did not result in any appreciable improvement of goodness of fit nor produce any noticeable change in slopes and intercepts of the linear trends shown in Figure 5.6. In addition, the analysis of standard errors for the bivariate relationship (Table 5.3) indicated unacceptably large relative standard error of around 60 percent. Finally, it was concluded that 5-m average IRI values could not be used for developing transfer functions between any of the three ARAN systems due to the random nature of differences between reported values. Therefore, only 0.1-mi (160-m) average IRI data was further analyzed to develop transfer functions between older (Van 7) and newer (Vans 8 and 9) ARAN systems.

Development of Transfer Functions for Network-Level IRI Data

This analysis started with the generation of PMS reports for matched 0.1-mile sections using post-processed IRI data. As shown in the previous sections, the final 39 km out of the 127 km Big Loop (i.e., Segment E) showed poor correlation of Van 8 data with the other two vans. Therefore, this segment was excluded from the analysis for development of transfer functions. In addition, the post-processed IRI data on Big Loop Segment A (20 km) could not be recovered from the Van 7 database. Eventually, segments B, C, and D (total length of 68 km) were chosen for development of transfer functions.

Figure 5.7 shows longitudinal 0.1-mi average IRI profile for those sections, left wheel path (top) and right wheel path (bottom), separately. Some notable outlying IRI values were encountered: such as RWP IRI for Van 8 between km 21 and 23 (severe alligator cracking, rutting, and potholes on Rt. 17), LWP and RWP IRI for Van 8 around km 30.2 (stop at intersection Rt.17/Rt.66), and LWP IRI for Van 7 between km 66.0 and 67.0 (series of potholes and overpass joint on ramp from Rt. 2 to Rt. 3)). Figure 5.8 shows example photos at these outlier locations. Thus, it is evident that the post-processing routine only eliminate the outstanding IRI values due to slowdowns and accelerations but does not eliminate extreme IRI values due to localized roughness and sharp turns.

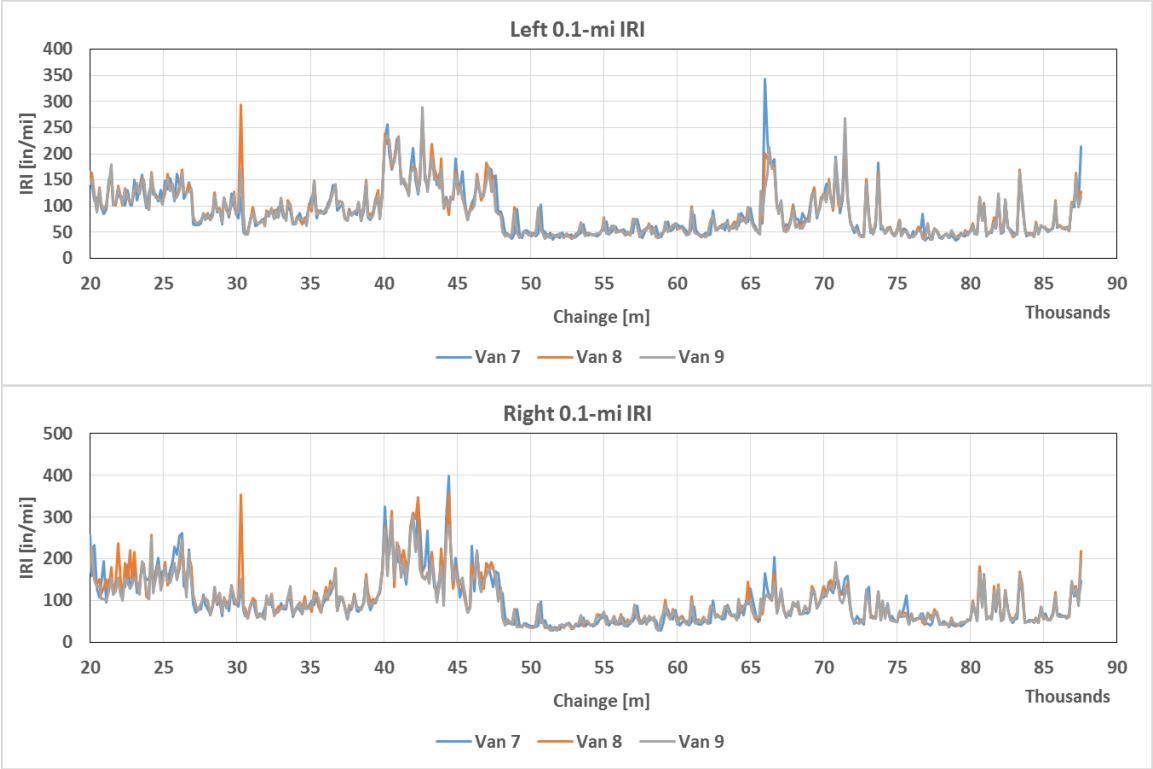


Figure 5.7 Longitudinal 0.1-mi IRI profiles for left (top) and right (bottom) wheelpaths for Big Loop Segments B, C and D.

Figure 5.8 Roadware Vision ROW images of sections with extreme IRI values.

Before the exploration into the development of transfer functions could be undertaken, the relationships between left and right wheelpath IRI were compared. The basic assumption was that, besides the extreme situations as shown in Figure 5.8, there should not be much difference in IRI between the wheelpaths, except right IRI has been traditionally found to be slightly higher than left IRI due to the inherent roadway structure. If this assumption proved true for this studies' dataset, it would allow for elimination of multiple transfer functions, while focusing on mean IRI (MRI) measure, which is usually used by the CTDOT for quality control and acceptance of newly paved roads. The bivariate plots of RWP vs. LWP IRI in Figure 5.9 indicate a relatively small bias of 5 to 10 in/mi toward higher RWP IRI exists, which translates to 9 to 12 percent of relative difference between the two wheelpaths. It can also be seen that the RWP and LWP IRI have a rather nonlinear relationship as indicated by the increased goodness of fit as compared with the linear trend (0.81 vs. 0.6, respectively) (see equations in Figure 5.9).

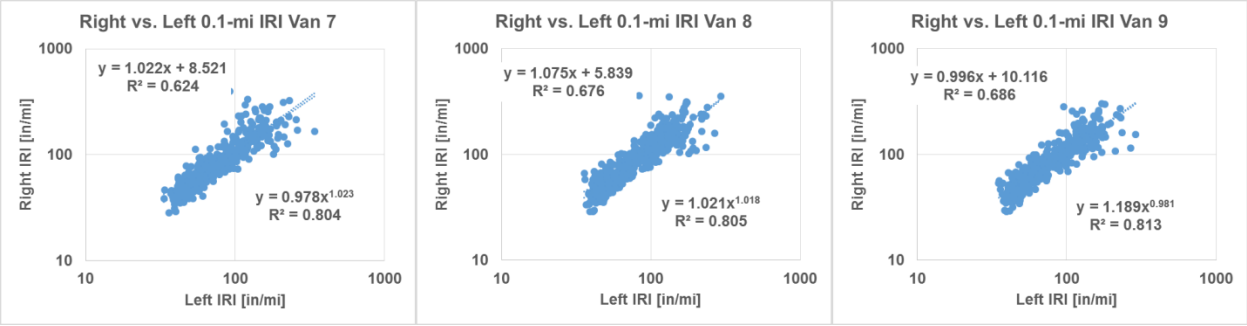


Figure 5.9 Bivariate plots of Right vs. Left Wheelpath 0.1-mi IRI

As shown above, CTDOT ARAN systems yielded a 10-percent relative difference between left and right wheelpath in measured 0.1-mi average IRI values. Such a difference would not exceed 10 in/mi on smooth pavements (IRI<95 in/mi). However, on fair and poor surfaces, it can grow up to 30 in/mi, which is not desired from the standpoint of precision and quality control. Therefore, the preliminary simple regression equations for each wheelpath were developed to understand the significance of differences between the three ARAN systems.

Table 5.4 summarizes slopes and statistical significance (R-squared, standard error) of bivariate relationships for each wheelpath, separately, as well as for MRI. The intercepts are fixed to zero, whereas a slope serves as measure of accuracy and R-squared and standard error indicate relative and absolute precision of the linear trend. Accordingly, one can infer that all three ARAN systems report 0.1-mile IRI at 96 to 99 percent accuracy of each other. The precision of wheelpath IRI between Vans 8 and 9, however, varies significantly, where Van 8 data is less correlated with the other two in both wheelpaths, especially in the left wheelpath. The MRI data appears to be the most precise on an absolute scale (standard error of 7 to 13 in/mi) due to an averaging effect. It should be noted that if at least 95% reliability is required for estimated precision, the standard error value should be doubled, thus reaching 26 in/mi for MRI and about 60 in/mi for individual wheelpaths.

Table 5.4 Summary of preliminary transfer functions between 0.1-mi average IRI datasets (based on June 2016 survey)

	Left IRI			Right IRI			Mean IRI		
From/to ARAN ID	**Slope**	**R-squared**	**Standard Error [in/mi]**	**Slope**	**R-squared**	**Standard Error [in/mi]**	**Slope**	**R-squared**	**Standard Error [in/mi]**
From Van 7 to Van 8	0.97	0.71	23	0.97	0.74	29	1.00	0.93	13
From Van 7 to Van 9	0.96	0.84	17	0.96	0.84	22	0.97	0.93	13
From Van 8 to Van 9	0.99	0.71	23	0.99	0.78	27	0.97	0.97	7

The transfer functions shown above were developed from the data collected in June 2016 by all three ARAN vehicles. Those functions were examined for influencing factors, such as speed, pavement surface rating, age, and functional classification. The importance of each factor was assessed through an analysis of variance (ANOVA) with F-value as a measure of the factor's contribution to variability. In summary, the speed was found to be inter-related with such factors as surface condition and functional classification as it was established to be the only statistically significant influencing factor in explaining the difference between IRI measured by ARAN Van 9 and the other two vehicles.

The relatively low correlation between Vans 8 and 9 led to doubt about the precision and accuracy of the new ARAN systems, which, of course could not be established based on a single run. Therefore, additional multiple runs were conducted to establish repeatability and reproducibility of the IRI measurements from Vans 8 and 9 as discussed in the next Section.

Precision and Accuracy of IRI Measurements (based on November 2016 survey)

The repeatability runs were conducted in November 2016 on three sections of in-service local roads: Brook Street Eastbound and Westbound (~2.7 km each) and Thornbush Road (~ 1 km) all in Rocky Hill, CT. The collection of this data was necessitated by adjustments made by Fugro to the ACS earlier in the year after the first set of data was collected for this study in 2016. Vans 8 and 9 were run simultaneously five times over each section. The precision of IRI measurements was estimated separately in the left and right wheelpaths for each ARAN system. The standard deviation from mean IRI measured over 5 repeated runs for each 0.1-mile section ("within" variation) was used for the analysis. The standard deviation value was reported as absolute precision, whereas the ratio of the standard deviation over mean (coefficient of variation) was reported as relative precision. The absolute and relative accuracy of IRI measurements was estimated as the standard deviation and coefficient of variation, respectively, from the mean of measurements of two Vans for each 0.1-mile section ("between" variation). The aforementioned estimates were made in the absence of ground-truth, or reference values, for the surveyed

segments. Therefore, only reproducibility between Van 8 and Van 9 could be reported. Note that the raw roughness data was processed through the "Roughness Processor" and then, the 5-m IRI data was post-processed to remove measurements at low speeds.

Figures 5.10 through 5.12 show left and right wheelpath 0.1-mile average IRI profiles for each run on Brook St. EB, WB, and Thornbush Rd. respectively. Note that on average, those three segments yielded 115, 105, and 175 in/mi, respectively, indicating relatively good ride quality on Brook St. and fair to poor ride quality on Thornbush Rd. While the corresponding wheelpath profiles for each ARAN dataset follow similar patterns (Figures 5.10 and 5.11), differences between runs can be seen, especially in the beginning and toward the end of the Brook St. sections as well as around station 1120 m on Eastbound, most likely due to slowdowns and accelerations. Figure 5.12 shows much larger differences between runs on Thornbush Road as compared with Brook Street, most likely due to variable fair to poor pavement conditions along this section.

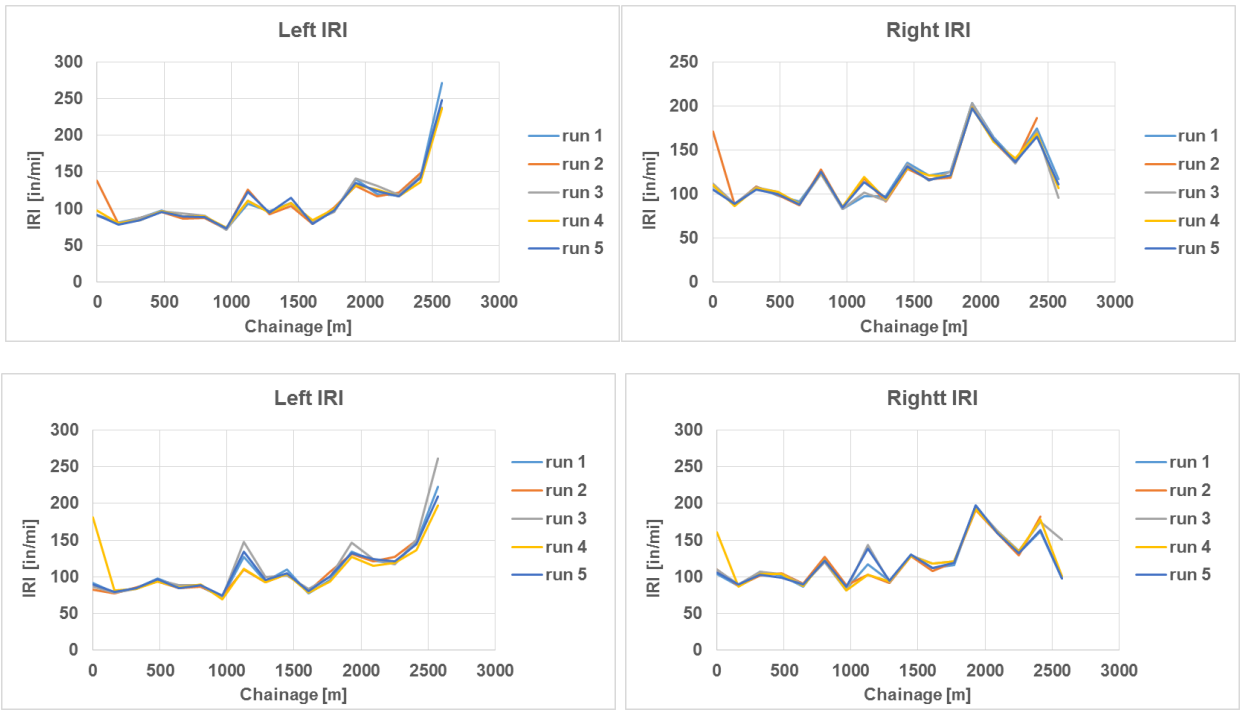
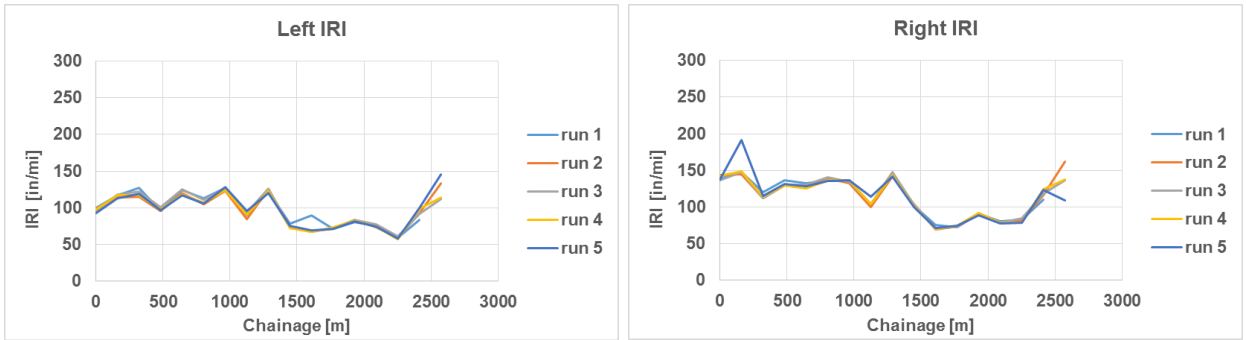


Figure 5.10 Longitudinal profiles of repeatability runs on Brook Street Eastbound by Van 8(top) and Van 9 (bottom).



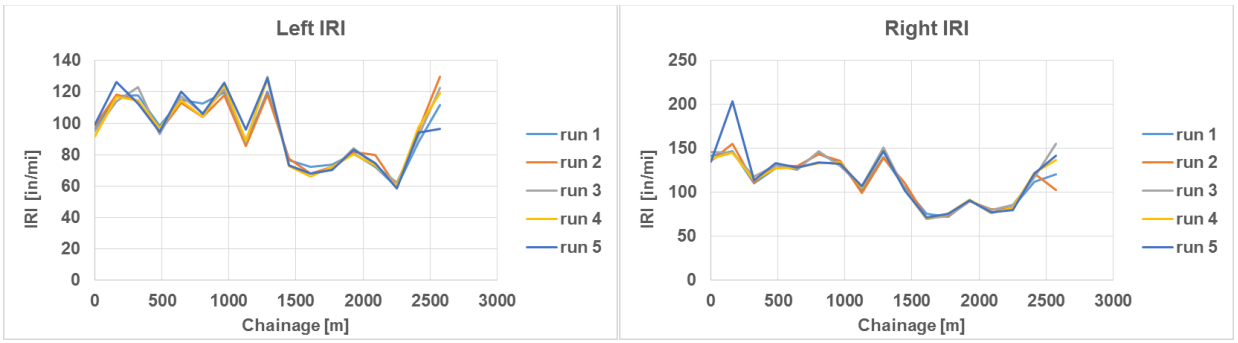


Figure 5.11 Longitudinal profiles of repeatability runs on Brook Street Westbound by Van 8(top) and Van 9 (bottom).

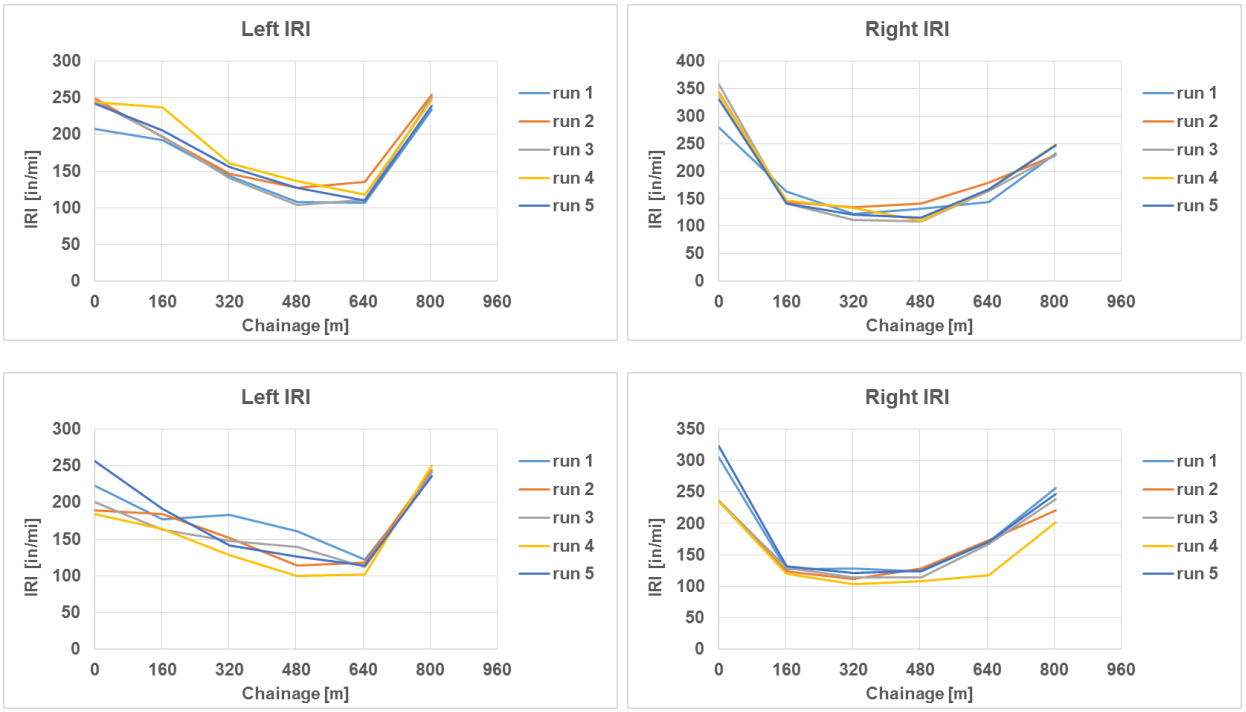


Figure 5.12 Longitudinal profiles of repeatability runs on Thornbush Road by Van 8(top) and Van 9 (bottom).

Table 5.5 summarizes a comparison of precision for Vans 8 and 9 for the three roadway sections shown in Figures 5.10 through 5.12. Note that after low-speed outlier values are removed, each precision value in Table 5.5 is an average of the number of 0.1-mi sections (i.e., average of 14 values for Brook St. and average of 6 values for Thornbush Rd.). Hence, each precision value has a standard deviation of its own. Therefore, a 95-percent confidence value (upper threshold) is computed as the mean plus two times the standard deviation of the standard deviation (mean+2*St.Dev of St.Dev.) and is shown in parentheses in Table 5.5. The absolute precision on segments with good ride quality remains within 3 in/mi (6 in/mi at 95% reliability). The relative precision on those segments is also high enough to pass a usual QC/QA threshold, which is 5

percent. On the other hand, the segment with poor ride quality (Thornbush Rd.) yields deviations up to 12 in/mi (25 in/mi at 95% reliability), which corresponds to relative precision of about 9 percent (18 percent at 95% reliability). It can also be noted that precision in the right wheelpath is lower (i.e., deviation is higher) as compared with the left wheelpath, most likely due to the pavement being sloped to match catch basins and driveway entrances, which is common on two-lane undivided roads.

Table 5.5 Summary of precision of 0.1-mi IRI reported on repeatability runs

Segment	Parameter	Van 8		Van 9	
Brook Street Eastbound	MRI, in/mi	116.1		115.7	
	Absolute Precision (upper 95% CI), in/mi	Left IRI	Right IRI	Left IRI	Right IRI
		2.3 (4.9)	2.3 (5.4)	2.8 (6.1)	2.4 (6.0)
	Relative Precision (upper 95% CI)	2.1% (3.9%)	1.8% (3.4%)	2.6 (4.9%)	2.0% (4.2%)
Brook Street Westbound	MRI, in/mi	105.9		105.0	
	Absolute Precision (upper 95% CI), in/mi	Left IRI	Right IRI	Left IRI	Right IRI
		2.9 (6.9)	2.4 (4.9)	2.6 (4.5)	2.7 (5.3)
	Relative Precision (upper 95% CI)	2.5% (5.1%)	2.75 (7.9%)	2.8% (4.3%)	3.2% (9.3%)
Thornbush Road	MRI, in/mi	183.0		170.0	
	Absolute Precision (upper 95% CI), in/mi	Left IRI	Right IRI	Left IRI	Right IRI
		10.8 (17.2)	10.0 (14.0)	12.4 (24.7)	12.1 (25.8)
	Relative Precision (upper 95% CI)	7.0% (12.0%)	7.0% (11.0%)	9.2% (18.8%)	8.8% (16.7%)

Table 5.6 summarizes reciprocal accuracy of Van 9 and Van 8 on the three sections under discussion. It is easily noticed that absolute accuracy between the two vans on all segments is well within acceptable limits (under 2.5 in/mi [5 in/mi at 95% reliability). The relative repeatability is between 1 and 2 percent (2 and 4 percent at 95% reliability). Thus, it can be inferred that the accuracy of the newer CTDOT ARAN measurements at the network level is higher than their precision.

Table 5.6 Summary of accuracy of 0.1-mi IRI reported on repeatability runs

Segment	Parameter	Left IRI	Right IRI
Brook Street Eastbound MRI=115.9 in/mi	Absolute Accuracy (upper 95% CI), in/mi	0.8 (1.8)	1.0 (2.6)
	Relative Accuracy (upper 95% CI)	0.8% (1.8%)	0.8% (1.8%)
Brook Street Westbound	Absolute Accuracy (upper 95% CI), in/mi	0.9 (2.4)	0.6 (1.5)
	Relative Accuracy (upper 95% CI)	0.9% (2.4%)	0.5% (1.2%)

MRI=105.5 in/mi			
Thornbush Road MRI=176.5 in/mi	Absolute Accuracy (upper 95% CI), in/mi	1.8 (4.1)	2.3 (4.7)
	Relative Accuracy (upper 95% CI)	1.3% (3.3%)	1.6% (3.8%)

Reproducibility of IRI measurements on Big Loop during November 2016 Survey

Due to relatively low correlation between IRI measurements observed on the 127-km Big Loop during the first survey in June, one more run by Van 8 and 9 was conducted in November 2016. Unfortunately, Van 7 had been decommissioned by that time, and, therefore, only accuracy between the newer ARAN systems is analyzed in this section. Figure 5.13 shows bivariate plots of RWP vs. LWP IRI for Vans 8 and 9 (left and right graph respectively). Both vans consistently report RWP IRI being 5 to 10 percent higher than LWP IRI with bias of 5 to 7 in/mi at similarly high correlation (R-Squared of 0.75). Those results are similar to those from June 2016 survey (see Figure 5.9), which validates the assumption that the road conditions did not change significantly during the summer months. This is important when consideration is given to establishing further data quality control checks.

As far as accuracy is concerned, Figure 5.14 presents bivariate plots of the 0.1-mi average IRI measurements in each wheelpath, separately for Van 9 and Van 8. It is noted that the LWP IRI values reported by Van 9 and Van 8 at the network level (0.1-mi average) are identical, whereas RWP IRI reported by Van 9 is within 97 percent of that reported by Van 8. The correlation of the measurements of IRI between the two newer ARAN vans is very strong in both wheelpaths, which indicates vast improvement from the June 2016 measurements (consider R-Squared of 0.96 to 0.99 in November as compared with 0.71 to 0.78 in June for left and right wheelpaths, respectively). Furthermore, the standard errors, i.e. accuracy for November measurements were computed as 7 in/mi and 9 in/mi for left and right wheelpaths, correspondingly (compare with 23 and 27 in/mi for left and right WP, respectively, in June 2016).

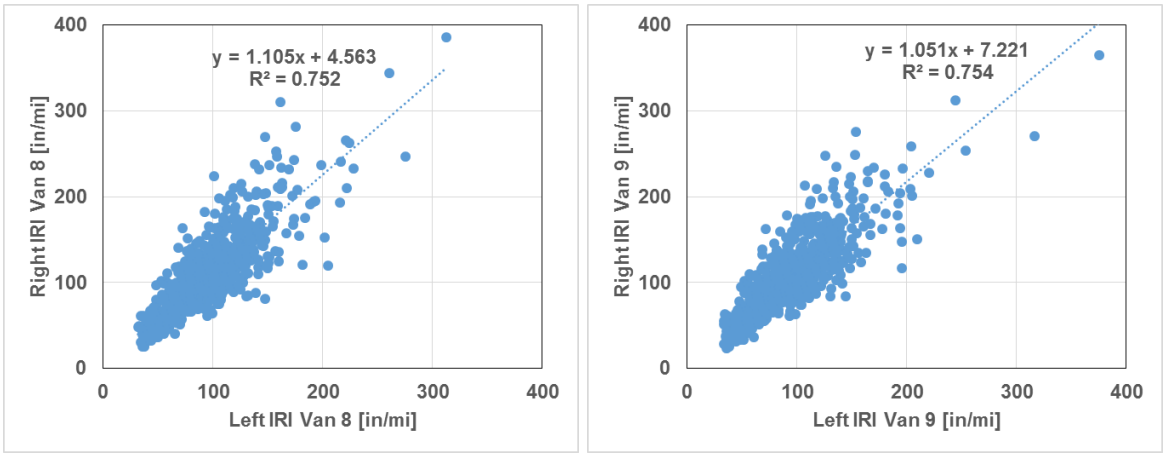


Figure 5.13 Bivariate plots of Right vs. Left average 0.1-mi IRI reported by Van 8 (left) and Van 9 (right) in November 2016.

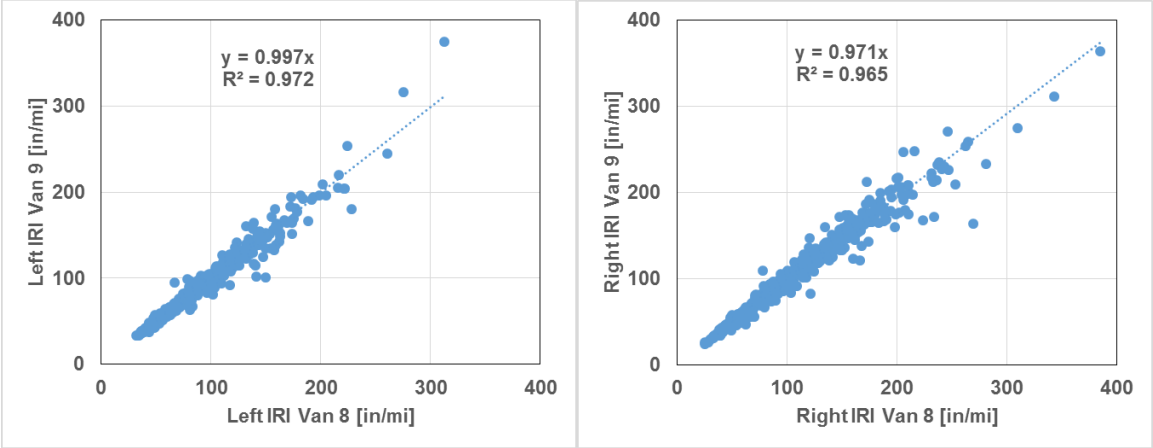


Figure 5.14 Bivariate plot of Van 9 vs. Van 8 average 0.1-mi IRI for Left and Right wheelpaths reported in November 2016.

To conclude the discussion on reproducibility of Vans 9 and 8 with respect to network-level IRI on the Big Loop, Table 5.7 differentiates between accuracy (absolute and relative errors in measurements) overall, at regular average speeds greater than 30 mi/hr, and at low average speed (less than 30 mi/hr). It shows that at 95-percent reliability, one should not expect more than 17-in/mi difference (14%) in 0.1-mi average IRI between Van 9 and Van 8 at regular speeds. However, up to a 35-in/mi (30%) difference at slowdowns and accelerations could be expected. Obviously, those differences may be higher on sections with multiple obstacles and severe surface distresses.

Table 5.7 Summary of reproducibility of network level IRI between new ARAN systems

Speed Level	Parameter	Left IRI	Right IRI
Overall	Absolute Reproducibility (upper 95% CI), in/mi	3 (15)	5 (21)
	Relative Reproducibility	3% (12%)	5% (16%)
Regular (>30 mi/hr)	Absolute Reproducibility (upper 95% CI), in/mi	2 (9)	4 (17)
	Relative Reproducibility	3% (9%)	4% (14%)
Low (<30 mi/hr)	Absolute Reproducibility (upper 95% CI), in/mi	11 (35)	13 (43)
	Relative Reproducibility	9% (26%)	9% (30%)

Chapter 6 Analysis of Transverse Profiles and Rutting

Transverse Data Analysis

Traditionally, pavement management personnel have been using transverse profile data, collected with the CTDOT ARAN systems, to determine rut depth in the left and right wheelpaths, as well as cross-slope (transverse slope). With the transition from the older (Van 7) to the newer vans (Vans 8 and 9), the differences in technology as shown in table 6.1 must be considered:

Table 6.1 CTDOT ARAN Transverse Profile Technology Comparison

Vehicle ID ► / Attribute ▼	CTDOT Van 7	CTDOT Van 8 and 9
Height measurement technique	Ultrasonic 1-D	3-D laser triangulation
Transverse distance interval	100-mm	10-mm
Profile width	2.4 m	4.2 m
Longitudinal data collection interval	5-m	0.1-m

It would be anticipated that the differences in technology shown in Table 6.1 could affect the comparative accuracy between the older and newer ARAN systems. Therefore, the first comparison attempted focuses on identifying the correlation between cross-slope values reported by ARAN Vans 7, 8, and 9 on the same segment/section of roadway. Significant discrepancies between the three ARAN systems in the cross-slope reports were found and, therefore, it was deemed necessary to conduct a detailed analysis of the transverse profile data. The myriad of transverse data collected with Vans 8 and 9 on the Big Loop and on repeatability runs precluded analysis of accuracy for the full length of survey. Instead, transverse data from two 1000-m segments of Big Loop Segment A and Thornbush Road were sampled for accuracy, as discussed below.

Analysis of Cross-Slope Data

In this section, the cross-slope values from the Big Loop segment, reported by Vans 7, 8, and 9, are compared. Note that the cross-slope data from Van 7 came directly from Rut Reports (crossfall [XFALL]) generated by its ARAN system, and was reported in hundred units of slope (100*rise/run). All transverse data from the newer ARAN systems, on the other hand, were processed through Roadware Vision software, which reportedly computes cross-slope in degrees, from the best-fit line of the smoothed profiles. Note that the Vision algorithms are proprietary to Fugro, Inc., and, therefore, are not readily accessible. To allow for direct comparison, the cross-slope data from all three ARAN systems was converted to units of slope (rise to run) before any visual and/or statistical analysis was performed. Additional manipulation was also required in order to match cross-slope data along the surveyed sections. The 5-m average cross-slope data was analyzed first. Next, the average cross-slope for matched 160-m intervals was computed from the 5-m average values, for all three vans. This was done because 0.1-mi (160.93-m) data could not be retrieved directly from the Van 7 SQL databases.

Figure 6.1 shows bivariate correlation plots of the 5-m average cross-slope data for Vans 7, 8, and 9. It can be immediately observed that Van 8 exhibits significant bias of about 1 percent slope against both Vans 7 and 9, whereas only 0.05-percent bias exists between Van7 and 9, which can be ignored. The one percent difference in slopes translates into a 3.6-cm (1.4-in) difference in heights measured over the 3.6-m (12ft) wide lane. As a result, the bias of Van 8 cross-slope data was further investigated as part of the analysis of transverse profiles as discussed in the next section. The analysis indicated a "rotation" of the right side of the transverse profile that was collected with Van 8. This unexpected result likely was due to a misalignment of LCMS lasers.

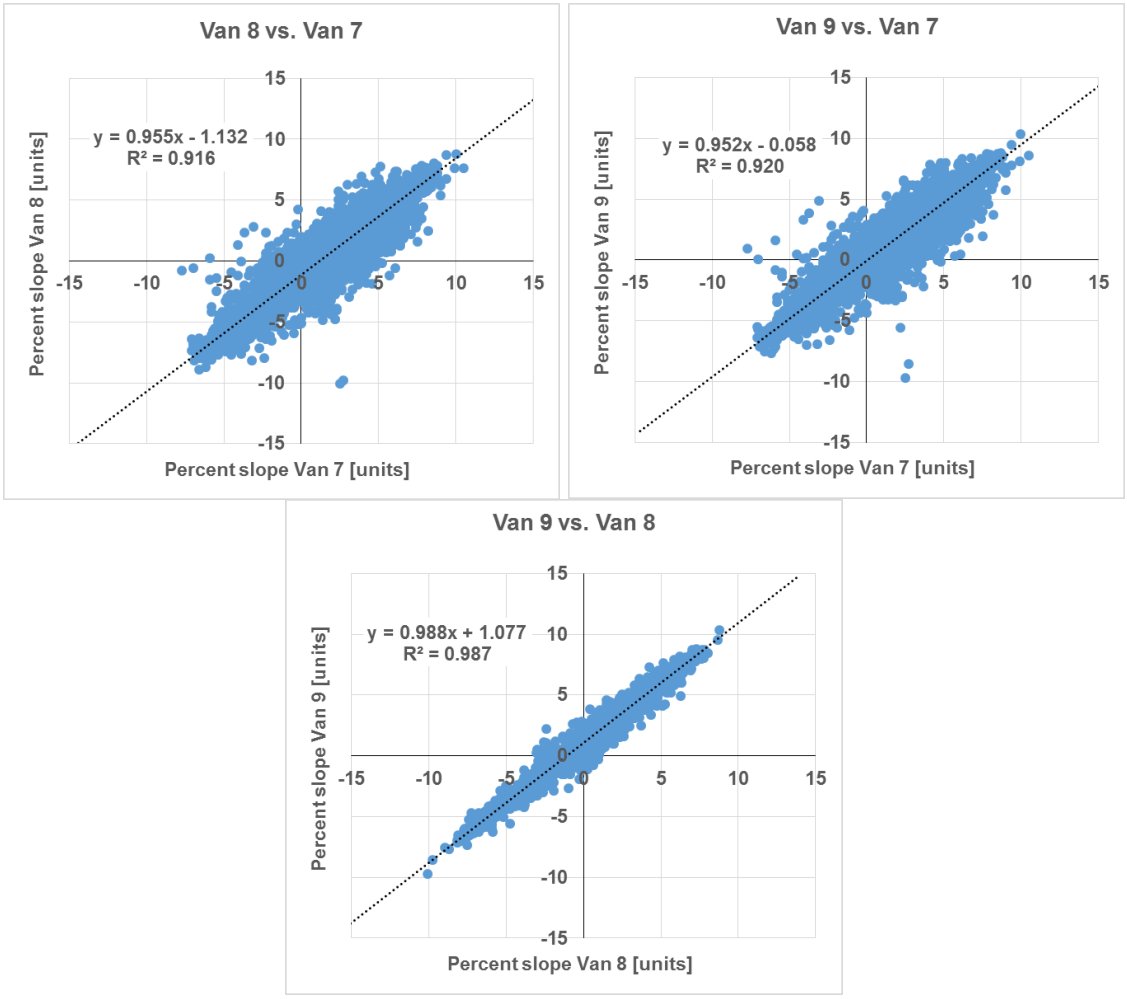


Figure 6.1 Correlation plots for 5-m average cross-slope data on the Big Loop

The slope values of 0.95 result in average 5-percent (0.05 unit slope) accuracy between the newer and older ARAN systems that is well within acceptance levels for quality control. A closer look at the correlation plots in Figure 6.1 reveals a relative spread width of ± 2 units of slope around the fitted line, which results in a relatively low R-squared of 0.92. One underlying reason for that discrepancy may be due to a difference in vertical movement of the ARAN vehicles relative to the pavement surface caused by differences in their chassis suspensions. This difference also led to a variable reaction to roughness of the surveyed pavement surface. As shown in Table 6.2, the

comparative accuracy between the ARAN systems (as measured by R-squared of fitted line and standard error) was higher on smoother pavement segments.

Table 6.2 Summary of comparative accuracy of 5-m average cross-slopes reported by CTDOT ARAN systems

Parameter	From to ARAN ID	Big Loop Segment A	Big Loop Segment B	Big Loop Segment C	Big Loop Segment D	Big Loop Segment E	Overall
Mean IRI (MRI) [in/mi]		123.3	126.5	58.2	78.0	119.0	99.6
Slope of fitted line (from Fig. 6.1)	From Van 7 to Van 8	0.94	0.85	0.99	0.97	0.87	0.92
	From Van 7 to Van 9	0.95	0.86	0.99	0.96	0.88	0.99
	From Van 8 to Van 9	0.98	0.98	1.00	0.99	0.98	0.92
Standard error (unit slope)	From Van 7 to Van 8	0.56	0.92	0.31	0.54	0.86	0.75
	From Van 7 to Van 9	0.50	1.05	0.30	0.51	0.82	0.72
	From Van 8 to Van 9	0.31	0.32	0.18	0.21	0.34	0.29

Transfer Functions for Network-Level Cross-Slope Data from PMS Reports

As might be expected, the analysis of correlations between network-level (160-m average) cross-slope data showed better correlation and much lower standard errors than for the 5-m average cross slopes. This is due to the effect of averaging. It still confirmed significant bias, or systematic error, of Van 8 data relative to the other two ARAN systems. This is indicated by the intercepts of the fitted lines in Table 6.3. Thus, it was concluded that only cross-slope data from Van 9 was reliable and could be correlated with historical cross-slope data contained within the CTDOT PMIS, which had been collected with Van 7. A special mention is made here that reference ground-truth values of cross-slope were not available for this project.

Table 6.3 Summary of correlation between cross-slope data reported by CTDOT ARAN systems on the Big Loop

From/to ARAN ID	Slope	Intercept (bias)	R-Squared	Standard Error (unit slope)
From Van 8 to Van 7	1.002	1.215	0.98	0.343
From Van 9 to Van 7	1.008	0.315	0.98	0.309
From Van 8 to Van 9	1.002	-1.072	1.00	0.126

Sampling Transverse Profiles

To investigate in more detail the differences in collected transverse profiles between the three ARAN systems, a 1-km section of Big Loop Segment A was chosen. On this section, 10,000 profiles were collected with an interval of 0.1 m by Vans 8 and 9 (LCMS), whereas Van 7 collected 200 profiles with a 5-m interval. As noted in Table 6.1, Van 7 surveyed a 2.4-m (~ 8-ft) width, whereas Vans 8 and 9 reported transverse data for a 4.2-m (~14-ft) width. Therefore, some adjustments were warranted to ensure comparability of transverse datasets, including: matching the locations of the profiles by GPS coordinates; smoothing of LCMS profiles using a moving average; and offsetting vehicle transverse wandering within the travel lane.

Transverse Profiles Issues Between Vans 8 and 9

Figure 6.2 shows an example of as-reported transverse profiles on Big Loop Segment A (at chainage 100 m) before any adjustment. As determined from GPS coordinates, the centers of profiles are matched within 2 cm in the longitudinal direction, and with transverse offsets of 4 cm between Vans 7 and 8, and 20 cm between Vans 8 and 9. Approximate lines of slopes are plotted in Figure 6.2 to illustrate an obvious rotation of the Van 8 transverse profile by about 1% (or 1 degree) with respect to Van 9. One very noticeable difference between Van 7 and the other two profiles is that Van 7 shows a possible rut in the left wheelpath between offsets negative 0.7 and 1.0, whereas Van 8 and 9 show rut on the opposite side of the road within the same positive offset values from the center of profile.

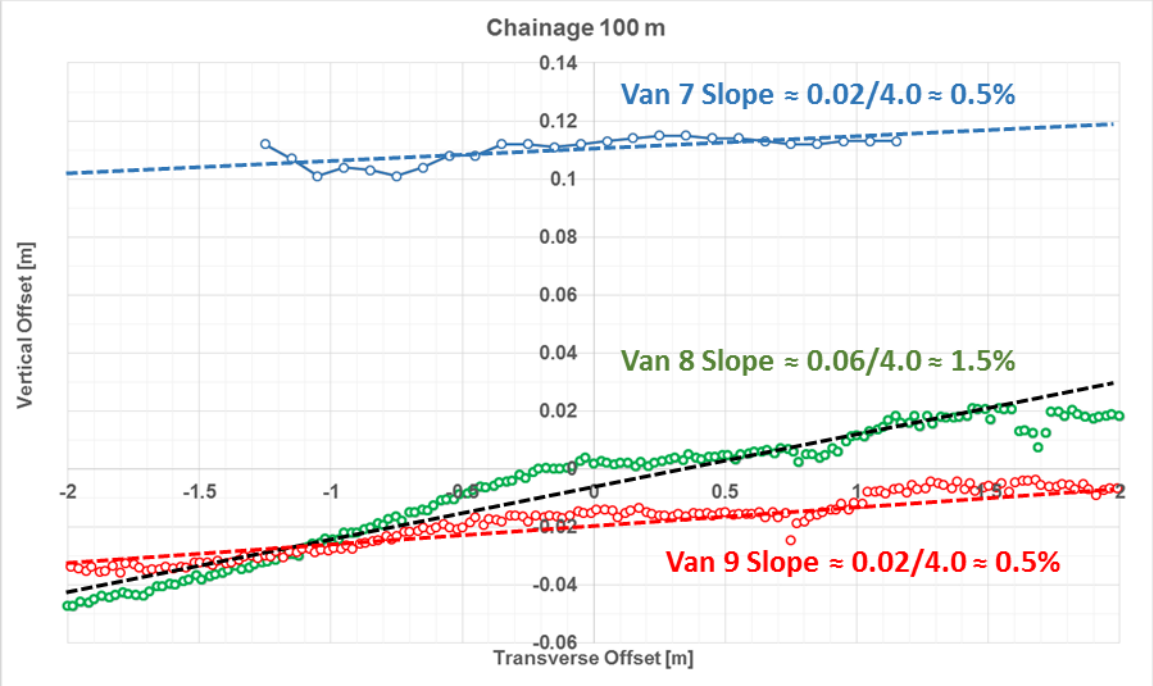


Figure 6.2 Examples of as-reported transverse profiles collected with three CTDOT ARAN systems

Further visual analysis of transverse profiles shown in the Roadware Vision interface suggested that all the transverse coordinates of the profile points are reversed, that is left and right wheelpaths are flipped. Note that, while the crack in Figure 6.3 appears close to the center of the lane (offset of -0.1 m), the profile chart displays the steep slope toward the curb on the left side, although the curb is actually located on the right side of the road.

More issues were found when the Transverse Profile Explorer option was used. As shown in Figure 6.4, this option allows viewing both collected and "smoothed" profiles with cross-slope values reported on the right side of the screen. Careful examination of those values revealed that the slopes from Transverse Profile Explorer are half as steep as that reported by the PMS. The smoothed profile appears to be correct. However, no smoothed (or "corrected") profile data could be accessed in the version of Roadware Vision available to the research team.

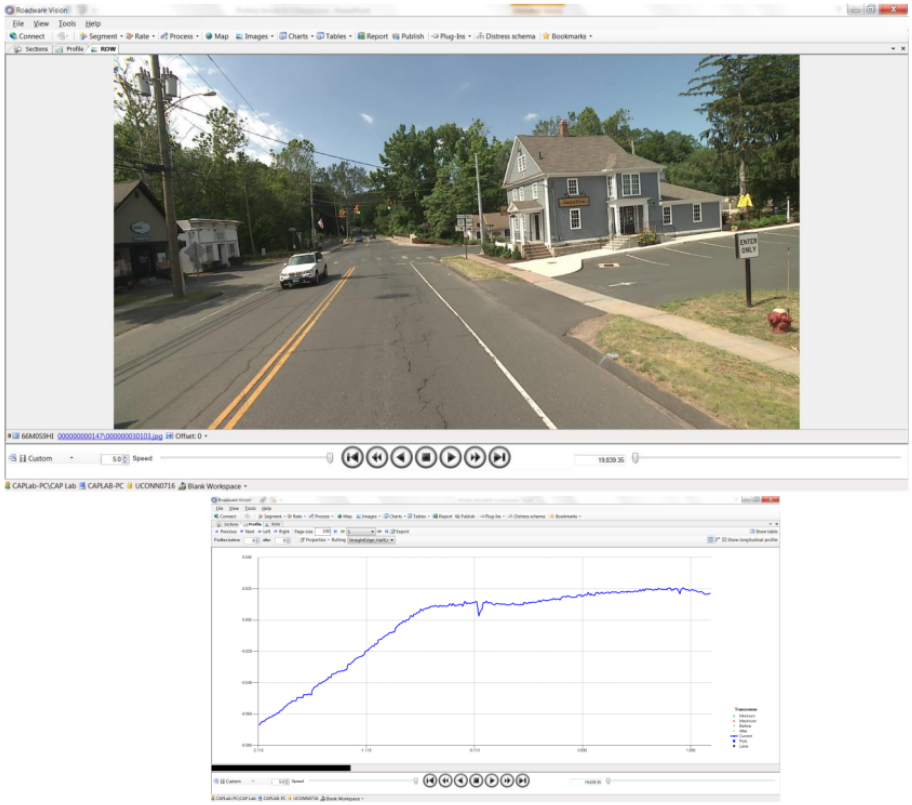


Figure 6.3 Screenshot comparison of right-of-way image and profile chart from Roadware Vision Profile viewer interface.

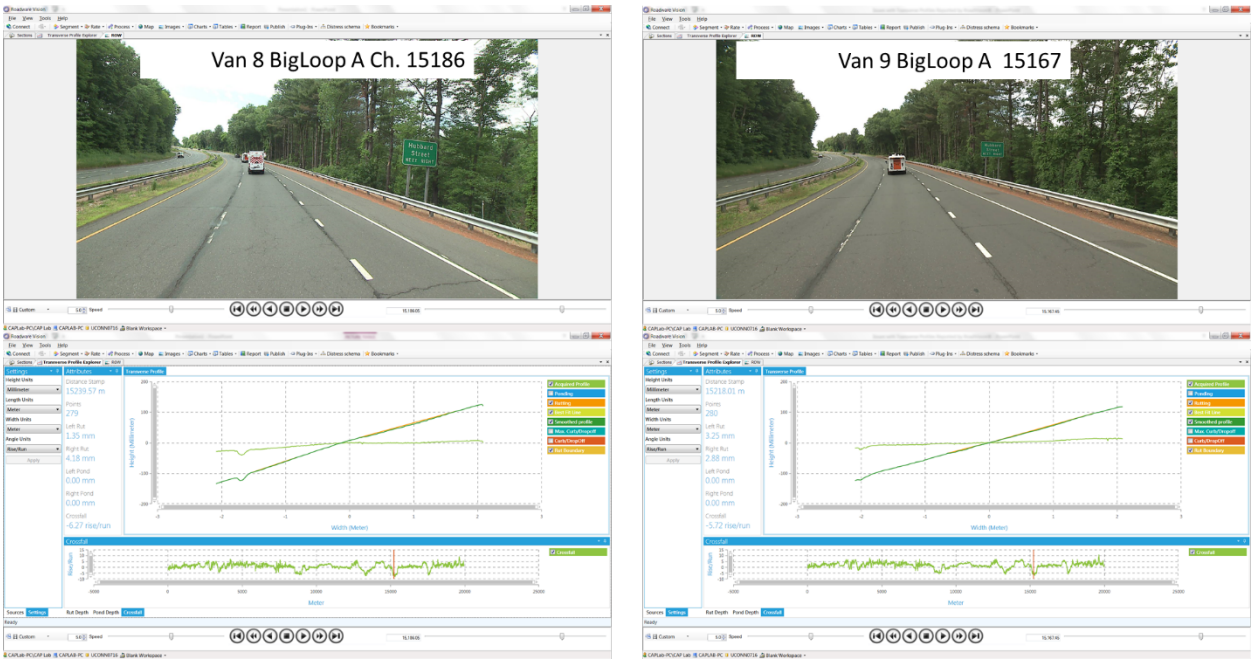


Figure 6.4 Comparison of screenshots from Transverse Profile Explorer

Cross-Correlation of Transverse Profiles

While no correlation between Van 7 transverse data and the corresponding data from Vans 8 and 9 could be established, the new ARAN systems would be expected to yield highly correlated if not identical results with respect to transverse profile elevations. To test this hypothesis, 10,000 profiles from the 1-km long section of Big Loop Segment A were analyzed for cross-correlation. The analysis compared fitted lines (slopes and intercepts) for the matched transverse profiles reported by the two new ARAN systems. The fitted lines were computed for the whole width as well as for the left and right half-width separately, as depicted on the diagram in Figure 6.5. Once determined, the best-fit line slope values were analyzed for correlation between corresponding values from the Van 8 and 9 datasets. Note that the full-width values were expected to be comparable to the cross-slopes in the PMS reports.

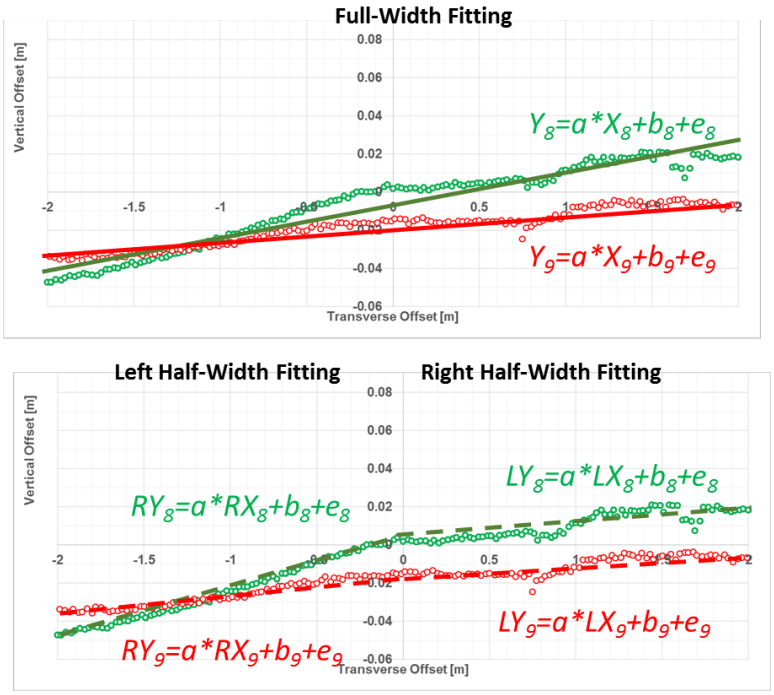


Figure 6.5 Diagram of line fitting for transverse profile data reported by LCMS

Figure 6.6 shows bivariate correlation plots for left half (top left), right half (top right), and full-width profile slopes. One trend worth noting is the difference in ranges of slopes reported by the ARAN systems. The intercepts of fitted line for left and right halves of the profile indicate consistent bias of negative 1% and 0.7% slope for Van 9 in the left and right half-widths, respectively, as well as similar slope of 0.86 to 0.87. This confirms the hypothesis that Van 8 reports consistently "rotated" transverse profiles through the whole width. The small bias of 0.2% at 0.55 slope indicates that, on average, Van 8 reports half as steep transverse slope as Van 9 does. All this leads to an assumption that the left and right LCMS lasers may not be properly aligned on Van 8.

Also notable is that the slopes in the left half range between 0 and 4 percent, whereas in the right half they do not exceed positive 1 percent slope. This trend is counterintuitive, because the road is expected to be sloped toward the right side shoulder or curb, except of course on sections containing superelevation.

Considering all the aforementioned issues and trends, the transverse profile data by itself did not appear reliable and, therefore, the cross-slope values reported by all three ARAN vans were not deemed to be potentially suspect. In the future, a reference ground truth slope on some control sections should be surveyed via geodetic equipment, and transverse profile algorithms that are embedded in Roadware Vision, should be checked and calibrated to the reference data.

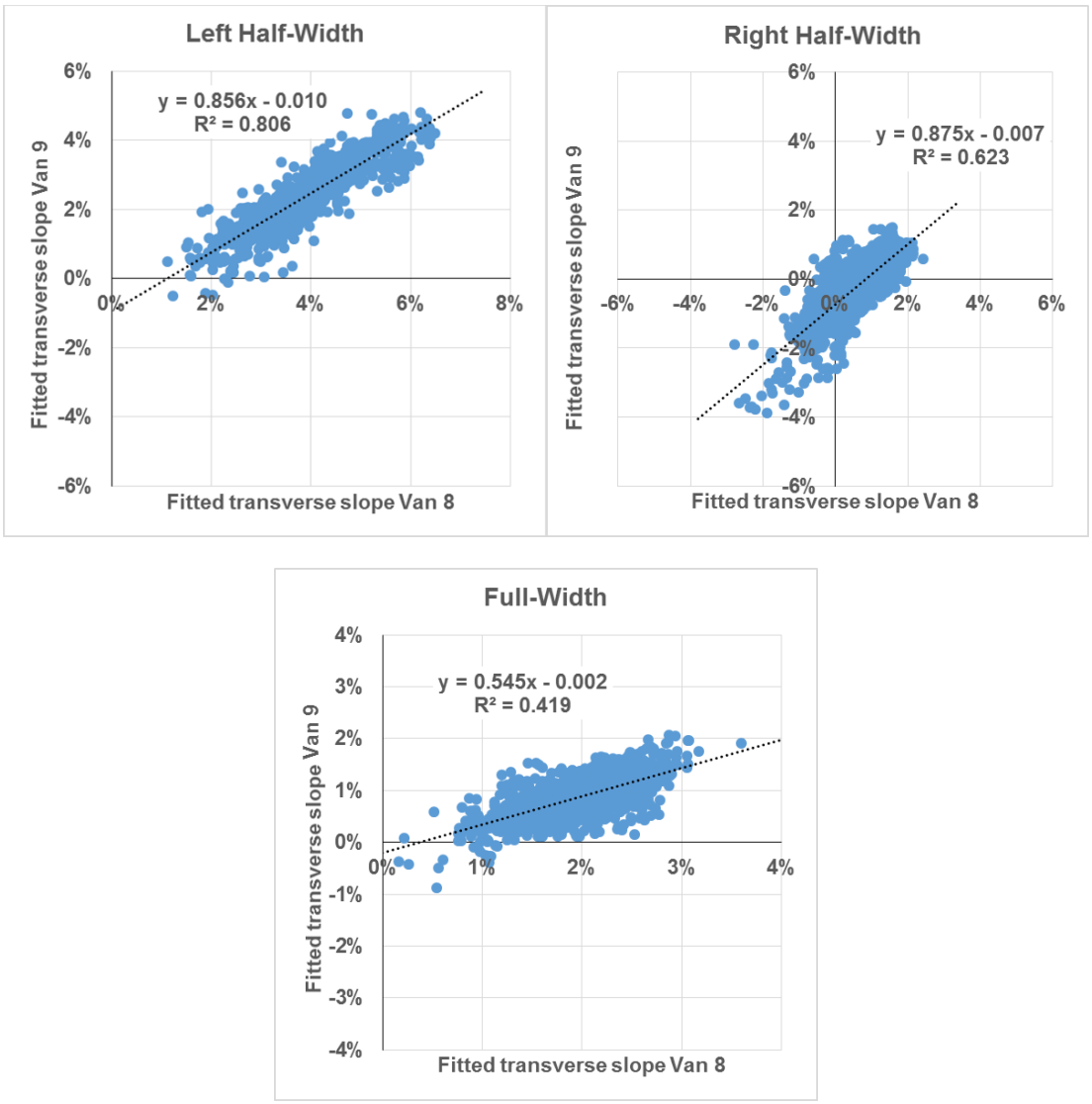


Figure 6.6 Correlation plots of fitted transverse slopes from ARAN datasets

In summary, with respect to transverse profile data and cross-slope reports, the following is recommended:

- Reporting cross-slope data from processed transverse profiles through Roadware Vision should be abandoned. Instead, crossfall (XFALL) values from vehicle orientation (VEHICLEORIENTATION) databases should be used for PMS reports since all three ARANS report virtually identical XFALL values with very high reliability of at least 97 percent.
- Cross-slope units in future PMIS reports should be checked for consistency. Units of slope (100 times rise over run), as in historical PMIS reports, are preferable units of measure.
- The LCMS equipment on both new ARAN vehicles (Van 8 and 9) should be checked for alignment to avoid misalignment of transverse profiles and, therefore, misreporting or misclassifying rutting and cracking.

Rutting Data Analysis

For a number of reasons, the rut depth values reported by the newer ARAN systems (Vans 8 and 9) are not easily comparable with those reported by Van 7. First, the Van 7 system measures and reports discrete rut depth values in millimeters at 5-m intervals along the surveyed road, whereas the new Pavemetrics system computes an average rut depth value for the desired interval. Thus, for a 5-m length, 500 values are collected and averaged, (each extracted from a transverse profile collected with 0.1-m interval). Second, Van 7 collects transverse profiles over a 2.4 m (~8ft) lane width resulting in 25 readings, whereas Pavemetrics' resolution is 0.025 to 0.01 m, which results in between 170 and 400 readings. Those two fundamental differences in longitudinal and transverse resolution make matching profiles a challenging task, as discussed earlier in this chapter. Furthermore, over-smoothing of profiles by the new ARAN system leads to (1) underestimating the overall rut severity and (2) very poor correlation with historically collected data.

The original goal of this project was to develop transfer functions for translating newly collected data into historically reported rutting data. Therefore, the first attempt at analysis was made by direct comparison of average rut depth values reported by the three ARAN systems. The Van 7 data was retrieved from *RutReport files generated by the on-board Van 7 system, whereas for Vans 8 and 9, rut depths were extracted from PMS reports generated by Roadware Vision. All three systems utilized the half-lane straightedge (HLSE) algorithm to compute rut depths as shown in Figure 6.7. Before comparison, all rut depth values were converted to inches.

Analysis of Distributions

Figure 6.7 compares sample distributions of 5-m average rut depths on Big Loop Segments A and B for the left and right wheelpaths separately. While all three systems report normally distributed rut depth values, Van 7 data exhibits noticeable skewness toward medium and high rutting (> 0.5 in [>12 mm]) in the left wheelpath, while showing identical or near identical rut depths with Vans 8 and 9 in the right wheelpath. The distributions of Vans 8 and 9 values are centered around 0.2 in (5 mm) in both wheelpaths for the whole length of survey whereas Van 7 reports consistently higher values in the left path, which centered around 0.4 in (10 mm).

Analysis of Correlations

As reported in Table 6.4 for Big Loop Segments A and B, at the project level (5-m average), very poor correlation was established between the older (Van7) and newer (Vans 8 and 9) ARAN systems. On average, Vans 8 and 9 reported 0.12-in (3-mm) deeper ruts as compared with Van 7. The correlation between Vans 8 and 9, although being significantly better than with Van 7, was found to be poorer than expected (R-squared between 0.5 and 0.7), especially in the right wheelpath. For the moment, it is believed that fundamental differences in measurement and in computation algorithms for transverse profiles are the main reasons for poor correlation between rut values reported with 5-m intervals. Note that despite the very low correlation, all R-squared values reported herein were found to be statistically significant with a 99-percent reliability (p-value <0.01). The rest of the discussion of rutting data is focused on the network-level analysis, which considers 0.1-mi (160-m) average values of rut depth.

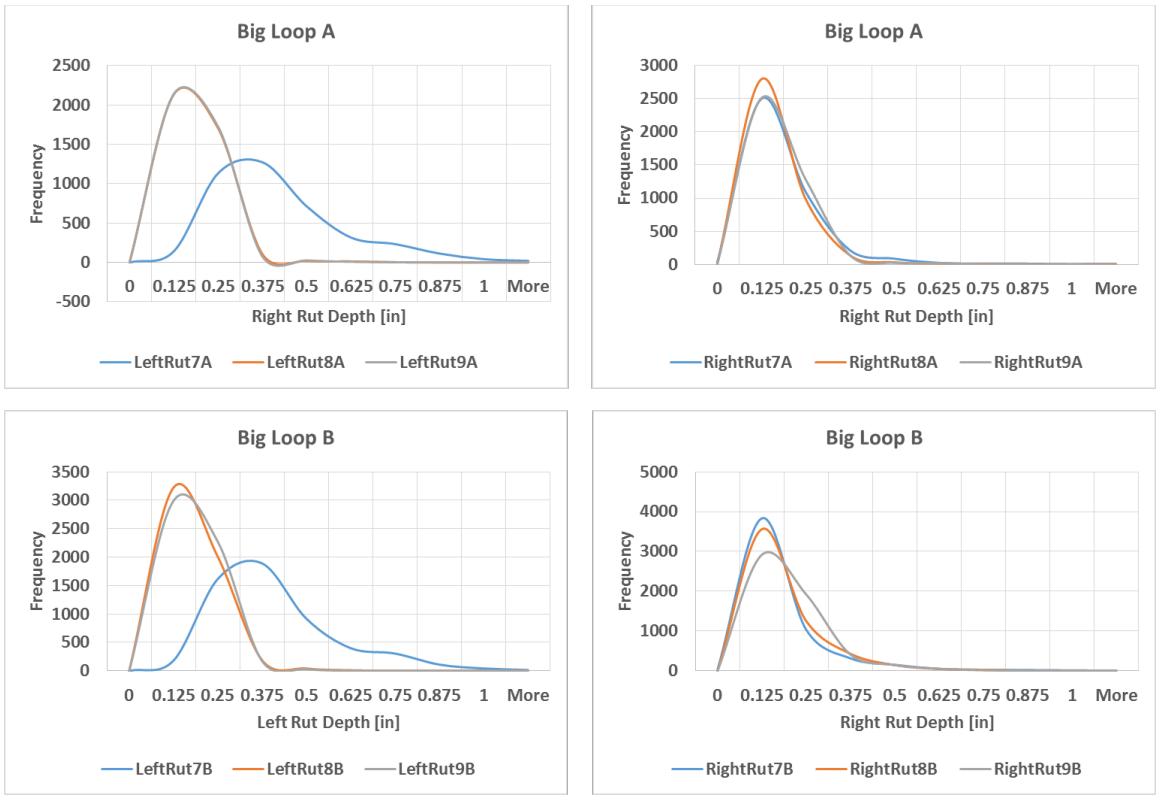


Figure 6.7 Sample histograms of 5-m average rut depth values reported by CTDOT ARAN systems.

Table 6.4 Summary of bivariate correlations between ARAN-reported 5-m average rut depths.

Section ID	Wheelpath	From/to ARAN ID	Slope	Intercept [in]	R-Sq.
Big Loop Segment A	Left	From Van 7 to Van 8	0.05	0.12	0.024
		From Van 7 to Van 9	0.05	0.11	0.028
		From Van 8 to Van 9	0.77	0.03	0.650
	Right	From Van 7 to Van 8	0.28	0.10	0.075
		From Van 7 to Van 9	0.28	0.11	0.108
		From Van 8 to Van 9	0.46	0.07	0.316
Big Loop Segment B	Left	From Van 7 to Van 8	0.06	0.11	0.028
		From Van 7 to Van 9	0.04	0.12	0.020
		From Van 8 to Van 9	0.78	0.03	0.722
	Right	From Van 7 to Van 8	0.38	0.11	0.119
		From Van 7 to Van 9	0.36	0.13	0.131
		From Van 8 to Van 9	0.66	0.06	0.516

Attempt at Transfer Functions for Network Level Rut Depth Data

Table 6.5 summarizes statistical parameters for transfer functions between the rut depth datasets for the three ARAN systems. It appears that averaging the rut depths over 160 m (0.1 mi) significantly improves the correlation for all interrelationships. The values of standard errors between the newer (Vans 8 and 9) and older ARAN systems shown in Table 6.5 range from between 0.02 to 0.07 in (0.5 to 2 mm). Considering the mean rut depth of 0.4 in (10 mm), those errors translate to 5 to 20 percent on average, which brings them to 10 to 40 percent at 95-percent reliability (double standard error). Thus, the linear relationships shown below would likely prove to be of not much use to PMS personnel. While the ground truth reference for rut depths has not been yet established, it seems that Van 7 reports rutting levels that are closer to reality than the two newer ARAN vehicles. This can be stated because ultrasonic radar reports real-time ruts every 5 meter, whereas, the Pavemetrics 3-D system averages about 50 values over a 5-m interval.

Table 6.5 Summary of correlation between 160-m average rut depths reported by the ARAN systems

Section ID	Wheelpath	From/to ARAN ID	Slope	Intercept [in]	R-Sq.	Standard Error [in]
Big Loop Segment A	Left	From Van 7 to Van 8	0.18	0.08	0.106	0.07
		From Van 7 to Van 9	0.17	0.08	0.106	0.07
		From Van 8 to Van 9	0.88	0.01	0.837	0.02
	Right	From Van 7 to Van 8	0.66	0.08	0.257	0.04
		From Van 7 to Van 9	0.72	0.08	0.429	0.03
		From Van 8 to Van 9	0.59	0.05	0.490	0.04
Big Loop Segment B	Left	From Van 7 to Van 8	0.48	-0.05	0.535	0.06
		From Van 7 to Van 9	0.20	0.07	0.102	0.07
		From Van 8 to Van 9	0.90	0.02	0.891	0.01
	Right	From Van 7 to Van 8	0.67	0.05	0.184	0.02
		From Van 7 to Van 9	1.00	0.09	0.533	0.04
		From Van 8 to Van 9	0.81	0.04	0.627	0.03

Precision and Accuracy of Rut Depth Measurement with Pavemetrics System

Regardless of the degree of correlation that exists between the historical (ultrasonic RutBar) and newly implemented technology (3D LCMS), it is important to determine the precision of accuracy of the new system in order to establish repeatability and reproducibility of rut depth measurements. This section discusses precision as "within" variation between five repeatability runs. In the absence of ground-truth or reference values, the accuracy as measured using the data generated from Van 8 and 9 is reported herein as "between" variation between the means of five repeatability runs. The results on precision and accuracy of rut depth measurements are given separately for project (5-m average) and network (0.1-mi [160-m]) levels.

Precision and Accuracy of Project-Level Rut Measurements

Five repeatability runs were conducted on Thornbush Road (Southbound) and Brook Street (Eastbound and Westbound). Tables 6.6 through 6.8 summarize precision analysis for Thornbush Rd. Brook St. EB and Brook St. WB, respectively. The summary includes ranges of reported values as well as absolute (standard deviation in inches) and relative (ratio of standard deviation over mean) values of precision and accuracy. One can notice overall low to medium rut depths on all three test sections. The absolute precision ranges between 0.02 and 0.04 in (0.5 to 1 mm), which translates to 13 to 27 percent of mean rut value measured over 5 runs (32 to 65 percent at 95-percent confidence). This level of precision is deemed acceptable.

Table 6.6 Precision of 5-m average rut depths reported by ARAN systems on Thornbush Road

Statistics		Van 8		Van 9	
		Left Rut [in]	Right Rut [in]	Left Rut [in]	Right Rut [in]
Range of Values		0.07 - 0.56	0.06 - 0.58	0.07 - 0.62	0.06 - 0.45
Absolute Precision	Mean	0.02	0.04	0.03	0.04
	95% Confidence	0.08	0.11	0.07	0.13
Relative Precision	Mean	13%	22%	15%	23%
	95% Confidence	32%	65%	34%	55%

Table 6.7 Precision of 5-m average rut depths reported by ARAN systems on Brook Street EB

Statistics		Van 8		Van 9	
		Left Rut [in]	Right Rut [in]	Left Rut [in]	Right Rut [in]
Range of Values		0.07 - 0.28	0.05 - 0.27	0.06 - 0.26	0.05 - 0.27
Absolute Precision	Mean	0.02	0.02	0.02	0.03
	95% Confidence	0.06	0.07	0.07	0.08
Relative Precision	Mean	13%	17%	19%	27%
	95% Confidence	34%	50%	45%	56%

Table 6.8 Precision of 5-m average rut depths reported by ARAN systems on Brook Street WB

Statistics		Van 8		Van 9	
		Left Rut [in]	Right Rut [in]	Left Rut [in]	Right Rut [in]
Range of Values		0.06 - 0.26	0.04 - 0.93	0.05 - 0.27	0.05 - 0.60
Absolute Precision	Mean	0.02	0.02	0.02	0.03
	95% Confidence	0.05	0.09	0.06	0.08
Relative Precision	Mean	14%	17%	18%	22%
	95% Confidence	36%	46%	46%	47%

The absolute and relative reproducibility values between Van 8 and Van 9 are shown in Table 6.9. The absolute accuracy ranges between 0.01 and 0.02 in (0.2 to 0.5 mm) and does not exceed 1.5 mm with 95-percent confidence. Noticeably, the accuracy in the left wheelpath is higher than in the right wheelpath. In addition, it is noticed that accuracy of rut measurements is slightly higher than precision (the variation is lower) as was also shown to be the case with IRI measurements. The precision may be degraded due to operational factors (e.g., vehicle wander within the lane, speed variations).

Table 6.9 Summary of reproducibility between Van 8 and 9 of rut depth measurements on repeatability runs

Statistics		Thornbush Road		Brook Street	
		Left Rut [in]	Right Rut [in]	Left Rut [in]	Right Rut [in]
Range of Values		0.07 - 0.59	0.07 - 0.51	0.06-0.21	0.05-0.17
Absolute Accuracy	Mean	0.01	0.02	0.01	0.02
	95% Confidence	0.02	0.06	0.03	0.04
Relative Accuracy	Mean	5%	12%	10%	16%
	95% Confidence	13%	31%	23%	36%

As a result of the low correlation found for IRI (as reported in another section of this report) and technical issues with on-board ARAN systems, additional field runs by Vans 8 and 9 were conducted on the Big Loop in November 2016. This provided the opportunity to re-evaluate reproducibility of rut measurements at the project level using the two surveys in June and November 2016. It can be inferred from Table 6.10 that despite the slight increase in rutting, which is expected over a few months of road service, the reproducibility between Van 8 and 9 of rut surveys did not change. Indeed, it stayed within 0.02 in (0.07 in at 95% confidence), or 0.5 mm (1.8 mm at 95% confidence).

Table 6.10 Summary of reproducibility between Van 8 and 9 of rut depth measurements on Big Loop.

Statistics		June 2016 Survey		November 2016 Survey	
		Left Rut [in]	Right Rut [in]	Left Rut [in]	Right Rut [in]
Range of Values		0.04- 0.83	0.03 - 0.97	0.05 - 0.95	0.04 - 1.42
Absolute Accuracy	Mean	0.01	0.02	0.01	0.01
	95% Confidence	0.04	0.07	0.04	0.07
Relative Accuracy	Mean	9%	13%	9%	12%
	95% Confidence	25%	39%	24%	33%

Precision and Accuracy of Network-Level Rutting Reports

The network-level rutting is computed by averaging rut depth measurements over a 0.1-mi length, that is, by averaging 1,600 values. Due to the Pavemetrics System over-smoothing both the longitudinal and transverse profile, the computation likely leads to reporting lower rut severities than actually exist. Nevertheless, 0.1-mi average rut depth is an important performance measure included in PMS reports and it affects pavement condition rating. Therefore, it was of interest to investigate repeatability (precision) and reproducibility (reciprocal accuracy) of rutting reports at the network level.

Table 6.11 summarizes precision of 0.1-mi average rut depth reports on Thornbush Road and Brook Street. Note that 95-percent confidence values are reported in parentheses. On average, very shallow rut depths of less than 0.2 in. (5 mm) are reported for all 0.1-mile (160.9-m) segments. It can be observed that both Van 8 and Van 9 showed high precision within 0.05 in. (1.25 mm) at 95 % confidence with the variation in the left wheelpath being lower than that in the right wheelpath (6 to 8% in LWP against 8 to 16% in RWP). Lastly, Van 9 shows higher variability in RWP rut depths as compared with Van 8.

The results of reproducibility at the network level (Table 6.12) indicate an overall high accuracy of rut depth reports for the LWP, within 0.03 in (0.8 mm) and within 0.06 in. (1.5 mm) at 95-percent confidence. Note that variability in the right wheelpath remains higher than that in the left wheelpath.

The discussion on accuracy of network-level rut depth reporting is concluded with a comparison of reproducibility between Van 8 and Van 9 as determined from two surveys in June and November 2016 (see Table 6.13). At the overall low rutting level reported (max 0.5 in [13mm]), the absolute accuracy did not change between the two surveys, while remaining at 0.01 in (0.03-in at 95% confidence) on average.

Table 6.11 Summary of precision of 0.1-mi average rut depth reports.

Statistics		Van 8		Van 9	
		Left Rut	Right Rut	Left Rut	Right Rut
Thornbush Road	Range of Value, in.	0.11 - 0.22	0.10 - 0.29	0.11 - 0.21	0.11 - 0.31
	Absolute Precision (95% upper CI), in.	0.01 (0.02)	0.02 (0.05)	0.01 (0.03)	0.01 (0.05)
	Relative Precision (95% upper CI)	6% (12%)	14% (32%)	6% (14%)	16% (27%)
Brook Street EB	Range of Values, in.	0.10 - 0.18	0.07 - 0.13	0.09 - 0.15	0.08 - 0.15
	Absolute Precision (95% upper CI), in.	0.008 (0.016)	0.006 (0.0016)	0.01 (0.02)	0.01 (0.03)
	Relative Precision (95% upper CI)	6% (12%)	7% (18%)	8% (14%)	14% (25%)
Brook Street WB	Range of Values, in.	0.08 - 0.12	0.06 - 0.14	0.06 - 0.10	0.09 - 0.19
	Absolute Precision (95% upper CI), in.	0.008 (0.016)	0.006 (0.0016)	0.01 (0.02)	0.01 (0.03)
	Relative Precision (95% upper CI)	7% (13%)	8% (16%)	8% (17%)	12% (23%)

Table 6.12 Summary of accuracy between Van 8 and 9 of 0.1-m rut depth reports

Statistics		Thornbush Road		Brook Street	
		Left Rut [in]	Right Rut [in]	Left Rut [in]	Right Rut [in]
Range of Values		0.11 - 0.22	0.10 - 0.33	0.06-0.17	0.06-0.18
Absolute Accuracy	Mean	0.01	0.02	0.01	0.01
	95% Confidence	0.02	0.06	0.02	0.03
Relative Accuracy	Mean	4%	11%	9%	15%
	95% Confidence	11%	31%	19%	33%

Table 6.13 Summary of reproducibility between Van 8 and 9 of network-level rut reports on Big Loop.

Statistics		June 2016 Survey		November 2016 Survey	
		Left Rut [in]	Right Rut [in]	Left Rut [in]	Right Rut [in]
Range of Values		0.06 - 0.33	0.05 - 0.50	0.06 - 0.36	0.05 - 0.50
Absolute Accuracy	Mean	0.006	0.011	0.008	0.01
	95% Confidence	0.02	0.04	0.02	0.03
Relative Accuracy	Mean	5%	9%	7%	8%
	95% Confidence	14%	27%	16%	22%

Chapter 7 Analysis of Pavement Surface Images and Cracking Data

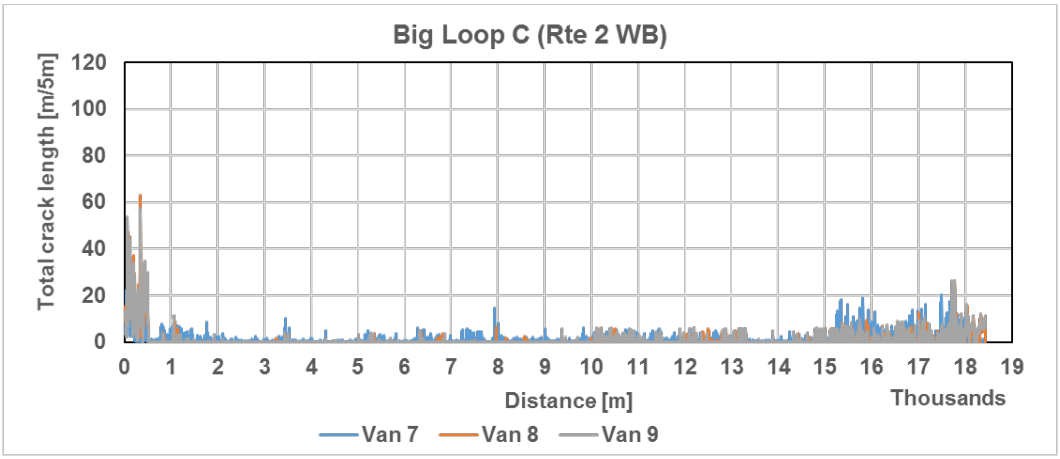
The original goal for the analysis of cracking data was to develop transfer functions between historically-detected cracking obtained from Strobe video cameras on Van 7 and cracking detected with the 3-D LCMS systems installed on Van 8 and Van 9. The first attempt at direct comparison of cracking datasets using the 2015 Big Loop survey was unsuccessful due to issues with processing Van 7 data through Roadware Vision software (see Chapter 4). The second survey of the Big Loop in June 2016, however, resulted in approximately 40 lane-km (out of 127 km surveyed) of comparable pavement surface images suitable for processing and analysis of cracking data using all three ARAN vans. In addition, the precision and accuracy of the cracking survey was evaluated using five repeatability runs from Vans 8 and 9 on Thornbush Road and Brook Street. This chapter first discusses the attempt at direct comparison of cracking data reported by the older and newer ARAN systems. Next, the differences in detection capability of image processing software are analyzed. Lastly, efforts on developing transfer functions between the various cracking datasets are described.

Evaluation of PMS Reports

The first step in making a direct comparison of ARAN cracking datasets included matching the limits of survey and generating PMS reports for project (5-m interval) and network (160-m interval) levels of detail. The PMS reports for Van 7 were generated with WiseCrax software, whereas reports for Vans 8 and 9 were created using Roadware Vision templates. The Roadware Vision units of measurements were converted from feet to meters prior to the comparison.

Figure 7.1 shows ARAN cracking profiles in m/5m (meters of crack length per 5 meters of pavement length) for Big Loop Segments C and D (upper and lower graphs respectively). Note that Segment C contains a portion of Route 2 Westbound that was mostly repaved in 2015 and, thus, has little or no cracking. Segment D, on the other hand, exhibits a significant amount of cracking on Route 3 SB (first 12 km of the section) with much less cracking on I-91 SB, which was last paved in 2011.

Table 7.1 summarizes the total amount of cracking by class/orientation. Note that the Van 7 system classifies cracking as only longitudinal and transverse, whereas Vans 8 and 9 also report area cracking, albeit in linear units. A few observations are worth noting. First, it appears that on the pavement sections with little or no cracking, all three ARAN systems report similar total lengths of all cracks (about 4000 m on Big Loop Segment C). However, on significantly cracked surfaces the 3-D LCMS systems on Van 8 and Van 9 similarly detect much more cracking, mostly classified as area (about 28000 out of total 39000 m). In addition, one can easily notice that, while comparable longitudinal crack lengths are reported by all three ARANs on Segment D (between 6000 and 8000 m), Vans 8 and 9 report four times less transverse cracking than Van 7 (about 2800 vs. 11500 m, respectively).



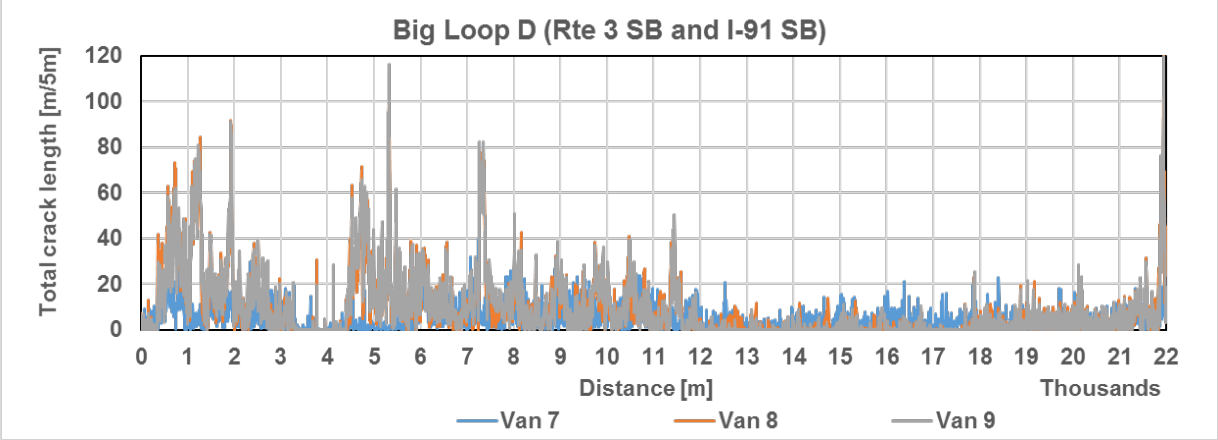


Figure 7.1 ARAN cracking profiles on Big Loop Segment C (top) and Big Loop Segment D (bottom) from June 2016 survey

Table 7.1 Sum of total crack lengths by class/orientation

Section ID	ARAN ID	Crack Class			
		Total Longitudinal [m]	Total Transverse [m]	Total Area (Linear) [m]	GRAND TOTAL [m]
Big Loop Segment C	Van 7	2162	1976	N/A	4138
	Van 8	1680	442	1768	3890
	Van 9	1965	424	1821	4210
Big Loop Segment D	Van 7	6014	11477	N/A	17491
	Van 8	8262	2783	27900	38945
	Van 9	7739	2889	28604	39232

With respect to severity ratings, Table 7.2 indicates that Van 7 rates more than 95 percent of cracks as high-severity distress (with zero low-severity cracks), which is definitely counterintuitive. Vans 8 and 9 report normally distributed severities. This result warranted further investigation into

severity thresholds used by WiseCrax and Roadware Vision. It was found that WiseCrax utilized 3- and 6-mm width as opposed to Roadware Vision's 6- and 12-mm width for low/moderate and moderate/high severity transitions, respectively. This finding rendered comparison of distress severities reported by the newer and older ARAN systems impractical. From this point forward, the investigation of cracking data focused on a comparison of the capability of the ARAN systems to detect ground truth through analysis of pavement surface images, as discussed in the next section.

Table 7.2 Sum of total crack lengths by severity

Section ID	ARAN ID	Crack Severity			
		Total Low [m]	Total Moderate [m]	Total High [m]	GRAND TOTAL [m]
Big Loop C	Van 7	0	82	4056	4138
	Van 8	1612	2012	266	3890
	Van 9	1839	2117	254	4210
Big Loop D	Van 7	0	541	16951	17491
	Van 8	13919	21581	3444	38945
	Van 9	14425	21892	2916	39232

Analysis of Pavement Images to Determine Ground Truth

It was understood in the beginning of this project that direct comparison of estimated crack lengths may not be successful, firstly, because of the differing nature of the data collection processes used by the Strobe (2-D images) and 3-D LCMS systems (reconstruction of images from 3-D surface measured by laser). Due to these differences, the images collected by Strobe and 3-D LCMS systems had to be processed separately through WiseCrax and Roadware Vision routines, respectively. The aforementioned crack length computation schemas had significantly different settings with respect to brightness thresholds for crack detection and crack width thresholds for crack rating by severity. Note that those settings were developed by ratings from a human rater. As noted earlier, one other difference between cracking datasets was that WiseCrax reported two classes of cracks by orientation, i.e. longitudinal and transverse, whereas Roadware Vision recognized so-called area cracks, albeit reporting them in linear units.

Figure 7.2 emphasizes the above-mentioned differences between pavement images generated by Strobe and 3-D laser systems, as well as the challenges found in identifying ground truth from pavement images. The following is a list of discrepancies found with respect to the shown images in Figure 7.2, and other investigated images:

- The WiseCrax and Roadware Vision report different lane widths for the analysis (3.22 and 3.39 m, correspondingly)
- The newer ARAN images included more of the right shoulder, thus indicating that they did not travel in the same wheelpath as Van 7 (Note location of the white line mark on the images).

- The visual estimates of cracks from WiseCrax and Roadware Vision images (highlighted by bright green color) vary due to differences in image brightness and contrast.
- The brighter WiseCrax images led to detection of more cracks as compared with the darker Roadware Vision images.
- Neither WiseCrax nor Roadware Vision automatic detection capabilities matched data gathered with human eyes.
- The Roadware Vision detection algorithm was extra sensitive to the presence of secondary cracks, thus, reporting higher linear estimates for area cracking.

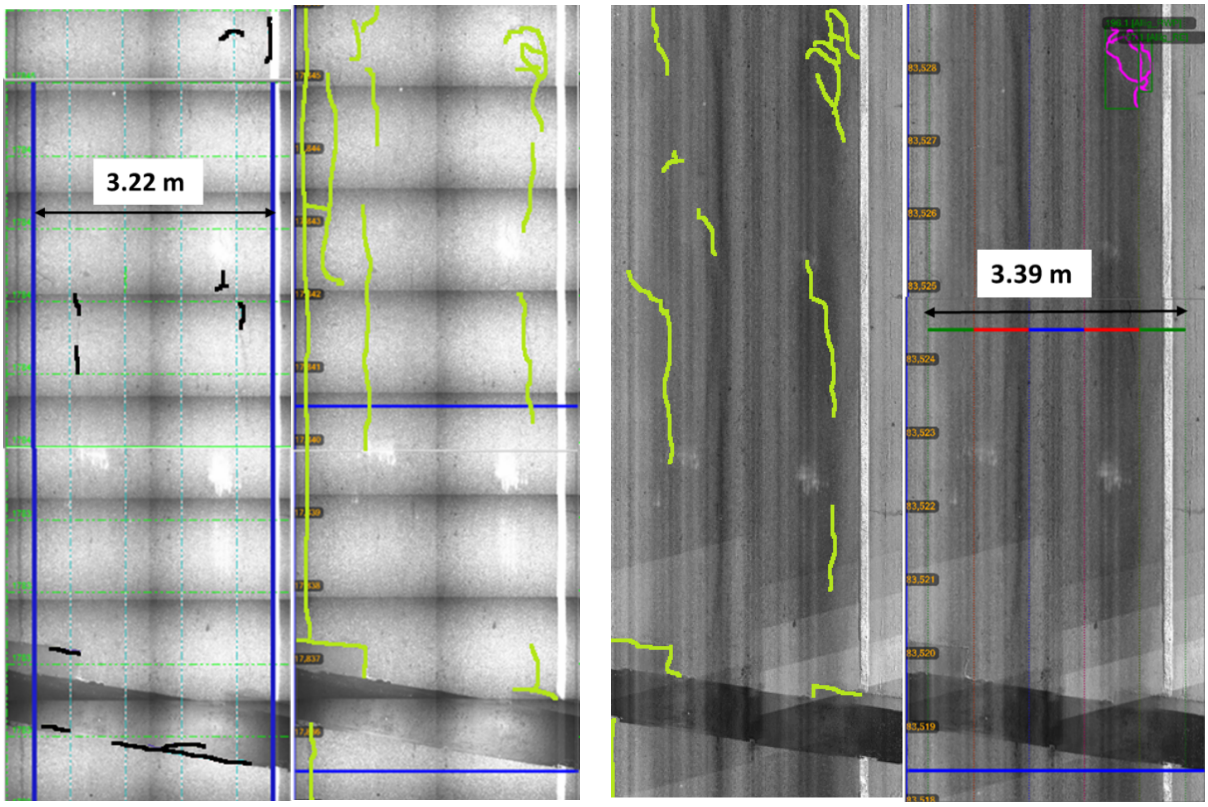


Figure 7.2 Pavement surface images (from left to right) with: WiseCrax-detected cracks from Strobe camera for Van 7; Visually-estimated cracks from Strobe camera for Van 7; Visually-estimated cracks from 3-D LCMS images for Van 8; and Roadware Vision-detected cracks from 3-D LCMS for Van 8.

Evaluation of Transfer Functions

In consideration of all of the above, the following three-step process was set up in an attempt to develop transfer functions for network-level cracking data:

Step 1. Sample randomly suitable pavement images generated by Strobe (Van 7) and LCMS (Van 8) systems.

Step 2. Visually estimate crack lengths per 0.1-mile by type (long., trans., area) from sampled *.jpg pavement images.

Step 3. Evaluate transfer functions between visual estimates and PMS reports by WiseCrax and LCMS in the form of a simple regression model as shown in Figure 7.3

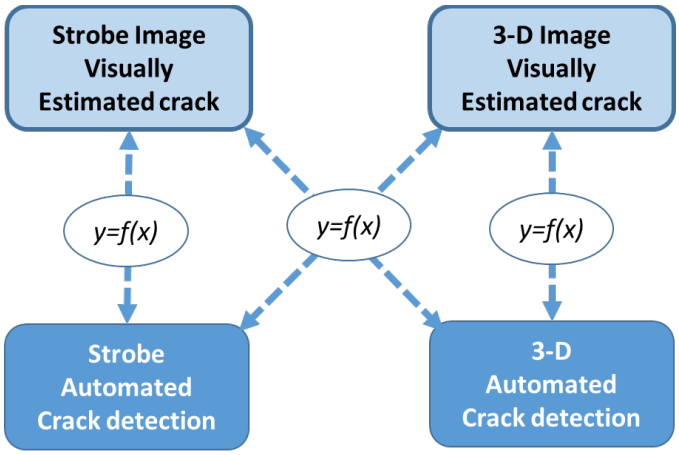


Figure 7.3 Flowchart for evaluation of transfer functions for cracking data

Analysis of Distributions and Random Sampling of Pavement Images

In order to select random yet representative samples of sections with low, moderate, and high amounts of cracking, first, the distributions of crack lengths per 0.1-mi as reported by Van 7 (old system) and Van 8 (representative of new system) were analyzed to determine three quantiles. The charts on Figure 7.4 show that the distribution followed lognormal (normal distribution of logarithms). However, the 3-D LCMS reported no cracks on 49 0.1-mi sections (newly paved), whereas WiseCrax did not report less than 10 ft of cracking on any 0.1-mi section. On Segment D, which has a significant amount of cracking, 3-D LCMS reported many more sections with more than 1000 ft of cracking per 0.1-mi than did WiseCrax. In general, one can notice the distribution of crack lengths reported by the 3-D LCMS on Segment D is very different from that reported by WiseCrax.

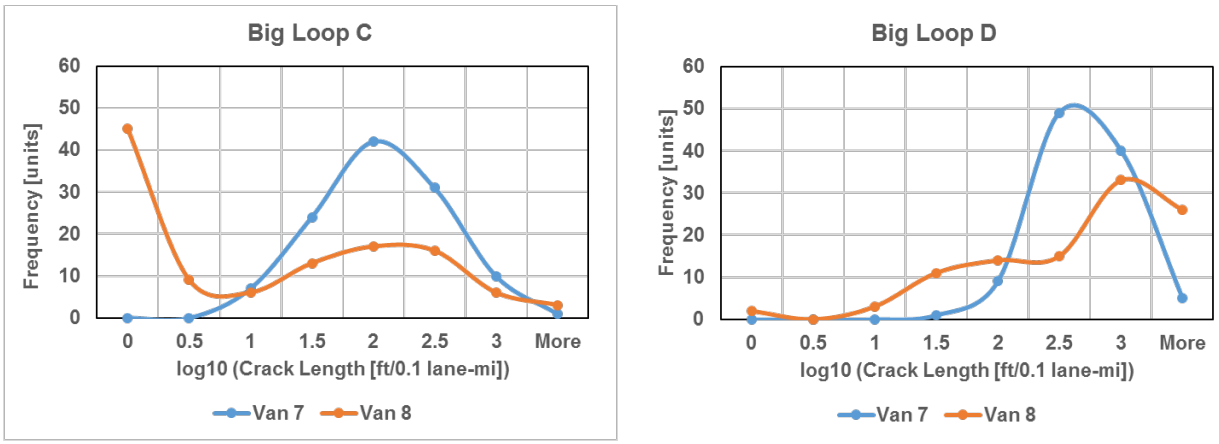


Figure 7.4 Summary of ARAN-reported crack length distributions on Big Loop Segment C (left) and Big Loop Segment D (right).

Once, the 33-percent quantiles were established, six 0.1-mi sections were randomly selected from each quantile. Thus, 18 samples were pooled from each of the two datasets. The charts in Figure 7.5 compare the cumulative percentage of the whole population of samples (115 from Segment C and 104 from Segment D) with that of pooled samples. A comparison of percentiles shown on Figure 7.5 (left) with those on Figure 7.5 (right) indicate that the random samples adequately represent the entire dataset for both Van 7 and Van 8.

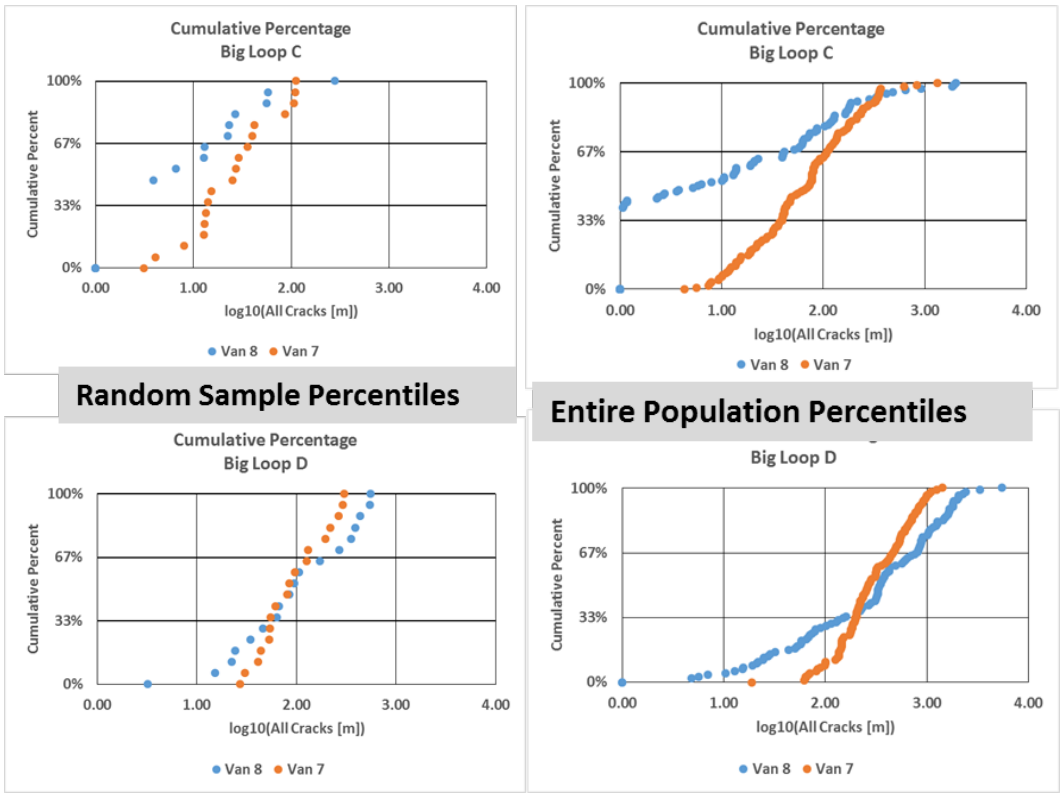


Figure 7.5 Summary of 33-percent quantiles for random samples (left) and entire dataset (right).

Visual Estimation of Cracking from Pavement Images

A total of 1744 images (18*32 5-m images from Van 7 and 18*16 10-m images from Van 8) were analyzed visually to develop ground-truth estimates. Due to differences in the quality of images between WiseCrax and 3-D LCMS (as seen in Figure 7.2), the visual estimates from both systems were checked for adequacy through bivariate analysis. Each point on Figure 7.6 corresponds to a total crack length for one 0.1-mi sample. A better agreement with visually estimated total crack lengths per 0.1-mi was found on Segment D (slope =0.95 at R-Squared = 0.94), where crack extent was higher. A much worse agreement was found for Segment C (slope=0.76 at R-squared=0.65) and was attributed to discrepancies in longitudinal crack reports due to the ARAN vehicles not travelling in the same wheelpath. Note that the lane shift from the 'correct' wheelpath on Segment D for the images from Van 8 was between 6 and 8 inches. Van 7 images indicated a shift toward the shoulder of between 16 and 18 inches on Segment D.

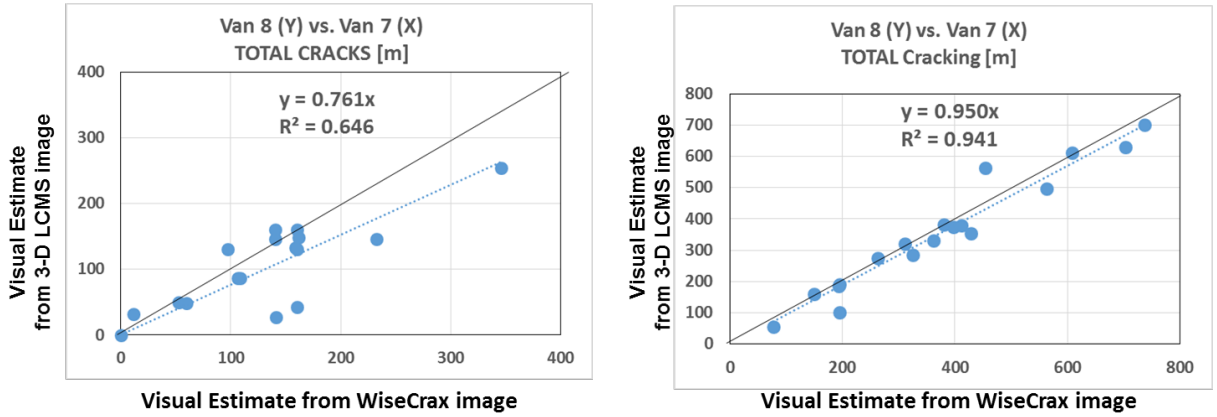


Figure 7.6 Comparison of visually estimated crack lengths per 0.1-mi extracted from WiseCrax and 3-D LCMS images on Big Loop Segment C (left) and Big Loop Segment D (right).

Correlation between WiseCrax and 3-D LCMS automatic crack detection systems

As stated earlier, the direct comparison of automatically detected crack lengths extracted from WiseCrax and 3-D LCMS project-level PMS reports was not successful. In anticipation of an improvement in the comparison when summarized at the network-level, the total lengths per 0.1mi (~160 m) for Segments C and D together (total 219 samples) were superimposed as shown in Figure 7.7. Note that area and longitudinal crack lengths from the 3-D LCMS were totaled to reflect the similar longitudinal orientation of both, and to make them more comparable with longitudinal cracking obtained from WiseCrax. The transverse crack lengths were compared directly because no or very little amount of transverse cracks were classified as area cracks by the 3-D LCMS. From the linear trend estimates shown in Figure 7.7, it can be inferred that 3-D LCMS largely overestimates the longitudinal cracking extent, reporting 2.4 times more cracking than WiseCrax detects. On the other hand, the 3-D LCMS detects only 20 percent of the transverse cracking reported by WiseCrax. Together, the 3-D LCMS reports about 40 percent more cracking overall as compared with WiseCrax. With regard to the significance of bivariate relationships shown in Figure 7.7, none of the R-squared values can be rejected, at least at 99-percent reliability,

although the inter-prediction of longitudinal/area cracks detected by both WiseCrax and 3-D LCMS is deemed impractical due to the very low correlation (R-squared=0.29).

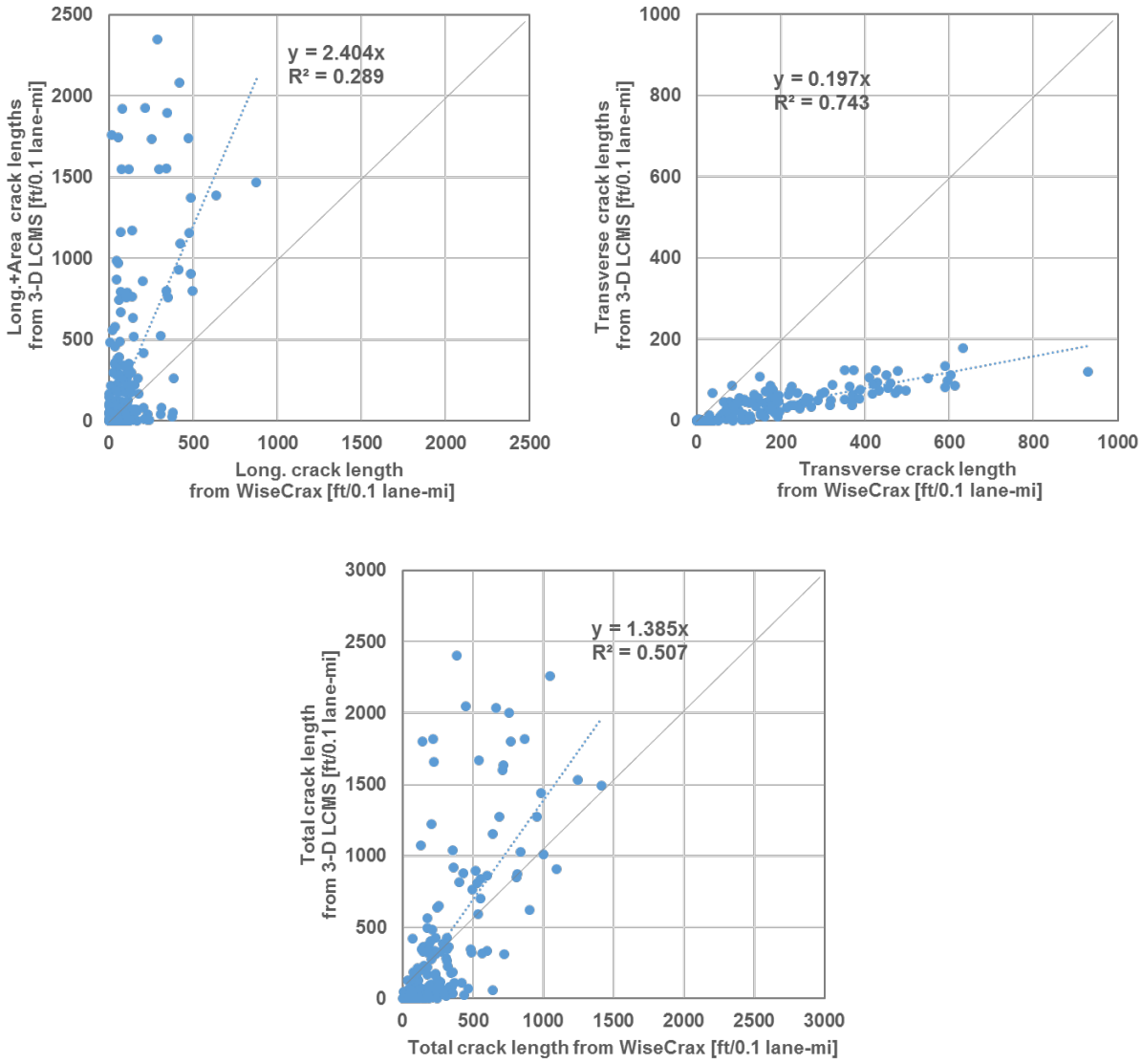


Figure 7.7 Correlation plots of automatically detected crack lengths for longitudinal (top left), transverse (top right), and total cracking (bottom).

Accuracy of automatic crack detection systems with respect to ground truth

The next step leading up to the development of transfer functions for cracking data was a determination of the accuracy of WiseCrax and 3-D LCMS crack detection with respect to data obtained visually from pavement images, which could be considered a form of reference ground truth. This analysis utilized 36 random samples (total of 1.8 lane-miles of surveyed surface) from Big Loop C and D segments. Those segments have differences in surface condition (good condition on C and fair to poor on D) and spatial orientation (westbound travel direction for C and southbound for D). Therefore, the correlation between automatically and visually detected crack

lengths is shown separately for those two segments (Figures 7.8 and 7.9 for Segments C and D, respectively).

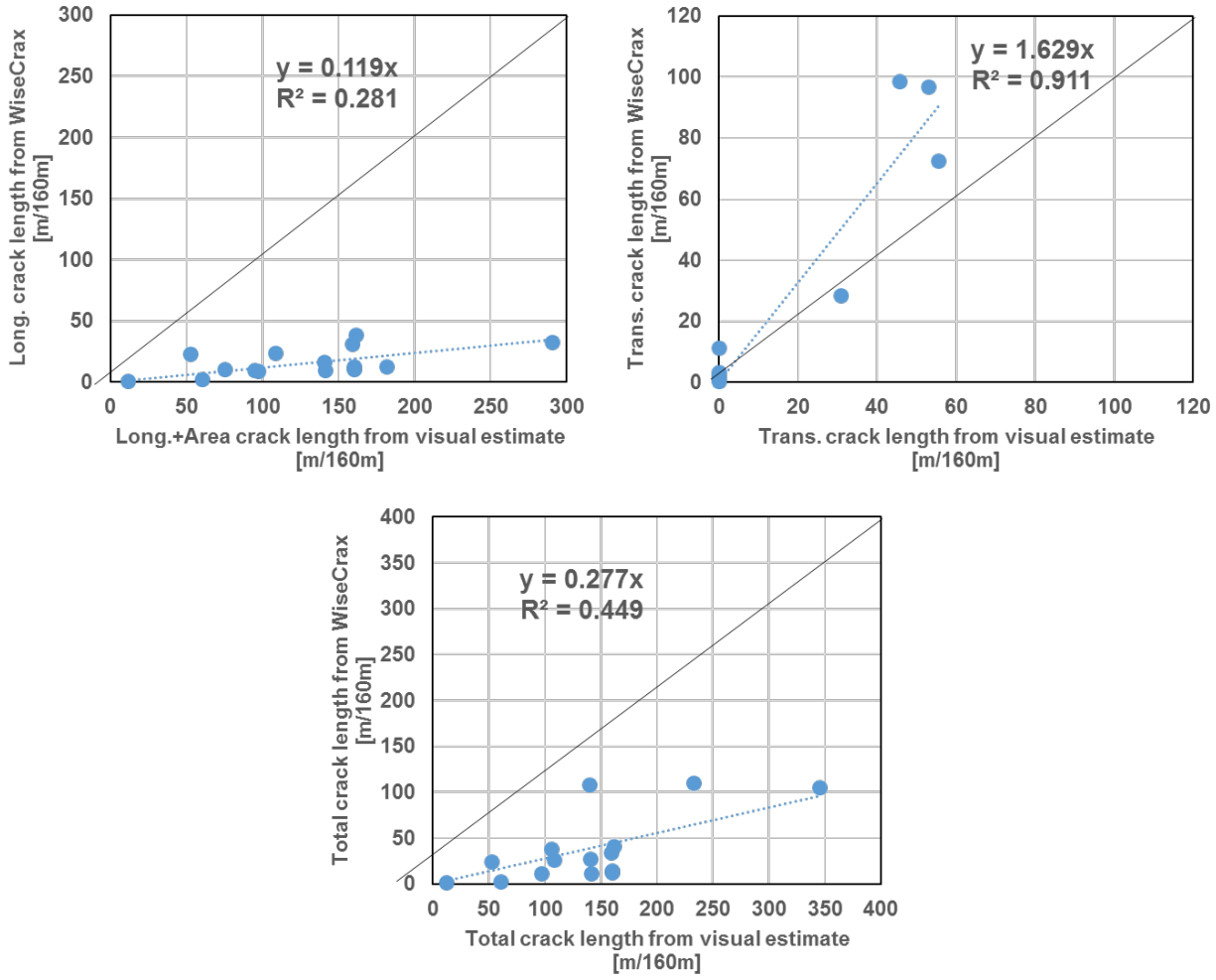


Figure 7.8 Correlation plots of WiseCrax-detected and visually estimated crack lengths on Big Loop Segment C.

The best-fit linear trends shown in Figures 7.8 and 7.9 indicate overall low accuracy of WiseCrax crack detection, based on the following observations:

- The WiseCrax reports about 15% of visually estimated longitudinal crack lengths.
- The WiseCrax tends to overestimate the extent of transverse cracking, reporting 30 to 60 percent more crack length than what was estimated visually.
- Overall, WiseCrax is able to detect about 30% of the visually estimated total crack lengths, most likely due to inadequate threshold settings in the detection algorithm.
- No distinctive difference in accuracy can be noted between data collected on Big Loop Segments C and D despite their different surface condition and geographic orientation. Therefore, the effect of those two factors on accuracy of automated crack detection can be stated as insignificant.

- Statistically, only about 50 percent of differences between calculated WiseCrax and visually estimated (ground truth) cracking can be explained with linear functions at a 95 percent confidence level.

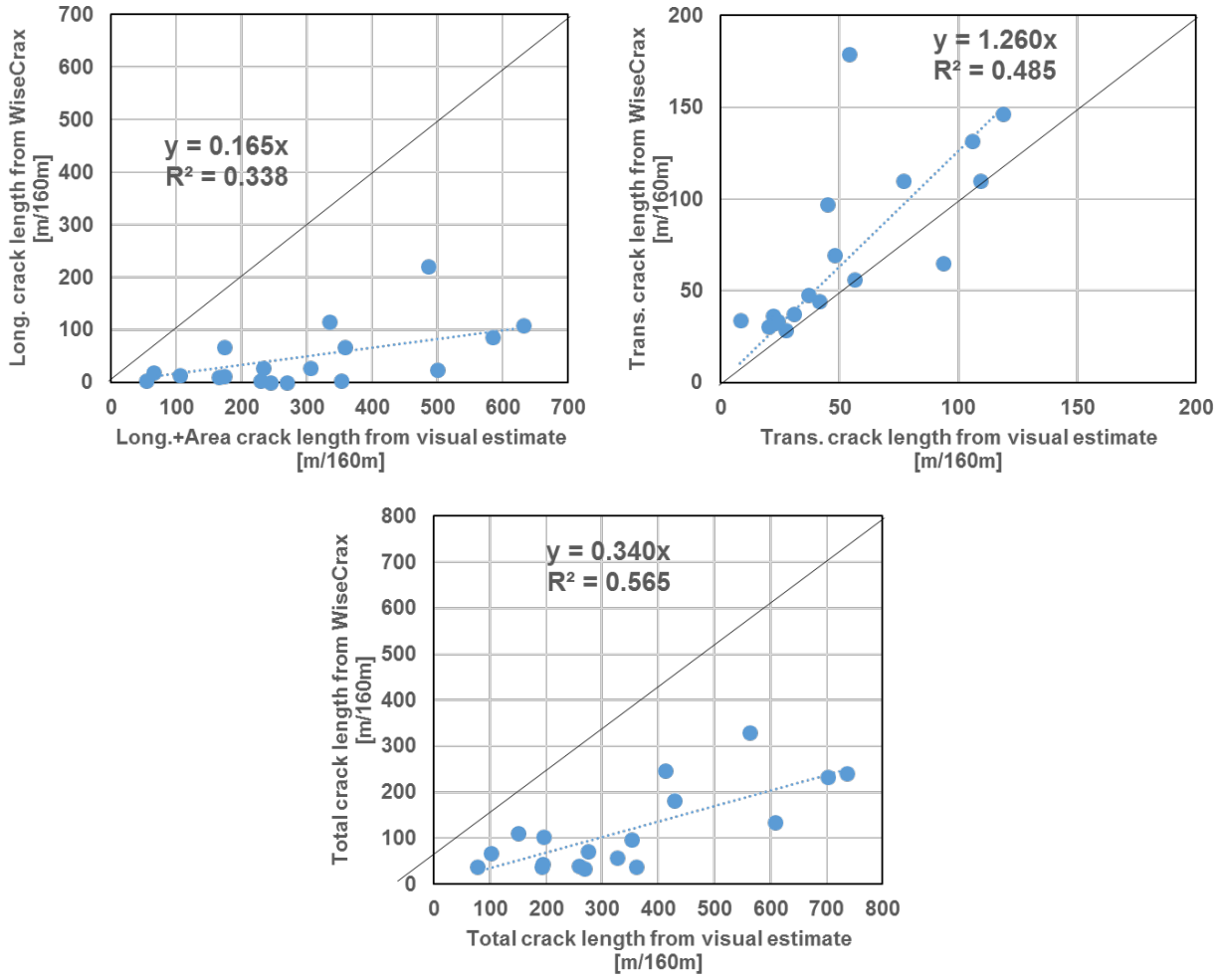


Figure 7.9 Correlation plots of WiseCrax-detected and visually estimated crack lengths on Big Loop Segment D.

A similar analysis of accuracy with respect to visually estimated cracking (reference ground truth) was conducted on 3-D LCMS pavement images of 1.8 lane-miles from each of the Big Loop C and D segments. Figures 7.10 and 7.11 show bivariate plots of correlation between automatically detected and visually estimated crack lengths for those two segments, respectively. Comparing the linear trends between the two segments for three classes of cracks, results in the following:

- The Roadware Vision software reported better accuracy for longitudinal and area crack detection for Big Loop Segment D (~65 percent visual estimate) as compared with Big Loop Segment C (~30 percent of estimate). This eventually affected the total crack detection (35 and 60 percent total visual estimates for Segments C and D, respectively).

- The difference in accuracy between Segment C and D sections may indicate higher sensitivity of 3-D LCMS to overall distress condition (i.e., amount of distress present) as compared with WiseCrax. 3-D LCMS underestimates cracking extent on "good" surfaces and overestimates it on "fair" and "poor" surfaces.
- The Roadware Vision software detected about 35 percent of the visually estimated transverse crack lengths.
- In terms of statistical significance, about 40 percent of the differences between automatically detected cracking by Roadware Vision and visually estimated total crack lengths can be explained with linear functions at a 95-percent confidence level.

Careful examination of lane limits for Van 8 in the Roadware Vision software revealed that most of the longitudinal cracks near the left edge of the pavement image fell exactly or slightly out (up to 4 in.) of software lane limits (see example on Figure 7.3 [right]). Hence, the extremely low correlation in the chart in Figure 7.10 (top left).

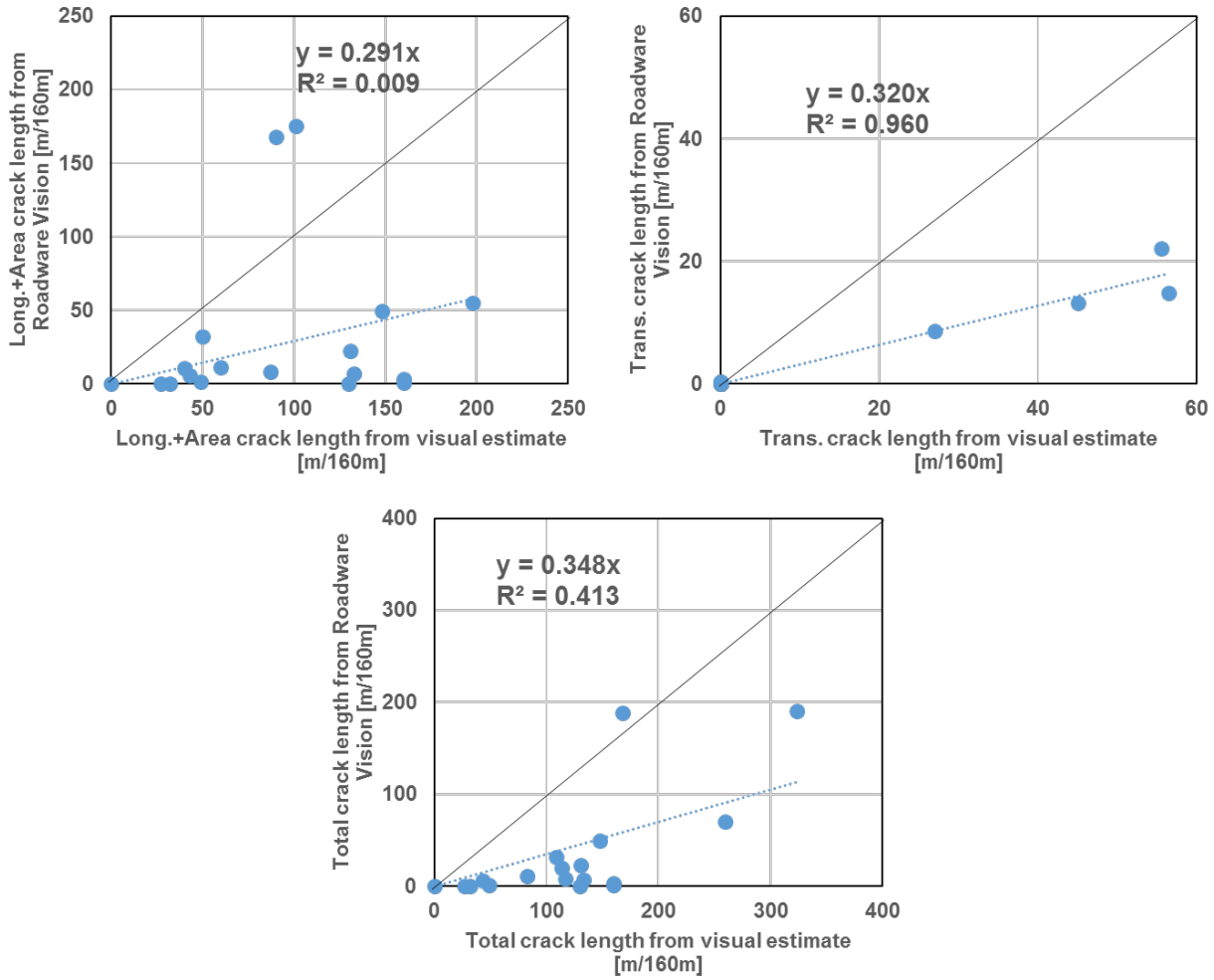


Figure 7.10 Correlation plots of Roadware Vision-detected and visually estimated crack lengths on Big Loop Segment C

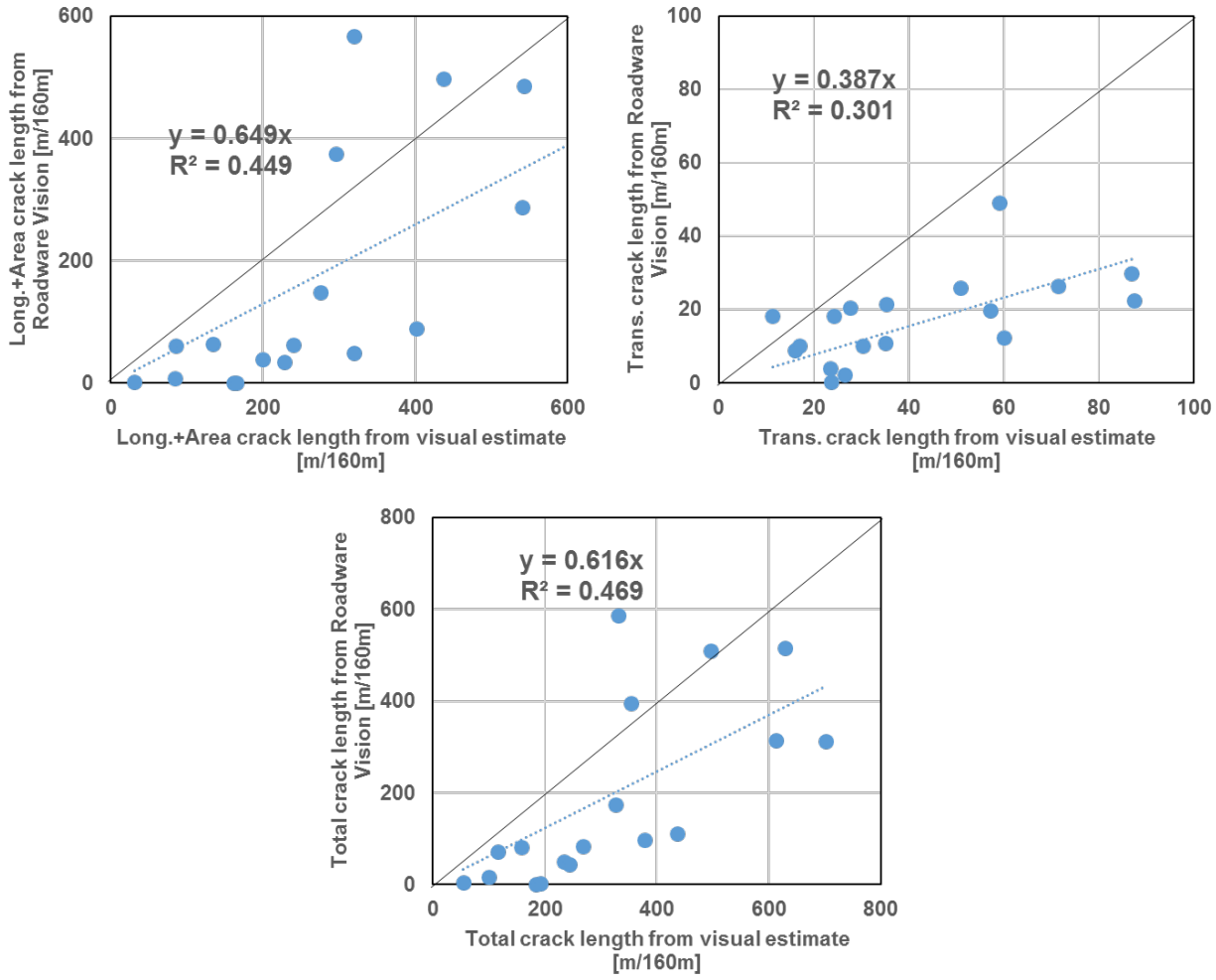


Figure 7.11 Correlation plots of Roadware Vision and visually estimated crack lengths on Big Loop Segment D

To sum up the discussion of accuracy of automated crack detection, an overall low level of detection (low bias at low precision) for both WiseCrax and 3-D LCMS algorithms was revealed by the statistical analysis of correlations. Nevertheless, the total crack estimates for the sample of 1.8 lane-miles shown in Tables 7.3 and 7.4 indicate that on average, the 3-D LCMS system was able to detect crack extent that was closer to the visually-estimated one (58 percent of total) than did WiseCrax (31 percent). This outcome, however, is not statistically reliable due to high variability and the relatively small sample size of the analyzed data.

Table 7.3 Summary of crack detection with WiseCrax algorithm.

Segment	Crack Type	Visual Estimate	Automated Detection	Percent Detected
Big Loop C	Longitudinal	2052	262	13%
	Transverse	185	331	179%
	TOTAL	2237	593	27%
Big Loop D	Longitudinal	5810	806	14%
	Transverse	943	1289	137%
	TOTAL	6753	2095	31%

Table 7.4 Summary of crack detection with 3-D LCMS algorithm.

Segment	Crack Type	Visual Estimate	Automated Detection	Percent Detected
Big Loop C	Longitudinal	1639	550	34%
	Transverse	184	60	33%
	TOTAL	1823	610	33%
Big Loop D	Longitudinal	5075	3065	60%
	Transverse	744	311	42%
	TOTAL	5819	3376	58%

Precision and Accuracy of Cracking Survey by 3-D LCMS System

The very low agreement between the crack lengths reported by the newer and older ARAN systems during the survey in June 2016 led to questioning their precision and reciprocal accuracy, or reproducibility. Accordingly, five repeatability runs on short segments of Thornbush Road and Brook Street, as well as an additional run on the Big Loop were conducted in November 2016 to establish repeatability and reproducibility of crack length surveys. Unfortunately, Van 7 had been decommissioned by that time and, therefore, only precision and reproducibility of Van 8 and Van 9 is presented in this section. It should also be noted that due to higher variability inherent in project-level data, the discussion is focused here on network-level data, which are comprised of the sum of crack lengths per 0.1 mile (~161 m).

Similar to the IRI analysis, the precision of the crack length survey was estimated by computing variability of data collected over five consecutive runs. The absolute precision, then, would be an average standard deviation from the mean of five values measured for each 0.1-mi section. The relative precision would be an average coefficient of variation, i.e. ratio of standard deviation of that mean. The reciprocal accuracy of crack length survey was estimated by the "between" variability, i.e. standard deviation from the average of the two vans' means (absolute accuracy) and the corresponding coefficient of variation (relative accuracy). The precision and accuracy terms were applied to crack classes by orientation (longitudinal, transverse, and area) and by location within a lane (edges, wheelpaths, and center) as well as to crack severity (low, medium, and high), as reported by Roadware Vision software.

Precision of Crack Survey on Repeatability Runs.

Five consecutive repeatability runs were conducted simultaneously with ARAN Van 8 and Van 9 on each of the two segments – approximately 1-km long Thornbush Road southbound and 2.7-km long Brook Street (both eastbound and westbound). As shown in Figure 7.12, the cracking on Brook Street varied between none and very little (zero to 40 ft/0.1 lane-mi) and, therefore, only Thornbush Road data (cracking of 1000 to 2000 ft/0.1 lane-mi) was analyzed for precision.

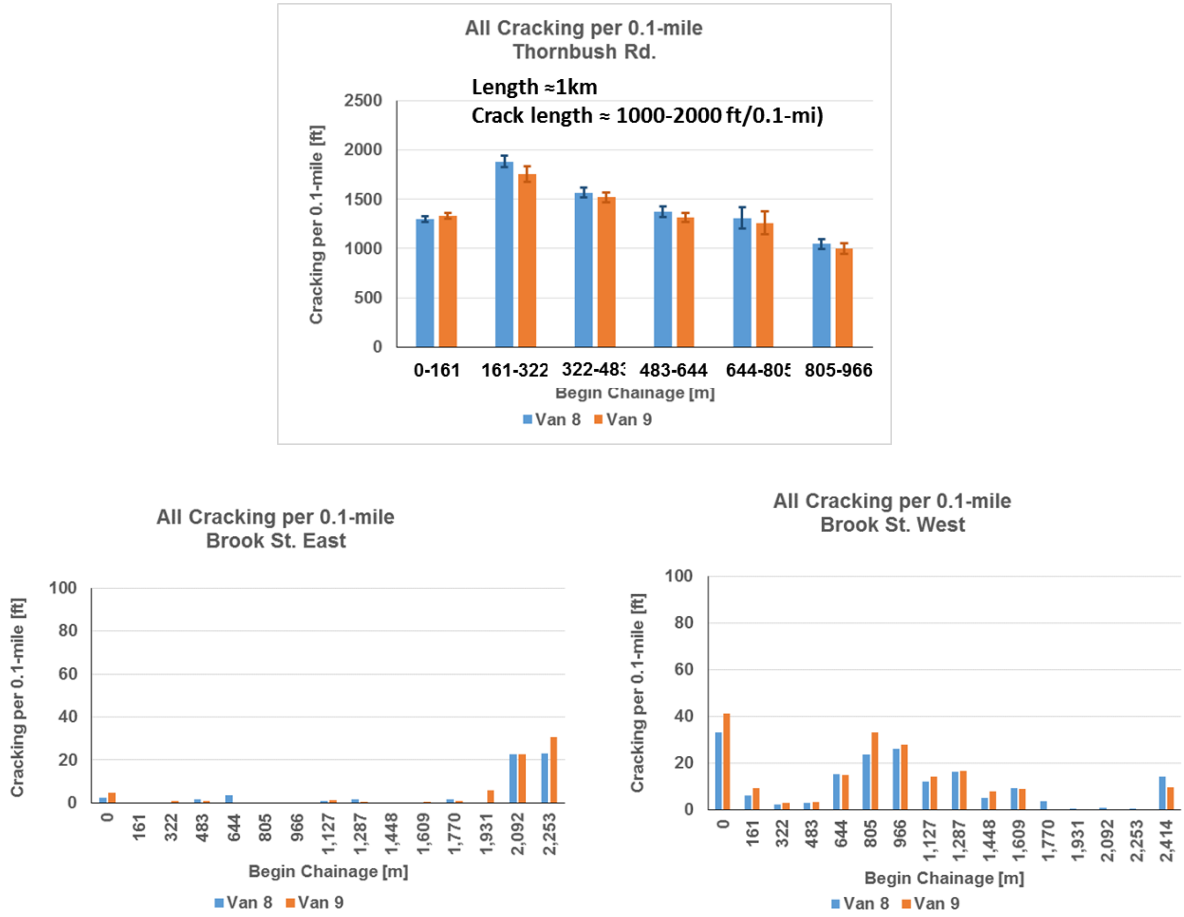


Figure 7.12 Summary of total crack lengths per 0.1-mi on Thornbush Road (top), Brook Street EB (bottom left), and Brook Street WB (bottom right).

Precision results for crack length (by orientation, location, and severity) for Vans 8 and 9 are summarized in Tables 7.5 and 7.6. Since each value in the tables is an average of six (for six 0.1-mi sections), their 95-percent confident upper threshold of precision is reported in parentheses. As the numbers in Table 7.5 suggest, both Van 8 and Van 9 systems produced highly repeatable results (10-percent precision at 95-percent confidence) with regard to longitudinal cracking, while yielding lower precision for transverse and area crack measurements (13 to 21-percent precision at 95 –percent confidence).

Table 7.5 Summary of crack survey precision by orientation (crack type).

Crack Type	Parameter	Van 8	Van 9
Longitudinal	Absolute Accuracy (upper 95% CI), ft/0.1 lane-mi	27 (48)	28 (52)
	Relative Accuracy (upper 95% CI)	5% (10%)	5% (10%)
Transverse	Absolute Accuracy (upper 95% CI), ft/0.1 lane-mi	19 (38)	17 (24)
	Relative Accuracy (upper 95% CI)	11% (21%)	10% (13%)
Area	Absolute Accuracy (upper 95% CI), ft/0.1 lane-mi	53 (79)	67 (108)
	Relative Accuracy (upper 95% CI)	8% (13%)	10% (16%)
All Cracks	Absolute Accuracy (upper 95% CI), ft/0.1 lane-mi	58 (106)	62 (118)
	Relative Accuracy (upper 95% CI)	4% (8%)	5% (9%)

The precision results with respect to crack location within the pavement lane (Table 7.6) are highly variable with the best precision achieved in the left wheelpath (16-percent precision at 95-percent confidence). All other zones show low to very low precision of 30 to 90-percent variation from the mean at 95-percent confidence. One can argue that for local roads such as Brook Street 50-percent reliability level for pavement survey is acceptable. Still, 25 to 45-percent precision is much lower than 5 to 10-percent variation from mean usually accepted by many agencies (Pierce et al., 2013).

Table 7.6 Summary of crack survey precision by location within lane.

Crack Location Zone	Parameter	Van 8	Van 9
Left Edge	Absolute Accuracy (upper 95% CI), ft/0.1 lane-mi	40 (92)	19 (34)
	Relative Accuracy (upper 95% CI)	46% (94%)	40% (74%)
Left Wheelpath	Absolute Accuracy (upper 95% CI), ft/0.1 lane-mi	42 (74)	35 (48)
	Relative Accuracy (upper 95% CI)	10% (16%)	9% (16%)
Center	Absolute Accuracy (upper 95% CI), ft/0.1 lane-mi	15(34)	22 (39)
	Relative Accuracy (upper 95% CI)	13% (25%)	20% (38%)
Right Wheelpath	Absolute Accuracy (upper 95% CI), ft/0.1 lane-mi	64(144)	82 (133)
	Relative Accuracy (upper 95% CI)	16% (32%)	24% (34%)
Right Edge	Absolute Accuracy (upper 95% CI), ft/0.1 lane-mi	60 (119)	60 (117)
	Relative Accuracy (upper 95% CI)	19% (43%)	19% (49%)

Where the precision of crack rating by severity is considered (Table 7.7), one can notice very similar precision levels shown by Van 8 and Van 9. For low- and medium-severity cracks, both

systems yield 10-percent precision on average (16-percent at 95-percent confidence). A relatively low precision found for high-severity cracking (17-percent at 95-percent confidence) may be related to an overall low amount of those distresses on the surface (See Table 7.2).

Table 7.7 Summary of crack survey precision by severity level.

Crack Severity	Parameter	Van 8	Van 9
Low	Absolute Accuracy (upper 95% CI), ft/0.1 lane-mi	44 (75)	46 (74)
	Relative Accuracy (upper 95% CI)	7% (14%)	9% (16%)
Medium	Absolute Accuracy (upper 95% CI), ft/0.1 lane-mi	57 (93)	59 (88))
	Relative Accuracy (upper 95% CI)	11% (18%)	11% (16%)
High	Absolute Accuracy (upper 95% CI), ft/0.1 lane-mi	23 (52)	24 (56)
	Relative Accuracy (upper 95% CI)	16% (36%)	17% (38%)

To sum up the discussion on precision of 3-D LCMS crack surveys, it can be concluded that Van 8 and 9 exhibit very similar precision levels. Extensively cracked sections (more than 1000 ft/lane-mi) produced much better repeatability as compared with sections with low cracking extent (less than 100 ft/lane-mi). This is not an unexpected trend for precision. However, both Vans exhibit low precision with regard to specific locations or zones within a pavement lane. This likely occurs due to differences in lateral position within the lane between the two vehicles because of not driving in exactly the same path. Differences in inclination angle of downview cameras may also have led to these results.

Reproducibility of Crack Surveys between Van 8 and Van 9

In the absence of reference values (or ground truth) for the crack surveys on the repeatability runs, only reproducibility between Van 8 and Van 9 could be determined. The absolute and relative values of the accuracy are reported for Thornbush Road (high cracking extent) and Brook Street (low cracking extent) in Tables 7.8 through 7.10, which distinguish between precision for crack survey by orientation, location, and severity, respectively. In addition, the differences between network-level cracking data reported by the two supposedly identical LCMS systems during repeated surveys in June and November 2016 on the Big Loop are summarized in Tables 7.11 through 7.13.

Reproducibility of crack survey on Thornbush Road and Brook Street

When comparing the levels of accuracy on Thornbush Road and Brook Street, one can immediately notice that relative accuracy decreases drastically (e.g. the variation increases sharply) with decrease in cracking level. It is also noticeable that the reproducibility between the ARAN vans is very similar or slightly better than their respective precisions. This may indicate that most of the time the errors in measurements are random, that is, no systematic error between Van 8 and Van 9 survey can be established. Further examination of the precision levels in Tables 7.8 through 7.10 reveals an overall good accuracy in classifying cracks by orientation (3 to 7 percent [11 to 18 percent at 95-percent confidence]) and a fair accuracy with respect to severity

ratings (7 to 13 percent at 50-percent reliability). However, identifying and placing cracks into specific lane zones is not very accurate, especially at the edges (expect 40 to 70 percent variation at 95-percent confidence).

Table 7.8 Summary of crack survey accuracy by orientation (crack type).

Crack Type	Parameter	Thornbush Rd.	Brook St.
Longitudinal	Absolute Accuracy (upper 95% CI), ft/0.1 lane-mi	19 (56)	1 (3)
	Relative Accuracy (upper 95% CI)	4% (11%)	33% (83%)
Transverse	Absolute Accuracy (upper 95% CI), ft/0.1 lane-mi	12 (32)	1 (4)
	Relative Accuracy (upper 95% CI)	7% (18%)	41% (89%)
Area	Absolute Accuracy (upper 95% CI), ft/0.1 lane-mi	43 (122)	1 (7)
	Relative Accuracy (upper 95% CI)	6% (17%)	14% (46%)
All Cracks	Absolute Accuracy (upper 95% CI), ft/0.1 lane-mi	47 (122)	2 (7)
	Relative Accuracy (upper 95% CI)	3% (11%)	15% (47%)

Table 7.9 Summary of crack survey accuracy by location within lane.

Crack Location Zone	Parameter	Thornbush Rd.	Brook St.
Left Edge	Absolute Accuracy (upper 95% CI), ft/0.1 lane-mi	24 (82)	0.2 (1.2)
	Relative Accuracy (upper 95% CI)	29% (71%)	32% (81%)
Left Wheelpath	Absolute Accuracy (upper 95% CI), ft/0.1 lane-mi	24 (61)	1 (3)
	Relative Accuracy (upper 95% CI)	6% (16%)	23% (70%)
Center	Absolute Accuracy (upper 95% CI), ft/0.1 lane-mi	11 (33)	1 (5)
	Relative Accuracy (upper 95% CI)	10% (28%)	17% (47%)
Right Wheelpath	Absolute Accuracy (upper 95% CI), ft/0.1 lane-mi	52 (!56)	1 (4)
	Relative Accuracy (upper 95% CI)	14% (40%)	30% (70%)
Right Edge	Absolute Accuracy (upper 95% CI), ft/0.1 lane-mi	41 (115)	0.5 (3)
	Relative Accuracy (upper 95% CI)	12% (35%)	30% (86%)

Table 7.10 Summary of crack survey accuracy by severity level.

Crack Severity	Parameter	Thornbush Rd.	Brook St.
Low	Absolute Accuracy (upper 95% CI), ft/0.1 lane-mi	38 (89)	1 (4)
	Relative Accuracy (upper 95% CI)	7% (16%)	34% (86%)
Medium	Absolute Accuracy (upper 95% CI), ft/0.1 lane-mi	51 (111)	1 (5)
	Relative Accuracy (upper 95% CI)	10% (21%)	31% (82%)
High	Absolute Accuracy (upper 95% CI), ft/0.1 lane-mi	21 (62)	1 (6)
	Relative Accuracy (upper 95% CI)	13% (37%)	20% (60%)

Reproducibility of crack survey on Big Loop

Since the Big Loop was surveyed twice (in June and November 2016), it is of interest to compare reproducibility of the crack survey over a longer period to detect any changes that might occur in equipment function (as a result of preventive maintenance), between surveys or due to seasonal variation. When comparing accuracy results for all categories of cracking between the two surveys, it can be seen that no significant change in accuracy occurred between the surveys. With respect to crack orientation (Table 7.11), relatively poor accuracy is reported on average (18 to 26 percent) and even poorer if 95-percent confidence in results is desired (note 60 to 80-percent variation). Similar ranges of variation between the two ARAN vans are reported with respect to crack location and severity (16 to 36 percent on average with 50 to 90-percent variation at 95 percent confidence).

Table 7.11 Comparison of crack survey accuracy on Big Loop by orientation (crack type).

Crack Type	Parameter	June 2016 Survey	November 2016 Survey
Longitudinal	Absolute Accuracy (upper 95% CI), ft/0.1 lane-mi	14 (53)	13 (50)
	Relative Accuracy (upper 95% CI)	26% (80%)	23% (68%)
Transverse	Absolute Accuracy (upper 95% CI), ft/0.1 lane-mi	3 (12)	3 (13)
	Relative Accuracy (upper 95% CI)	18% (60%)	17% (55%)
Area	Absolute Accuracy (upper 95% CI), ft/0.1 lane-mi	23 (117)	22 (126)
	Relative Accuracy (upper 95% CI)	20% (59%)	19% (57%)
All Cracks	Absolute Accuracy (upper 95% CI), ft/0.1 lane-mi	30 (130)	30 (140)
	Relative Accuracy (upper 95% CI)	20% (70%)	20% (64%)

Table 7.12 Comparison of crack survey accuracy on Big Loop by location within lane.

Crack Location Zone	Parameter	June 2016 Survey	November 2016 Survey
Left Edge	Absolute Accuracy (upper 95% CI), ft/0.1 lane-mi	21 (113)	19 (116)
	Relative Accuracy (upper 95% CI)	27% (82%)	24% (72%)
Left Wheelpath	Absolute Accuracy (upper 95% CI), ft/0.1 lane-mi	7 (54)	5 (33)
	Relative Accuracy (upper 95% CI)	17% (49%)	18% (52%)
Center	Absolute Accuracy (upper 95% CI), ft/0.1 lane-mi	6 (47)	5 (28)
	Relative Accuracy (upper 95% CI)	17% (49%)	16% (44%)
Right Wheelpath	Absolute Accuracy (upper 95% CI), ft/0.1 lane-mi	15 (89)	13 (67)
	Relative Accuracy (upper 95% CI)	25% (70%)	25% (72%)
Right Edge	Absolute Accuracy (upper 95% CI), ft/0.1 lane-mi	18 (89)	14 (66)
	Relative Accuracy (upper 95% CI)	36% (94%)	30% (84%)

Table 7.13 Comparison of crack survey accuracy on Big Loop by severity level.

Crack Severity	Parameter	June 2016 Survey	November 2016 Survey
Low	Absolute Accuracy (upper 95% CI), ft/0.1 lane-mi	17 (85)	15 (67)
	Relative Accuracy (upper 95% CI)	23% (74%)	23 (73%)
Medium	Absolute Accuracy (upper 95% CI), ft/0.1 lane-mi	22 (99)	24 (124)
	Relative Accuracy (upper 95% CI)	26% (80%)	24% (69%)
High	Absolute Accuracy (upper 95% CI), ft/0.1 lane-mi	10 (71)	5 (27)
	Relative Accuracy (upper 95% CI)	38% (94%)	39% (97%)

The poor reproducibility of crack surveys on the Big Loop can possibly be explained by the accuracy being derived from only two single runs, whereas on Thornbush Rd. and Brook Street multiple runs were made. The other important factor could be the varying wander between the two vehicles' paths, particularly over the longer survey section (consider 40-km versus 1.0 and 2.7-km lengths of survey on Big Loop Segments C&D and Thornbush Rd. and Brook St., respectively). It seems that the increase in length of survey created cumulative relative error between the two ARAN datasets.

Effect of cracking extent (pavement condition) on variability in crack survey

As stated before, the analysis of crack length data from all surveys under this project produced crack length values (in ft/0.1 lane-mile) over a range of at least three orders of magnitude, and having *lognormal* statistical distributions. Furthermore, a detailed analysis of the ranges of relative accuracy for all categories showed variability that was highly dependent on the magnitude of crack length. In formal statistics, such a trend is called heteroscedasticity or unequal variance over the population. To illustrate this trend, Figure 7.13 compares average values of relative accuracy between Van 8 and Van 9 on the Big Loop for three cracking conditions: less than 100, 100 to 1000, and more than 1000 ft per lane-mile of cracking. One can easily notice a significant decrease in variability with an increase in cracking extent for all categories of cracks.

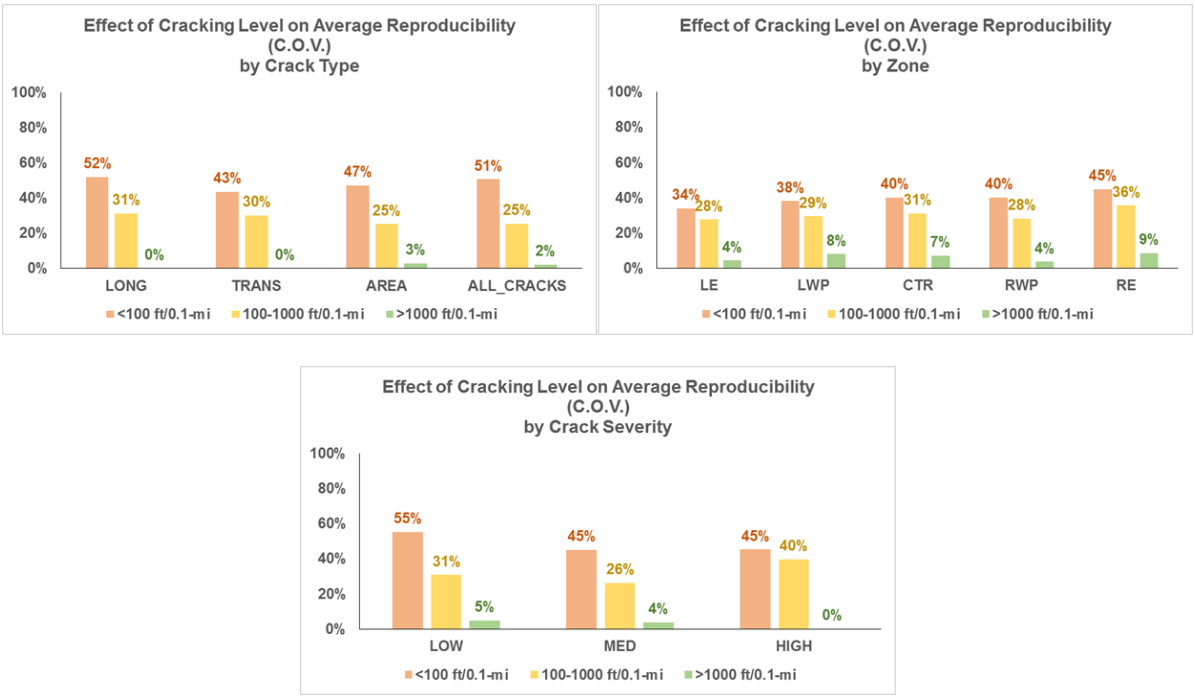


Figure 7.13 Effect of cracking extent on variability in crack survey on the Big Loop.

To investigate possible reasons for the variability trends shown in Figure 7.13, sample pavement images with detected, classified, and rated cracks were pooled for equivalent locations. An example is shown in Figure 7.14, where it can be seen that Van 9 fails to detect some cracks reported by Van 8 (note presence of cracks on the left image and their absence on the right image as highlighted by ellipses). In addition, different lengths and amount of cracking is assigned to the same distress (note differences within lower right corner of both images, for instance). Lastly, the same crack may be rated differently (note green, yellow, and red lines for low, medium, and high severity, respectively). Although the reasons are unknown, such an outcome is believed to be caused by differences in brightness/contrast of the image and/or algorithms used in the Roadware Vision statistical routines for detecting, classifying, and rating the cracks.

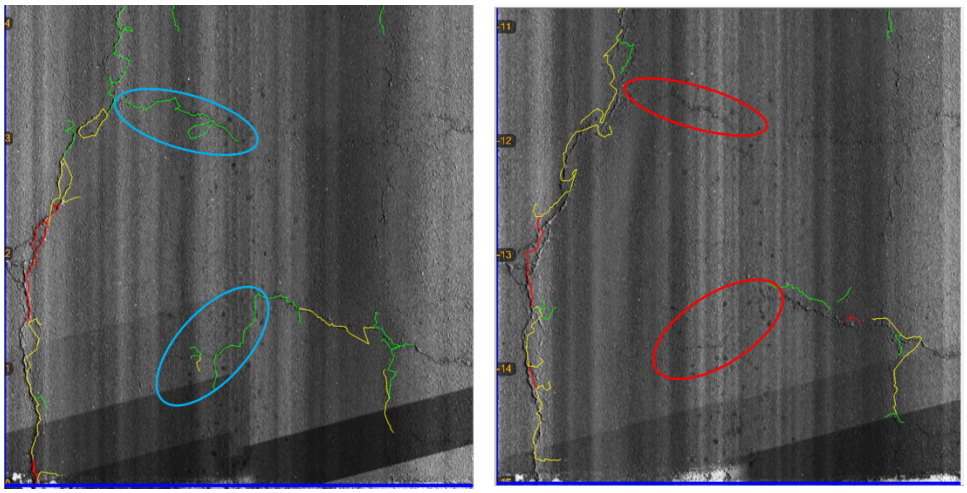


Figure 7.14 Screenshots of crack detection, classification, and rating results from Van 8 (left) and Van 9 (right). Location: Big Loop Segment A, Chainage 0.0 m. Note differences in detection (highlighted within ellipses) and in rating (difference in colors assigned by Roadware Vision for the same cracks).

Effect of Difference in Cracking PMS Reports on Pavement Serviceability Rating

The Connecticut DOT utilizes cracking data to compute an aggregate index of pavement condition. One of these indices that has been used extensively in the past is called Pavement Serviceability Rating (PSR). In this PSR rating, the cracking index is the highest weighting factor at 25% (see Equation 7.1). Therefore, it is important to understand how the variability in crack surveys between the different ARAN systems affects the PSR value.

$$Index\_PSR = 0.25*Index\_Cracking + 0.15*Index\_Distortion + 0.3*Index\_Disintegration + 0.2*Index\_Drainage + 0.1*Index\_Ride \quad (7.1)$$

The computation of the cracking index incorporated into the PSR (range of 1 to 9 for very poor to very good rating) is a complicated routine developed empirically by CTDOT PMS personnel, and it underwent a few changes between 2008 and 2014. However, a regression analysis revealed a highly reliable simple relationship between *Index_Cracking* and sum of all cracks in ft per 10 lane-meters (length of pavement images generated by Roadware Vision). The regression equation in the form of $Y=a-b*Ln(X)$ was derived by the UConn research team for each year between 2008 and 2014 for a dataset of about 45,000 0.1-mile long sections (obtained from the CTDOT PMIS), and the average regression coefficients that were determined are shown in Figure 7.15.

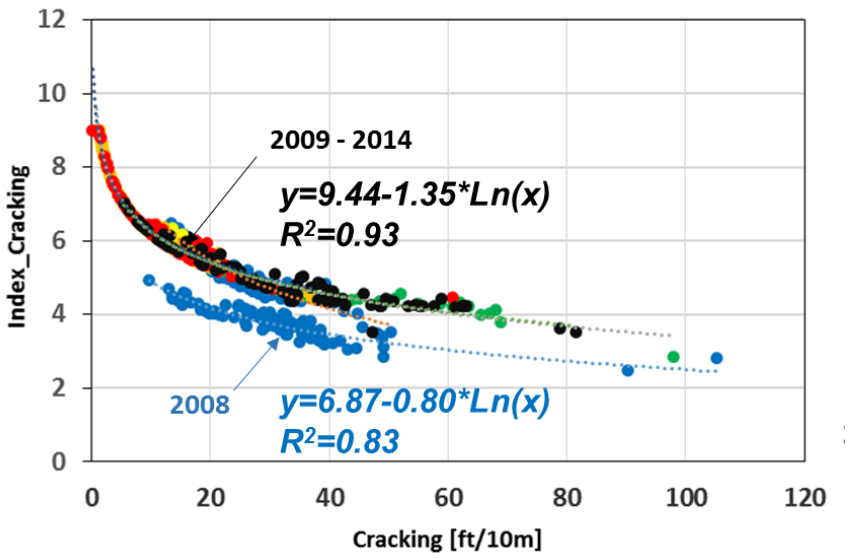


Figure 7.15 Regression trends of Index_Cracking vs. Total Cracking in ft/10 lane-m.

The regression equation for 2009-2014 data was then applied to the crack survey datasets generated by 2-D WiseCrax and 3-D LCMS analysis on Big Loop Segment D. This section was chosen because it contains a wide range of pavement conditions. Next, the bivariate plots of total cracking and *Index_Cracking* were generated as shown in Figure 7.16 left and right, respectively. The best-fit linear trend for total cracking plot on the left indicates that on average 3-D LCMS reports 45 percent more crack length than 2-D WiseCrax does. On the other hand, when translated to the *Index_Cracking* value (see plot on the right), this difference is practically negligible (note slope of 1.02, i.e. 2% difference on average). However, the very low goodness of fit, i.e. R-squared of 0.36, for both indicators, confirms the low precision of cracking survey and, as a result, low reliability of cracking index incorporated into the PSR metric.

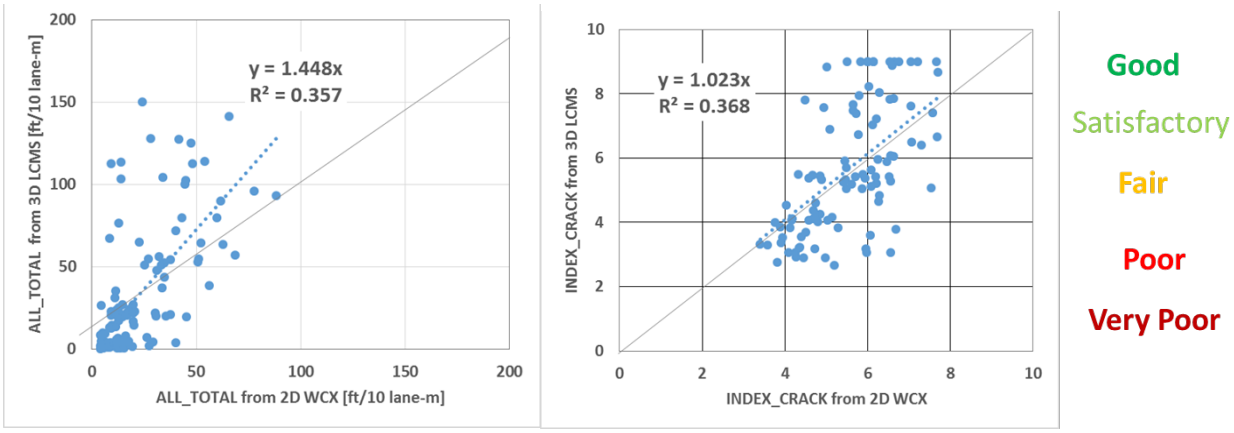


Figure 7.16 Bivariate plots of total crack length (left) and corresponding Index_Cracking (right) for 3-D LCMS vs. 2-D WiseCrax systems.

If the pavement ratings are assigned as shown in Figure 7.16 (right), the 3-D LCMS appears to overestimate quantities for medium to high extent cracking sections, thus leading to the

assignment of a "poor" rating to pavements rated as "fair" by 2-D WiseCrax. On the other hand, the 3-D LCMS system underestimates the amount of low-extent cracking and it may assign historically "fair" pavements as "satisfactory" and/or "good." The fact that none of the ARAN systems are close to the visually estimated extent of cracking (ground truth for this project) warrants further investigation into automated cracking surveys with possible re-evaluation of the relationship between ground-truth cracking and PSR values used for PMS purposes.

Summary of Recommendations of Further Research on ARAN Cracking Data

Collecting and Reporting:

- Replace 5-m reporting interval with 10-m interval to match PMS units of cracking.
- Reduce number of reported crack classes to longitudinal/transverse and wheelpath/non-wheelpath cracking
- Re-evaluate severity thresholds for width and depth so as to better match ground truth datasets

Improving crack detection capability:

- The most likely reason for an overall low level of detection is the distress schemas used for 2D WiseCrax and 3D LCMS, such as brightness/contrast thresholds and severity definitions.
- Another very important reason for low agreement between any two automated systems is lateral shift within the travel lane during the data collection, i.e., either by not driving consistently within the wheelpaths and/or using an incorrect LCMS laser/camera orientation.
- The level of detection of 3D LCMS should be further investigated based on a larger sample from 2016 data, as well as future verification runs during the next 3 years.

Impact on PMS:

- Investigate effect of increase in reported cracking on Cracking Index and PSR values.
- Re-evaluate weighting factors for different crack classes (types and location) and severities of cracking used for the PSR routine.

Chapter 8 Summary, Conclusions, and Recommendations on Implementation and Further Research

Summary of Undertaken Research

The SPR-2297 project titled "Implementation of a 3-D Sensing Technology for Automated Pavement Data Collection in Connecticut" was awarded to the UConn CTI research team by the CTDOT in order to investigate the effect of changes in ARAN technology on reported metrics of pavement performance. The project started with conducting simultaneous runs of the three ARAN vehicles (Vans 7, 8, and 9) over 127 lane-km of roads representative of typical pavement surface conditions in Connecticut. Van 7 represented older ARAN technology that had been used over the last 15 years, whereas Vans 8 and 9 were equipped with newly purchased 3-D LCMS (Pavemetrics) laser systems and Ro-Line height sensors.

The myriad of collected measurements including longitudinal and transverse profiles and pavement surface images were reviewed and analyzed to establish reproducibility between the older and newer ARAN systems with respect to roughness, rutting, and cracking extent. Additional repetitive runs of Vans 8 and 9 were conducted on short 1 to 3-km segments to establish precision and accuracy of the newer equipment with respect to profile, roughness, rutting, and crack measurements.

The analysis of roughness, rutting, and cracking data was conducted with two types of PMS reports generated through Roadware Vision software (Fugro, Inc.). A more detailed project level analysis focused on values averaged (IRI, rutting, cross-slope, and grade) or summed up (crack lengths by orientation (type), location in the lane, and severity) over 5-m intervals. The network level analysis was performed on the aforementioned values and reported for 0.1-mile (~161 m) intervals.

The analysis of the collected data targeted statistical significance of differences in IRI, rutting, and crack length values reported by the older (Van 7) and newer (Vans 8 and 9) ARAN systems. In addition, the correlation and linear regression trends between the two systems were evaluated to determine the feasibility of using transfer functions and/or correction factors between them. The transfer functions were to be developed to ensure continuity with historically established trends for pavement performance in Connecticut.

Major Findings and Conclusions

Overall, this project revealed the new 3-D laser scanning technology produces a much larger dataset and ultimately a much greater degree of resolution on the profiles due to an increase in the number of data points measured in both longitudinal and transverse directions. This substantially larger dataset created a smoothing effect (averaging of hundreds of values in the old system versus thousands of values in the new system) and under-estimating of the rut depths by the new system, for example. The new height sensors and accelerometers were also found to be extremely sensitive to undesirable vehicle movements that occur in response to obstacles, such as manhole covers, railroad crossings, and bridge joints. Speed gradient, including low speed and high acceleration or deceleration also adversely affected the output. This led to reporting unrealistic IRI values for as much as 5 to 10-percent of roughness data. The same undesirable movements apparently affected 3-D profile measurements that were used to generate 3-D images of pavement surfaces, thus,

leading to differences in brightness, color, and range (depth) of images generated by the two supposedly identical ARAN systems (Van 8 and Van 9). Nevertheless, the research team believes that the precision and accuracy of the ARAN measurements can be positively improved by the adjustment of the settings and configurations of both hardware and software, the development of repeatable/reproducible system/vehicle operation routines, and the stricter use of quality assurance principles during data collection and data processing. Below are presented the major findings and conclusions provided separately for profile, roughness, rutting, and cracking analyses.

Longitudinal Profiles

Due to differing default settings for the ARAN height sensors, the longitudinal profiles were initially collected with different frequency by all three ARAN vehicles, which precluded a direct comparison of both profile and roughness data. This warranted the use of a post-processing algorithm (Roughness Processor) developed by Fugro, Inc. for the newer ARAN data (Vans 8 and 9) to ensure comparability of measurements. Note that the old ARAN data (Van 7) did not require any post-processing.

The newer ARAN accelerometers were found to be extremely sensitive to operation at low speed and with braking (acceleration/deceleration) of the vehicle, as well as to undesired vertical movements caused by localized roughness, such as unadjusted manhole covers, railroad crossings, and bridge joints. The above-cited conditions led to abnormally high and unrealistic reported elevation and IRI values. One of the new ARAN vehicles (Van 9) exhibited greater sensitivity to localized roughness and acceleration/deceleration events than the other (Van 8).

On short sections (less than one mile in length) having no stops, slowdowns or obstacles, a relatively high correlation of 80 to 90 percent was achieved between the older and newer ARANs, while a correlation of 90 percent was achieved between the two newer ARAN systems. The degree of correlation decreased with an increase in length of section and where greater variability in surface roughness existed.

Surface Roughness (IRI)

Analysis of 5-m average IRI values (before post-processing) obtained from the older and newer ARAN-reported roughness revealed an overall poor correlation (R-Squared of 0.42), with a relative error of 62 percent between the two systems. The correlation between the two newer ARAN systems was much better (R-squared of 0.75 with 37-percent relative error) but lower than the desired 95 percent with 5-percent relative error (*24*).

After post-processing via the FUGRO Vision Roughness Processor and eliminating unrealistically high IRI values (higher than 630 in/mi [10m/km] (a limit found to be used in some other states)), the network-level IRI data (average per 0.1 mile) yielded comparable IRI values for all three ARAN vehicles (within 96 to 99 percent of each other). This translated to reproducibility between the older and newer ARANs of around 20 in/mi (0.3 m/km) and 25 in/mi (0.4 m/km) for the left (LWP) and right (RWP) wheelpaths, respectively. The use of mean IRI on the network level yielded a reduction in standard error between the older and newer ARANs, with accuracy of 13 in/mi (0.2 m/km) at R-squared of 93 percent. Post-processing greatly improved the comparability of the three vans for IRI.

The maximum achieved reproducibility, with respect to network-level IRI measurements between the older and newer ARAN systems, was lower than expected. This is likely due to differences in technology and operation, such as, but not limited to the following:

- Differences in sensitivity of profiling instrumentation to low speed vehicle operation with braking and localized roughness.
- Differences in the precision of distance and height measurements.
- Lateral deviations from the prescribed travel path (wandering within the travel lane).

The precision and the reproducibility between the two newer ARAN vans were evaluated on three short in-service roadway sections by analyzing five consecutive runs with each van over the same path. The following resulted:

- On relatively smooth sections with a mean roughness index (MRI) of 105 to 115 in/mi, average precision ranged within 2 to 3 in/mi in both wheelpaths (5 to 6 in/mi at 95-percent confidence). On the rougher sections with MRI of 170 to 180 in/mi, only 10 to 12-in/mi precision (14 to 25 in/mi at 95-percent confidence) could be achieved.
- The reproducibility between the two newer ARAN vans was very high (2 in/mi and 5 in/mi on smooth and rough surfaces, respectively, at 95-percent confidence). However, in the absence of ground truth IRI measurements as a reference, it is difficult to comment on the accuracy of the measurements.

Based on the above findings from data collected in November 2016, it can be concluded that the network IRI measurements by the newer ARAN systems (Vans 8 and 9) are in agreement with each other where a speed of at least 25 mi/hr is maintained as per ASTM E950 (*6*). The precision and accuracy of IRI measurements is highly dependent on precision and accuracy of collecting longitudinal profiles in accordance with standard procedures mentioned earlier.

Transverse Profiles

Overall, it was not possible to correlate the transverse profile data reported by the older and newer ARAN systems due to the following reasons:

- Differences in technology (ultrasonic vs. laser height measurements, 10-cm vs. 1-cm transverse data resolution for older vs. newer ARANs).
- Difference in width of collected profiles within the lane (2.4 m for the older ARAN vs. 4.2 m for the newer ARANs).
- Difference in longitudinal collection interval (5 m for older vs. 0.1 m for newer ARANs).
- Misalignment of reported transverse profiles due to vehicle wandering right and left within the lane (results in a lateral shift of the identified center of lane).

The analysis of cross-correlation between transverse profiles that were reported by the newer ARAN systems (Van 8 and 9) revealed a consistent “rotation” of the transverse profiles obtained with Van 8 by 0.7 to 1 percent slope. Furthermore, possible misalignment of left and right LCMS lasers/cameras was indicated by differences in correlation between left and right halves of profiles reported over the same cross-section. With the absence of reference ground-truth profile

measurements, it is difficult to identify the exact reasons for the discrepancies, such as hardware issues vs. errors in the processing algorithm, which is proprietary to Fugro, Inc.

The uncertainty of the transverse profile data may also have contributed to the poor precision and accuracy of cross-slope and rut depth values reported by the Roadware Vision software, as discussed next.

Cross-Slope Data

Pavement cross-slope or cross fall (vertical rise or fall per unit of horizontal distance (typically reported in percent)) is an important safety characteristic of road geometry, as well as a measure of its surface drainage performance. Comparison of reported average cross-slope values over 5-m intervals between the newer and older CTDOT ARAN systems revealed overall good correlation (R-Squared of 0.92) with no bias between Van 7 and Van 9 data. The main factor of variability in reported slopes was related to measured smoothness (IRI). It was noticed that the smoother the surface the higher the reproducibility was between the older and newer ARAN systems for cross slope.

The Van 8 cross slope data consistently indicated a bias of one (1) percent slope compared with both vans' 7 and 9 cross-slope values, which was most likely due to the issues with the transverse profiles stated earlier. This bias propagated to the network-level data. It should be noted that the cross-fall (XFALL) data from the vehicle orientation databases was found almost identical for all three ARAN vehicles. These XFALL values were found to be very close to historically reported ones in the CTDOT PMS and, therefore, they are recommended to be used for the purpose of comparison with historical data.

Average Rut Depth Data

Analysis of project-level data (5-m reporting interval) revealed very poor correlation between rut depth values reported by the older and newer CTDOT ARAN systems. On average, vans 8 and 9 reported 0.12-in (3-mm) deeper ruts as compared with Van 7. A fundamental difference in the measurement technology and computation algorithm for transverse profiles is believed to be the primary reason for the inability to correlate project-level rut depth values reported by the older and newer CTDOT ARAN systems.

On the network level, averaging rut depths over 0.1-mile increments did not show noticeable improvement of cross-correlation between the two newer CTDOT ARAN vehicles. The values of standard error between the newer (vans 8 and 9) and older (van 7) ARAN systems ranged between 0.03 and 0.07 in (0.7 to 2 mm) in the right and left wheelpaths, respectively. Considering a mean rut depth of 0.4 in (10 mm) reported in this study, those errors translate to 5 to 20 percent on average, which brings them to 10 to 40 percent at 95-percent reliability (double standard error).

The absolute precision of rut-depths reported by the 3-D LCMS at the project level (5-m reporting interval) ranged between 0.02 and 0.04 in (0.5 to 1 mm), which translate to 13 to 27 percent of mean rut value measured over 5 runs (32 to 65 percent at 95-percent confidence). This level of precision is deemed acceptable. At the network level (0.1- mi reporting interval), absolute and relative precision was 0.02 in and 25 percent at 95-percent confidence. The reproducibility

between the two new ARAN systems was found to be 0.03 and 0.06 in (0.8 and 1.7 mm) at 95-percent confidence in the left and right wheelpaths, respectively.

To conclude discussion on transverse profile characteristics, it should be noted that the AASHTO PP70 standard for collecting transverse profiles (*26*) and AASHTO R48 and PP69 for determining rut depth (*26, 27*) have been adapted for use with automatic data collection. The resultant algorithms are embedded into the Roadware Vision software and they differ significantly from the historical procedures, thus precluding direct comparison of historical and newly collected rut depth values.

Cracking Data and Pavement Surface Images

Table 8.1 shows the differences between the data collection systems for Van 7 and Vans 8 and 9. In recognition of the above differences, a direct correlation between WiseCrax and 3-D LCMS data was not attempted. The analysis of cracking data focused instead on comparison (correlation) of automatically detected crack reports from WiseCrax and 3-D LCMS with visual estimates of reference ground truth from surface images.

Table 8.1 Summary of Differences Between ARAN Crack Measurement Systems

	Van 7	**Van 8 and 9**
Capturing Pavement Image	2-D Strobe Black&White Video camera	3-D Laser Scan gray-scale camera
Method of image creation/capture	Stitch six 0.83-m long tilt images together to create 5-m long.	Continuous 3-D coordinate data is converted to gray-scale image and divided by 10-m lengths.
Image Analysis Method	WiseCrax algorithm detects variations in brightness and contrast for 2-D B & W images	Roadware Vision LCMS algorithm detects variations in surface profile depth
Crack Classification	Longitudinal and Transverse linear orientation,	Longitudinal and Transverse linear as well as area patterns (block and alligator cracking)
Crack Severity	3 and 6 mm thresholds for low/medium and medium/high	6 and 12 mm thresholds for low/medium and medium/high

Performing the analysis of cracking from WiseCrax and 3-D LCMS data was a challenge due to the following:

- The WiseCrax and Roadware Vision software provide different lane widths for the analysis (3.22 and 3.39 m, respectively).
- The newer ARAN images were frequently misaligned (laterally shifted) with respect to the older ARAN images, most likely, due to not travelling in the same path.
- The visual estimates of cracks from WiseCrax and Roadware Vision images varied due to differences in image brightness and contrast.

- Neither WiseCrax nor Roadware Vision automatic detection capabilities matched data extracted with human eyes.
- The Roadware Vision detection algorithm was extra sensitive to the presence of secondary cracks, thus, reporting significantly higher estimated quantities of area cracking.

The initial direct comparison of network-level (total length per 0.1- mile) cracking for a matched 40 lane-kilometers showed 3-D LCMS reported 2.4 times more longitudinal and area cracking, yet 5 times less transverse cracking on average, compared with the 2-D WiseCrax system. Overall, the 3-D LCMS system reported 40 percent more total crack length than WiseCrax. The reliability of this analysis did not exceed 50 percent.

With respect to visual estimates of ground truth, both WiseCrax and 3-D LCMS demonstrated an overall low level of detection (30 and 50 percent of total crack length on average, respectively). The 3-D LCMS system reported estimates closer to visual ground truth estimates on sections with fair to poor surface condition.

The precision and reproducibility of the two new 3-D LCMS systems (Van 8 and Van 9) with respect to reported cracking were estimated by simultaneously running the two ARAN vehicles five times over the same 3-lane-km stretch of roadway. This segment exhibited a wide range of conditions, with low (less than 40 ft/0.1-mi) to high (>1000 ft/0.1mi) cracking extent.

Both newer ARAN vehicles demonstrated overall high precision and accuracy of reported total crack lengths per 0.1-mile (10 percent at 95-percent confidence). The absolute and relative values for precision and accuracy were also computed for different crack classes and categories (longitudinal, transverse, and area cracks) within five lane zones (right and left lane edges, center, and right and left wheelpaths) and three crack severities (low, medium, and high). The precision results differed as follows:

- With respect to crack orientation, both Van 8 and Van 9 demonstrated a similarly high precision of 10 percent (reported at 95-percent confidence) for longitudinal cracks, and lower precision for transverse and area crack measurements (21 and 13 percent, respectively).
- Where specific lane zones (as defined in column 4 of Table 3.3) were analyzed, the precision varied significantly with fair results in wheelpaths and center zones (15 to 30 percent at 95 –percent confidence) and unsatisfactory results for the edges (45 to 90 percent at 95-percent confidence). This outcome was most likely due to wandering (lateral shift) between the two ARAN vehicles, which prevented the systems from consistently detecting cracks near the pavement edges.
- Both Van 8 and Van 9 demonstrated a fair level of precision with respect to rating crack severity, with 16 percent for low and medium severities and 37 percent for high severity, all at 95-percent confidence.
- It was found that highly cracked sections (>1000 ft/0.1-mi) produced both better repeatability and reproducibility as compared with sections with low cracking extent (<100 ft/0.1-mi)

The reproducibility between Van 8 and Van 9 with respect to crack reports by crack classification (extent by crack type), location (zone within the lane), and severity was similar to their individual precision and varied as follows (all values are reported at 95-percent confidence):

- On sections containing a high level of cracking (>1000 ft/0.1-mi), the accuracy ranged between 11 and 18 percent for longitudinal and transverse/area cracks, whereas on sections with very little or no cracking it was worse (with deviations of 90, 90, and 45 percent from the mean for longitudinal, transverse, and area cracking, respectively). These deviations for pavements with very low or no cracking should not be a cause for concern since they would all be categorized as good which renders these deviations as insignificant.
- With respect to location within a surveyed lane on significantly cracked sections, 16 to 40 percent accuracy was reported in the wheelpaths and center with 35 to 70-percent accuracy in right and left edge, respectively.
- On sections with noticeable levels of cracking (>100 ft/0.1-mi), 16, 21, and 37-percent accuracy of crack ratings was determined for low, medium, and high-severity cracks, respectively. The variability on sections with less than 100 ft/0.1-mi was higher (60 to 80 percent).

The above findings provide a basis for the need of verifying that the ARAN equipment and Roadware Vision software is compliant with existing and newly issued standard procedures for cracking distress survey, such as AASHTO R55, PP67 and PP68 (*28, 29, 30*). Ultimately, certification and calibration of the ARAN equipment in accordance with ASTM E1656 (*31*) should lead to improvements in precision and accuracy of distress measurement as compared with what was achieved during this project.

Recommendations on Implementation

Longitudinal Profile Measurements

- The RoLine height sensors and accelerometers should be re-calibrated on both the CTDOT ARAN vans (8 and 9) using an established reference value (ground truth) for pavement surface roughness obtained on a control/validation site.

- The settings in the on-board roughness calculation software, such as frequency of data collection and wavelength filter parameters, should be verified and set to be identical on both CTDOT ARAN vans (where appropriate) and compliant with AASHTO standards such as ASTM E950 and AASHTO R56 (*6, 24*) as well as AASHTO M328, R43, and R57 (*32, 33, 34*)

- The precision and accuracy of longitudinal profile surveys using the CTDOT ARAN vans should be re-evaluated on pre-approved control/validation sections in accordance with standard procedures listed above (*6, 24, 32, 33, 34*).

Surface Roughness (IRI) Reports

- For the time being, the IRI values averaged per 0.1-mile sections that are higher than 630 in/mi (10m/km) should be reviewed to determine if the pavement section is extremely rough or if they are outliers and should be eliminated from both project- or network-level reports.

- In order to preserve the continuity of historical IRI data, it is recommended to determine ground-truth IRI on some representative sections that were surveyed under this project in order to establish realistic correction factors (if any found) for both older (Van 7) and newer (Vans 8 and 9) CTDOT ARAN systems. While this may not be ideal since the Van 7 data will be several years old, it still provides a level of comparison between the old technology and new technology.

Transverse Profiles and Cross-Slopes

- The transverse profile data generated by the current version of the Roadware Vision software should be used cautiously until the issues with displaying and reporting correct transverse profiles in the Roadware Vision interface are resolved.

- In order to preserve continuity of cross-slope data, reporting cross-slope data as computed by the Roadware Vision should be avoided. Instead, crossfall (XFALL) values from vehicle orientation (VEHICLEORIENTATION) databases should be used to report cross-slope for PMS purposes. Note that all three ARANS report virtually identical XFALL values with very high reliability of at least 97 percent.

- Cross-slope units in the future PMIS reports should be checked for consistency, as the preferable units of measure are percent slope (100 times rise over run) as in historical PMIS reports.

- The LCMS equipment on both of the newer ARAN vehicles (Van 8 and 9) should be checked for proper and consistent alignment to avoid measurement of transverse profiles in an incorrect position of the lane, and, therefore, to also avoid misreporting or misclassifying rutting and cracking.

Rut Depth Measurements

- At this moment, rut depth data generated through Roadware Vision algorithms is not directly comparable with historical rut depth data due to differences in computational methods.
- Rut depth data should be validated against reference measurements using control/validation sites where physical measurements of the rut depths have been taken.

Cracking Surveys

This project leads to a conclusion that no transfer functions can be used to compare historical (pre-2015) and newly collected (2016 and later) cracking extent data in their current format. However, it is believed that 2-D WiseCrax and 3-D LCMS cracking outputs can be compared after conversion to common units of measure, compliant with existing provisional standards AASHTO PP67 and PP68 (*29, 30*). Those common units of measure should facilitate converting the crack lengths and areas to percent area cracks as prescribed by the new (December 2016)

HPMS Field Manual (*35*). Ultimately, the settings and configuration of the detection, classification and rating algorithms embedded in Roadware Vision software should be re-evaluated to ensure a good match with ground truth measurements. Much of the above will require additional research as outlined in the next section.

Overall Quality of Automatic Survey Data

- An important factor leading to reduced quality of data was inconsistency in data collection caused by driving outside the normal wheelpaths (offset center of lane)
- On roadway segments where significant variations in speed occurred, the longitudinal profile data was also compromised.
- The above-mentioned factors led to a significant portion of measurements (5 to 10 percent) identified as outliers and/or unrealistic values. This refers to both IRI and crack lengths reported by both Van 8 and Van 9.
- In addition to the above, significant portions of collected data could not be used for analysis due to issues with the newer ARAN hardware, such as inconsistent default settings for sensors and misalignment of lasers/and cameras, for example.
- Another ongoing project SPR-2309 is targeting improvement of the overall quality of the automated pavement survey through the development of quality management procedures.

Statement of Further Research Needs

In order to reflect the major findings and recommendations presented above, some additional research is warranted as proposed below:

- Investigation into cracking detection, classification, and rating algorithms embedded in the Roadware Vision software in order to verify their compliance with recommended standard procedures AASHTO R55, PP67, and PP68 (*28, 29, 30*). In addition, both historical (pre-2015) and newly collected (after 2015) crack extent measures (e.g., summed length vs. directional extent and percent area) should be re-evaluated to ensure comparability between historical and future distress reports as well as their compliance with FHWA reporting requirements in the HPMS Field Manual (*35*).
- Further investigation into the transverse profile measurement and rut depth calculations should be considered in order to improve the quality of the data. This research is warranted due to significant changes in computing the transverse profiles and the rut depth calculation algorithm embedded in Roadware Vision software to meet AASHTO standards PP70, R48, and PP69 (*25, 26, 27*). Determination of ground-truth transverse profiles is recommended to be done on sections representative of all rutting levels experienced in Connecticut, which were not available for this project.
- It is believed that current pavement performance models in Connecticut are based on the older ARAN technology and the revised PSR rating originally developed during the 1980s, and therefore, are obsolete. Research is needed to re-evaluate components of the current PSR metric and develop a new condition index that will be able to incorporate the new ARAN technology output, as well as meeting FHWA reporting requirements in the HPMS Field Manual (*35*).

References

1. Miller, J.S. and Bellinger, W.Y., 2003, Distress Identification Manual for the Long-Term Pavement Performance Program (Fourth revised Edition). Report No, FHWA-RD-03-031, Federal Highway Administration, USDOT, McLean, VA.

2. Papagiannakis, A., Gharaibeh, N., Weissmann, J., and Wimsatt. A., 2009, Pavement Scores Synthesis. Report No. FHWA/TX-09/0-6386-1, Texas Department of Transportation, Austin, TX.

3. McGee, K.H., 2004, Automated Pavement Distress Collection Techniques: A Synthesis of Highway Practice. NCHRP Synthesis 334, Transportation Research Board, Washington, DC.

4. Connecticut Department of Transportation (CTDOT), 2009, On the Move, Performance Metrics Report. http://www.ct.gov/dot/lib/dot/documents/dpublications/ctdot_on_the_move_performance_measures_report_011409.pdf, Accessed July 20, 2017.

5. Perera, R.W., Kohn, S.D., and Rada, G.R., 2002, LTPP Manual for Profile Measurements Operational Field Guidelines. Version 4.0, Federal Highway Administration, USDOT, McLean, VA.

6. American Society for Testing and Materials (ASTM), 2004, ASTM E950-98 Standard Test Method for Measuring the Longitudinal Profile of Traveled Surfaces with an Accelerometer Established Inertial Profiling Reference. ASTM International, West Conshohocken, PA.

7. Sayers, M.W. and Karamihas, S.M., 1998, The Little Book of Profiling. University of Michigan, Ann Arbor, MI.

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

9. Chang Arbitres, C.M., Smith, R.E., and Pendleton, O.J., 2007, Comparison of Automated Pavement Distress Data Collection Procedures for Local Agencies in San Francisco Bay Area, California. Transportation Research Record No. 1990, Transportation Research Board, Washington, DC

10. Vavrik, W., Evans, L., Sargand, S., and Stefanski, J., 2013, PCR Evaluation – Considering Transition from Manual to Semi-Automated Pavement Distress Collection and Analysis, Final Report. Ohio Department of Transportation, Columbus, OH.

11. Larsen, D.A., Sime, J.M., and Hudson, J.H., 1987, Theory to Practice: Photolog Laser VideoDisc and Its Application to Pavement Management in Connecticut. Second North American Conference on Managing Pavements, November 2-6, 1987, Toronto, Ontario, Canada.

12. Chan, S., Cui, S., and Lee, S., 2016, Transition from Manual to Automated Pavement Distress Data Collection and Performance Modelling in the Pavement Management System. Presented at the 2016 Conference of the Transportation Association of Canada, Toronto, Ontario, Canada.

13. Timm, D.H. and Turochy, R.E., 2014, Pavement Condition Model Based on Automated Pavement Distress Surveys. Alabama Department of Transportation, Montgomery, AL.

14. Wang, K.C.P., 2000, Designs and Implementations of Automated Systems for Pavement Surface Distress Survey. Journal of Infrastructure Systems, Vol. 6, pp 24-32.

15. Radoupoulu, S.C., Brilakis, I., Doycheva, K., and Koch, C., 2016. A Framework for Automated Pavement Condition Monitoring. Construction Research Congress 2016, May 31–June 2, 2016 | San Juan, Puerto Rico.

16. National Cooperative Highway Research Program (NCHRP), 2017, NCHRP Project 01-60 (anticipated), Calibration and Verification of Pavement Surface Images, Transportation Research Board, Washington, DC. http://apps.trb.org/cmsfeed/TRBNetProjectDisplay.asp?ProjectID=4350 Accessed August 3, 2017.

17. Serigos, P.A., Prozzi, J.A., Nam, B.H., and Murphy, M.R., 2012, Field Evaluation of Automated Rutting Measuring Equipment. Report No. FHWA/TX-12/0-663-1, Texas Department of Transportation, Austin, TX.

18. Fernando, E.G. and Walker, R.S., 2013, Impact of Changes in Profile Measurement Technology on QA Testing of Pavement Smoothness – Technical Report. Report No. FHWA/TX-13/0-6610-1, Texas Department of Transportation, Austin, TX.

19. Reggin, A., Shalaby, A., Emanuels, R., 2008, Urban Considerations for Using Road Roughness To Manage Road Networks. 7th International Conference on Managing Pavement Assets, June 23-28, 2008, Calgary, Alberta, Canada

20. Tomiyama, K., Nakamura, H., Mashito, H., Jomoto, M., and Watanabe, K., 2017, The Accuracy of Road Surface Profilers in TRUE Project. Paper 17-03150, Presented on 96th annual TRB meeting, Transportation Research Board, Washington, DC.

21. NCHRP Project 10-93(study active), Measuring, Characterizing, and Reporting Pavement Roughness of Low-Speed and Urban Roads. Transportation Research Board, Washington, DC. http://apps.trb.org/cmsfeed/TRBNetProjectDisplay.asp?ProjectID=3404 Accessed August 3, 2017.

22. Federal Highway Administration, Transportation Pooled Fund Study 5(063), (study active). Improving the Quality of Pavement Profiler Measurement. FHWA, USDOT, Washington, DC.

23. U. S. Government (USG), 2017, Code of Federal Regulations, Title 23 – Highways, Chapter I, Federal Highway Administration, Department of Transportation, Subchapter E –

Planning and Research, Part 490 – National Performance Management Measures, Subpart C - National Performance Management Measures for the Assessing Pavement Condition, Subsections 301-319, (23 CFR §490.301-319), U. S. Government Publishing Office, Washington, D. C.

24. American Association of State Highway and Transportation Officials (AASHTO), 2014, AASHTO R56-14 Standard Practice for Certification of Inertial Profiling Systems, AASHTO, Washington, DC.

25. AASHTO, 2017, AASHTO PP70-14 Standard Practice for Collecting the Transverse Pavement Profile, AASHTO, Washington, DC.

26. AASHTO, 2013, AASHTO R48-10 Standard Practice for Determining Rut Depth in Pavements, AASHTO, Washington, DC.

27. AASHTO, 2017, AASHTO PP69 Standard Practice for Determining Pavement Deformation Parameters and Cross Slope from Collected Transverse Profiles, AASHTO, Washington, DC.

28. AASHTO, 2013, AASHTO R55-10 Standard Practice for Quantifying Cracks in Asphalt Pavement Surfaces, AASHTO, Washington, DC.

29. AASHTO, 2017, AASHTO PP67-16 Standard Practice for Quantifying Cracks in Asphalt Pavement Surfaces from Collected Pavement Images Utilizing Automated Methods, AASHTO, Washington, DC.

30. AASHTO, 2017, AASHTO PP68-14 Standard Practice for Collecting Images of Pavement Surfaces for Distress Detection, AASHTO, Washington, DC.

31. ASTM, 2000, ASTM E1656-94 Standard Guide for Classification of Automated Pavement Condition Survey Equipment, ASTM International, West Conshohocken, PA.

32. AASHTO, 2017, AASHTO M328-14 Standard Specification for Inertial Profiler, AASHTO, Washington, DC.

33. AASHTO, 2013, AASHTO R43-13 Standard Practice for Quantifying Roughness of Pavements, AASHTO, Washington, DC.

34. AASHTO, 2014, AASHTO R57-14 Standard Practice for Operating Inertial Profiling Systems, AASHTO, Washington, DC.

35. FHWA, 2016, Highway Performance Monitoring System: Field Manual, FHWA, USDOT, Washington, DC.