



**THE UNIVERSITY OF TEXAS AT AUSTIN  
CENTER FOR TRANSPORTATION RESEARCH**

**0-6663-P2**

**RECOMMENDATIONS FOR SELECTION OF AUTOMATED  
DISTRESS MEASURING EQUIPMENT**

Pedro Serigos  
Maria Burton  
Andre Smit  
Jorge Prozzi  
MooYeon Kim  
Mike Murphy

*TxDOT Project 0-6663: Evaluation of Pavement Rutting and Distress  
Measurements*

**JANUARY 2014; PUBLISHED AUGUST 2014**

**Performing Organization:**

Center for Transportation Research  
The University of Texas at Austin  
1616 Guadalupe, Suite 4.202  
Austin, Texas 78701

**Sponsoring Organization:**

Texas Department of Transportation  
Research and Technology Implementation Office  
P.O. Box 5080  
Austin, Texas 78763-5080

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



## 1. Introduction

Phase 2 of Texas Department of Transportation (TxDOT) Research Project 0-6663: *Evaluation of Pavement Rutting and Distress Measurements* had the objective of evaluating the accuracy and precision of the new automated system developed by a TxDOT research group (composed of staff from the Construction Division's Materials and Pavement Section) for the high-speed measurement of pavement surface distresses, texture, and cross slope. In addition, equipment vendors participated in the study by providing equipment that represents the state of the practice—the automated distress collection vehicle. This equipment is used by other state DOTs for visual distress data collection through either vendor contracts or direct purchase. The implementation of an automated distress measuring system will allow the assessment of the highway condition at both the network and project levels and potentially eliminate the need for manual visual assessments to rate pavement distresses for network-level Pavement Management Information System (PMIS) applications. Eliminating any subjective elements in visual rating leads to more consistent and reliable data. Consistent and reliable data on the Texas road network will enhance pavement management and, ultimately, allow better utilization of ever-decreasing funds and overall state resources.

As part of this evaluation, the TxDOT system was compared to that of three automated system vendors in order to identify the best equipment for each pavement management data type. This comparison yields the information necessary for the researchers to help TxDOT in further evaluating the selection of automated distress measuring equipment. The high-speed measurements reported by each of the four automated systems that participated in the Phase 2 experiment were compared to manual measurements taken statically by experienced raters. The Phase 2 experiment comprised twenty 550-ft-long pavement test sections, including both flexible and rigid pavements, which were selected to represent the main pavement characteristics encountered on the Texas highway network.

The analyses of the automated measurement of surface distresses considered all distress types defined in the TxDOT PMIS and the Long-Term Pavement Performance Program (LTPP) protocols. Special focus was placed on the analyses of alligator cracking, longitudinal cracking, transverse cracking, patching, and failures. The only distress type not analyzed in Phase 2 was rutting, which was evaluated during Phase 1. The recommendations for the selection of an automated rut measurement system were provided in the project's product P1: *Recommendations for Selection of Rutting Measuring Equipment* (presented as Chapter 5 of report 0-6663-1).

Phase 2 analyses also included a qualitative comparison between the crack maps produced by the different automated systems at highway speeds and digital crack maps collected statically by manual measurement of the cracks. A comparative analysis of the digital crack maps allowed the researchers to obtain deeper insight into each system's quality of measurements and identify sources of error that cannot be detected by evaluation of summary statistics alone. For instance, this analysis allows for detecting cases for which the number of missed cracks practically compensated for false positives, producing apparently good overall summary statistics, thus creating misleading data for the interpretation of a system's true performance capabilities.

The current state of the practice in automated collection of pavement surface distresses is that, in general, transportation agencies have to choose between prompt delivery of results and enhanced accuracy. Faster distress data delivery is achieved by reporting the distresses detected and classified by the system's algorithms with minimal or no manual processing or corrections. Enhanced quality of results is achieved by the intervention of trained personnel that visually

inspect and correct the automated data produced by the system's algorithms. Since an ideal system would produce results with no need of further corrections, each service provider was requested to report their results with different levels of manual intervention in order to capture the difference in accuracy and identify the common types of errors produced by the systems' algorithms. In addition, since some technologies and algorithms were the best for certain types of distresses (e.g., the best system for detecting alligator cracking might not have been the best for detecting patching), the different distress types were analyzed separately, as carried out for texture and cross slope measurements. The recommendations for the selections of distress measuring equipment are based on the individual assessments for each distress type and time frame (level of manual processing), and the qualitative comparison of digital crack maps.

For Phase 3 of the project, two service providers (vendors) collected full network level data as per TxDOT PMIS specifications on the entire network in the Bryan District and in the Houston District. TxDOT asked the research team to evaluate the automated data collected, focusing on network-level processes and applications. As directed by TxDOT, the baseline data for this analysis will be the standard data collected for PMIS using the current methodologies that support the TxDOT PMIS. This phase is currently underway.

## **2. Summary Findings**

The researcher team selected twenty 550-ft field sections located in the Austin and Waco TxDOT Districts, distributed according to surface type into 15 asphalt concrete pavements (ACPs), 2 jointed concrete pavements, and 3 continuously reinforced pavements. Among the ACP test sections, the surface types were distributed as seven hot-mix asphalts, seven surface treatments, and one permeable friction course surface. The types of data collected at each section were distresses, texture, cross slope, and crack maps. The first types of data collected were manual measurements of distresses according to PMIS and LTPP protocols, conducted by two crews of raters with extensive years of practical experience in the respective protocols; longitudinal distribution of cross slope values using a FACE® Dipstick inclinometer; and longitudinal distribution of surface texture values for each wheel path, using a circular track meter (CTM).

Once the first set of manual measurements was completed, the different participating systems collected automated measurements at highway speeds on every test section of the study. In addition to the TxDOT 3D system, the following three vendors participated in the experiment: Dynatest (with an INO LCMS), Fugro-Roadware (with an INO LCMS), and Waylink-OSU (with an in-house developed 3D system). In order to capture the difference in accuracy for different levels of manual intervention, every participant was requested to report each data type within the following three different time frames:

- Fully automated with no manual post-processing, for data delivered at the end of a data collection run with no post-processing by the vendor;
- Semi-automated with minimum manual post-processing, for data delivered within 2 business days from the date that the vendor completes data collection on the last test section; and
- Semi-automated with higher manual post-processing, for data delivered within 4 weeks from the date the vendor completes data collection on the last test section.

Data reported for the first time frame represents the accuracy offered by the vendor if prompt delivery is a priority, whereas data reported for the third time frame represents the most accurate data interpretation possible. The number of days for each time frame was defined upon agreement with TxDOT and the participating vendors during a webinar conducted by the research team.

The last piece of data collected for Phase 2 experiment consisted of reference digital crack maps for the qualitative analyses. The crack maps were manually collected by the researchers at three 50-ft subsections per test section at 10 of the 20 test sections, thus collecting a total of 30 reference crack maps for the analyses. These subsections were selected in order to obtain sample cases for all main experimental variables in the study. The researchers marked the cracks visually detected in the field and categorized them into the following severity levels (crack width): cracks less than 3 mm (.12 in.) wide, between 3 mm and 6 mm (.24 in.) wide, and more than 6 mm wide. The crack width was measured using metallic rulers to determine the correct width category. Once the cracks were marked for the entire length of analysis, the next step consisted of taking digital pictures of the pavement surface every 5 ft. The reference digital photos taken by the research team were collected with a high-end digital camera mounted to a steel frame mounted to the front bumper of a truck. The camera was linked to a laptop operated from inside the truck, which provided controls to trigger the camera and collect the images. The steel frame was designed such that the camera was mounted approximately 12 ft above the pavement, pointing directly downward, and was therefore able to take photos of the entire lane width with minimal lens distortion. These individual pictures were further stitched and processed in order to obtain a unique digital crack map per subsection.

### **2.1. Digital Crack Maps**

The crack maps reported by the participants were evaluated qualitatively, as requested by TxDOT, by comparing them to digital crack maps manually collected with the objective of assessing the capabilities of the systems to properly detect cracks and identify their severity level to an accuracy consistent with the needs of the Department and the objectives of the research project. The comparative analyses of the digital crack maps allowed the researchers to detect patterns and sources of error that cannot be detected solely by analyzing the summary statistics.

Only one participant, TxDOT, reported digital crack maps within the first time frame (just after data collection). TxDOT, however, decided to not submit crack map data for the other two time frames (with manual processing) because the TxDOT automated equipment team considers this data as the most realistic and so the most appropriate for our analyses. Waylink-OSU reported crack maps for all test sections within the second time frame (minimum manual processing) but decided to not submit a dataset with higher manual corrections since they did not consider it necessary for improving the accuracy of their product. Dynatest and Fugro-Roadware reported crack maps with both minimal post-processing (within 2 days) and higher post-processing (within 4 weeks). Waylink-OSU, Dynatest, and Fugro-Roadware expressed that the crack maps data they reported for the second time frame consist of automated results produced by the systems' algorithms without manual correction. According to these three vendors, the processing performed during the 2 days after collecting the data was limited to only the amount of processing necessary for reporting the digital crack maps in the format requested by the researchers.

Among the crack maps with minimal or no manual post-processing, it was observed that TxDOT and Waylink-OSU tended to miss cracks more than reporting false positives regardless

of the surface type, whereas Dynatest and Fugro-Roadware presented maps with cases of both missed cracks and false positives. Therefore, TxDOT and Waylink-OSU system's algorithms tended to underestimate the crack lengths, TxDOT being the participant with the largest number of missed cracks. On several flexible pavements Waylink-OSU outperformed the other participants at detecting cracks; however, they tended to overestimate the crack width. In addition, Waylink-OSU was the only system that did not misidentify transverse or longitudinal joints on rigid pavements as cracks. The amount of missed cracks was greater for cracks less than 3 mm (.12 in.) wide for all participants and surface types. The very fine cracks observed on the rigid pavements were not detected by any automated system. The number of false positives observed from the Dynatest and Fugro-Roadware automated datasets was larger for the case of flexible pavements. In addition, TxDOT and Dynatest presented false positives caused by misinterpreting features such as vegetation, spots with different colors, and rumble stripes.

The automated results generated by Fugro-Roadware and Dynatest systems' algorithms were greatly improved after applying manual correction. In addition, the dataset corresponding to the third time frame also included types of distresses that were not reported in the fully automated data deliveries, such as patching and raveling. These observations show that applying manual processing to automated results even for current state-of-the-art equipment can improve distress detection and elimination of false positives identified by automated algorithms. However, another interesting observation was that manual corrections performed visually by trained raters were also a source of error in some cases. It should be noted that the vendors were not constrained to providing the detailed, manual ratings at 4 weeks (the maximum allowed for the third set of data) if results could be delivered sooner. However, the research team notes that both Dynatest and Fugro-Roadware used the full 4-week period for manual processing of thirty 50-ft-long test subsections. This suggests that manual interpretation of an entire pavement network might be time consuming and will be interesting to evaluate in Phase 3 of this study for the Bryan and Houston Districts.

## **2.2. Distresses Statistics**

The distresses statistics reported by each participant and by experienced manual raters were compared for every type of distress with the objective of identifying the differences and similarities among the different systems and to observe the changes between the fully automated (or with minimal manual post-processing) and semi-automated results. Due to the cost and time delays associated with developing software systems by the vendors to collect and report distress data in the TxDOT PMIS protocol and for TxDOT to report distress data in the LTPP protocol, it was not possible to compare all four participants directly. The three vendors, which already have data collection software and protocols for LTPP data, were evaluated according to the LTPP protocol whereas TxDOT was evaluated using the PMIS protocol. Each system was compared to statistics manually collected according to the corresponding LTPP or PMIS protocols by experienced raters. As for the case of digital crack maps, TxDOT delivered distress statistics only for the first time frame; Waylink-OSU did it only for the second time frame; while Fugro-Roadware and Dynatest reported summary statistics for the second and third time frames.

The comparative analyses among Waylink-OSU's, Fugro-Roadware's, Dynatest's, and manual raters' LTPP distress statistics were performed for each type of distress separately. There was no clear pattern between manual measurements and the different vendors for any type of distress in flexible pavements except for the case of transverse cracking, for which a good match among the manual raters', Waylink's, and Fugro-Roadware's semi-automated dataset was

observed. The lack of correlation for the majority of distresses types is explained, in part, by the differences in criteria used for distress classification; e.g., longitudinal cracks in the wheel path that were correctly detected by all participants might have been classified by either the vendors or manual raters as fatigue cracking rather than longitudinal cracking. In addition, it was observed that Waylink-OSU and Fugro-Roadware semi-automated datasets consistently reported fewer numbers and smaller patch sizes while Fugro-Roadware fully automated Dynatest's fully and semi-automated datasets did not report patching.

Interestingly, Dynatest and Fugro-Roadware (both using INO LCMS) consistently increased the fatigue cracking area after manual intervention by a similar proportion for every flexible pavement. This observation suggests that the Dynatest and Fugro-Roadware systems' algorithms tend to systematically underestimate fatigue cracking. Examples of other significant differences between the fully and semi-automated datasets were a decrease in Fugro-Roadware's transverse cracks and an increase in Dynatest's edge cracking for flexible pavements. Regarding the analysis of rigid pavements, the number and length of transverse cracks drastically increased after manual intervention for both Fugro-Roadware and Dynatest data sets. However, no clear pattern was observed for the different types of cracking among the vendors and manual raters. Also, Fugro-Roadware and Dynatest reported the different types of spalling and joint damage only for the case of semi-automated data.

TxDOT distress readings were compared with the manual PMIS measurements. Only longitudinal cracking and transverse cracking could be reported by TxDOT's current automated equipment set-up, whether the section was asphalt pavement or concrete pavement. On many sections in which TxDOT values were significantly higher than the reference, values became closer to the reference values after TxDOT's sealed crack counts were removed by the researchers during data analysis and interpretation, thus counting only non-sealed cracks. This process was not intended to change the TxDOT analysis results, but only to help understand differences between the reported, automated data and manual reference data.

### **2.3. Texture**

Pavement texture was reported as the mean profile depth in mm, every 50 feet for each wheel path. Only TxDOT reported texture data just after data collection. The other three participants reported texture data for the second time frame (within 2 days of completing data collection). In most sections, Dynatest and Fugro-Roadware were close to the reference measurements taken by the research team using a CTM, whereas Waylink-OSU and TxDOT's reported average reading were usually higher in magnitude. Waylink-OSU followed a similar trend in shape as the reference.

### **2.4. Cross Slopes**

As for the case of texture data, only TxDOT reported cross slope for the first time frame while the three vendors reported their cross slope values within 2 days of collecting data. The cross slope values were reported every 50 feet, in units of percent. For most (19 out of 20) sections, Dynatest measurements closely match or follow a similar trend to the reference in the graph-line shape and, though for fewer instances, slope magnitude. Fugro-Roadware (12 out of 20 sections) and Waylink (7 out of 20 sections) sometimes match or partially match the reference graph-line shape, though they (Waylink-OSU more than Fugro-Roadware) exhibit variations above and below the reference slope magnitude. TxDOT's average cross slope readings were often close to the reference and the other vendors, though readings with the AASHTO PP69 algorithm were

often farther from the reference than the other two algorithms (two point and line fitting) reported by TxDOT.

### **3. Final Recommendations**

The University of Texas at Austin has completed Phase 2 of TxDOT Research Project 0-6663, *Evaluation of Pavement Rutting and Distress Measurements*. During this phase,

- A field experiment consisting of 20 sections was developed,
- Static manual distress statistics, texture, cross slopes, and digital crack maps were collected,
- Four participants were invited to collect automated distress, texture, and cross slope measurements at highway speeds, and
- The results were analyzed and compared to assess the difference between automated and manual measurements and evaluate the change in accuracy between fully and semi-automated results.

As a result of the Phase 2 efforts, the research team reached the following preliminary conclusions:

- Among the datasets reported within 2 days, it can be observed that, under the conditions evaluated, the Waylink-OSU outperforms the remaining participating systems in terms of crack detection. However, Waylink-OSU tended to overestimate the crack widths, suggesting the need for further adapting and calibrating the system's algorithms for Texas conditions.
- Dynatest and Fugro-Roadware showed a significant improvement in the accuracy of their distress measurements after applying manual post-processing consisting of visual interpretation and correction of the results produced by their systems' algorithms. Additionally, the results reported within 4 weeks included more types of distresses. These observations show the current need for applying manual interpretation to the automated results produced by state-of-the-art equipment.
- The TxDOT crack maps were missing a large number of cracks, suggesting the need for calibrating the algorithms in order to increase system sensitivity for detecting narrower cracks. The researchers noted that adjusting system sensitivity to find more cracks can also result in a greater number of false positives—this is a trade-off that each participant must consider when calibrating their crack detection systems. It is also suggested that TxDOT consider the development of algorithms to quantify crack widths and thus report crack severity levels.
- From the comparative analyses among the distress statistics reported by each participant and the manual raters, no clear, obvious patterns emerged for all types of distresses and time frames. Thus, the researchers could not identify one automated system that was clearly superior to the other. This lack of clear patterns is in part due to the use of different distress classification criteria. It is recommended that an objective and programmable standard or protocol be developed for classifying distresses from automated data in order to increase the consistency of results.



- Several types of distresses, such as patching, punchouts, spalling, and joint damage, were reported only after manual post-processing of the crack maps by Fugro-Roadware and Dynatest, whereas Waylink-OSU reported some of these types of distresses on the 2-day time frame.
- TxDOT did not provide data for all PMIS distress types due in part to ongoing work to improve distress identification algorithms and reporting methods. Apparently, additional time and effort is needed to refine the TxDOT system to provide fully automated, short-time-frame results.
- It is suggested that TxDOT could improve crack identification accuracy by differentiating between sealed and unsealed cracks. The number of sealed cracks reported by TxDOT often caused the crack count to be significantly higher than the reference. It could be that either TxDOT is over-counting the sealed cracks, or the reference is under-counting the sealed cracks.
- Dynatest and Fugro produced texture results close to the reference in magnitude with minor error. It is suggested that Waylink-OSU and TxDOT consider updating or calibrating their systems since all measurements presented were greater than the reference values. Note that TxDOT texture results were reported as an average value for each 550-ft section, which is equivalent to the 0.10-mile subsection length used to store and calculate PMIS rating sections values. Revising the TxDOT algorithm to report values on a 50-ft. interval could have resulted in a different conclusion.
- Dynatest reported cross slope measurement results closest to the reference in graph-line shape and magnitude. Fugro results are fairly close to the reference in magnitude at certain points, although the graph-line shape is not always close with the reference. Waylink-OSU can sometimes deliver a graph-line shape similar to the reference, although often the magnitude is higher or lower than the reference.
- TxDOT cross slope was evaluated with average values per entire section, so a precise comparison could not be evaluated. The researchers suggest that further work is needed to improve analysis of cross slope data based on the results from the three algorithms used by TxDOT.