JOINT TRANSPORTATION RESEARCH PROGRAM

INDIANA DEPARTMENT OF TRANSPORTATION AND PURDUE UNIVERSITY

Algorithm and Software for Proactive Pothole Repair

Leila Sadeghi, Yaguang Zhang, Andrew Balmos, James V. Krogmeier, John E. Haddock

SPR-3908 • Report Number: FHWA/IN/JTRP-2016/14 • DOI: 10.5703/1288284316337

RECOMMENDED CITATION

Sadeghi, L., Zhang, Y., Balmos, A., Krogmeier, J. V., & Haddock, J. E. (2016). *Algorithm and software for proactive pothole repair* (Joint Transportation Research Program Publication No. FHWA/IN/JTRP-2016/14). West Lafayette, IN: Purdue University. http://dx.doi.org/10.5703/1288284316337

AUTHORS

Leila Sadeghi
Graduate Research Assistant
Lyles School of Civil Engineering
Purdue University

Yaguang Zhang
Andrew Balmos
Graduate Research Assistants
School of Electrical and Computer Engineering
Purdue University

James V. Krogmeier, PhD
Professor of Electrical and Computer Engineering
School of Electrical and Computer Engineering
Purdue University

John E. Haddock, PhD, PE
Professor of Civil Engineering
Lyles School of Civil Engineering
Purdue University
(765) 496-3996
jhaddock@purdue.edu
Corresponding Author

JOINT TRANSPORTATION RESEARCH PROGRAM

The Joint Transportation Research Program serves as a vehicle for INDOT collaboration with higher education institutions and industry in Indiana to facilitate innovation that results in continuous improvement in the planning, design, construction, operation, management and economic efficiency of the Indiana transportation infrastructure. https://engineering.purdue.edu/JTRP/index_html

Published reports of the Joint Transportation Research Program are available at http://docs.lib.purdue.edu/jtrp/.

NOTICE

COPYRIGHT

Print ISBN: 978-1-62260-413-5
ePUB ISBN: 978-1-62260-414-2

TECHNICAL REPORT STANDARD TITLE PAGE

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
FHWA/IN/JTRP-2016/14		

4. Title and Subtitle	5. Report Date
Algorithm and Software for Proactive Pothole Repair	April 2016
	6. Performing Organization Code
7. Author(s)	**8. Performing Organization Report No.**
Leila Sadeghi, Yaguang Zhang, Andrew Balmos, James V. Krogmeier, John E. Haddock	FHWA/IN/JTRP-2016/14
9. Performing Organization Name and Address	**10. Work Unit No.**
Joint Transportation Research Program Purdue University 550 Stadium Mall Drive West Lafayette, IN 47907-2051	**11. Contract or Grant No.** SPR-3908
12. Sponsoring Agency Name and Address	**13. Type of Report and Period Covered**
Indiana Department of Transportation State Office Building 100 North Senate Avenue Indianapolis, IN 46204	Final Report
	14. Sponsoring Agency Code

15. Supplementary Notes

Prepared in cooperation with the Indiana Department of Transportation and Federal Highway Administration.

16. Abstract

Potholes are a common pavement distress, particularly appearing during the spring freeze-thaw period in northern climates. Potholes reduce ride quality, and if left unrepaired can lead to rapid pavement deterioration. Typically, when a pothole appears a repair crew is dispatched to place patch mixture in the hole with the hope that the patch will last until such time as a more permanent repair can be made. This reactive approach to potholes can often be too late to prevent further pavement damage and also makes it difficult for repairs crews to be scheduled in the most cost effective manner.

In this study, the relation between traffic loads combined with weather records, such as temperature, freeze-thaw cycles and the numbers of potholes requiring patching was investigated in an attempt to develop a model to predict pothole formation and distinguish the routes which are prone to pothole formation before the potholes begin to form. If pothole prediction were possible, this proactive approach would enable agencies to plan and schedule maintenance activities more cost and time effectively thus increasing ride safety and mobility.

To achieve the objective, four years of maintenance data from Indiana routes were collected and statistically analyzed to develop a model to estimate the probability of occurrence of a pothole due to annual average daily traffic and climate. The model indicates how significant traffic loads combined with weather condition influence the pothole. Also, although traffic loads and weather conditions are the essentials for potholes to form, the effect of pavement condition on the initiation of new potholes cannot be disregarded.

Additionally, this study began the development of a basic roadway distress evolution model by employing several standard statistical tools, such as, the empirical cumulative distribution functions (CDF) and the Kolmogorov-Smirnov (KS), to a pavement condition dataset. The goal of the model was to predict and rank areas of probable future concern by likelihood and severity. The resulting analysis showed promise but the data resolution was too low to achieve predictions on the desired fine scale.

17. Key Words	18. Distribution Statement
pothole, predictive models, pavement, AADT, weather, cumulative distribution function	No restrictions. This document is available to the public through the National Technical Information Service, Springfield, VA 22161.

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	30	

Form DOT F 1700.7 (8-69)

EXECUTIVE SUMMARY

ALGORITHM AND SOFTWARE FOR PROACTIVE POTHOLE REPAIR

Introduction

Potholes are a common pavement distress that appear particularly during the spring freeze-thaw period in northern climates. Potholes reduce ride quality and if left unrepaired can lead to rapid pavement deterioration. Typically, when a pothole appears a repair crew is dispatched to place patch mixture in the hole with the hope that the patch will last until such time as a more permanent repair can be made. This reactive approach to potholes often leads to further pavement damage and also makes it difficult for repair crews to be scheduled in the most cost-effective manner.

In this study an attempt was made to develop a model to predict pothole formation and determine which routes are prone to pothole formation before the potholes begin to form. If pothole prediction were possible, this proactive approach would enable agencies to plan and schedule maintenance activities in a more cost- and time-effective manner, thus increasing ride safety and mobility.

Findings

- Traffic load plays a more important role in the formation of potholes in the urban and rural Interstate highways compared to the urban US highways and urban and rural state routes.
- Temperature is more important than traffic loads in pothole formation on rural routes.
- Applying the developed models could help agencies assign maintenance priority to highways predicted to develop comparatively more potholes.
- Analyses suggest that current condition data resolution may not be sufficient to predict a pothole of only a few feet in size, but it may be sufficient to predict which road segments of about 1 mile in length will or will not experience distress leading to potholes.
- Currently the resolution of the repair database is too low to prove or disprove the results of the analyses conducted. Higher resolution repair records are needed.

Implementation

While the findings of this study are encouraging, there are currently no results that are ready for implementation. The following recommendations are offered as possible future work:

- High-resolution repair records could be recorded with a mobile approach, like the PPTracker developed in this research.
- Schemes to automatically flag and fix quality control issues in the pavement databases could be developed and implemented.
- Other Indiana Department of Transportation databases that include information such as pavement age, pavement type, traffic volumes, complete weather data, and so forth could be more directly integrated into the algorithms.
- When increased resolution data becomes available, a true pothole formation prediction algorithm could be developed.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

1. BACKGROUND

A pothole is defined as a bowl-shaped hole in the pavement surface with a minimum plan dimension of 150 mm (Miller & Bellinger, 2003). Over time moisture penetrates into the pavement through cracks or joints and accumulates beneath or within the pavement structure. As freeze-thaw cycles occur the expansion and contraction of the moisture, combined with other loads such as traffic, results in the formation of potholes (Kuennen, 2004). Potholes are one of the most common pavement distresses and often require expensive maintenance activities to repair the damage (McDaniel, Olek, Behnood, Magee, & Pollock, 2014). They reduce ride quality, pavement life, performance, and accelerate pavement deterioration (Paterson, 1987), all of which lead to increased pavement life-cycle costs.

Jimoh (2012) observed that traffic loads enlarge existing potholes at a rate of 8 cm^2 per cumulative 80 kN standard axle. However, other factors, e.g., precipitation, age, and pavement conditions also influence pothole and cracking initiation and progression (Mubaraki & Thom, 2012; Paterson, 1987). In order to more completely understand the current state of a given road segment and predict the segment's future state, a model for road deterioration is necessary. Given that most of the contributing factors to roadway damage are natural phenomena, and therefore are inherently random, it is reasonable to use a stochastic model that treats future states and inputs as random variables that can be conditioned on and averaged over.

The current state-of-the-art includes several models that predict pothole formation and progression (Jimoh, 2012; Morosiuk, Riley, & Odoki, 2004; Mubaraki & Thom, 2012; Paterson, 1987). Paterson (1987) conducted a study based on data collected in Brazil, St. Vincent in the Caribbean, Ghana, and Kenya to predict the initiation and progression of potholes from wide cracks or raveling. The model's relationships were later updated by Morosiuk, Riley, & Odoki (2004). A third model was developed by Jimoh (2012) from data gathered on flexible pavements in semi-urban arterial in Nigeria and found a relationship between the enlargement rate of pothole area and cumulative traffic loads. Mubaraki and Thom (2013) modeled pothole formation for main streets in Saudi Arabia and found age of the pavements to be the most significant variable in the progression of the potholes.

All the aforementioned models are based on empirical data collections dependent on particular environments, usage, and maintenance schemes. In particular the models focused on developing countries in tropical climates. Given that the design, construction procedures, and maintenance plans vary between countries, as well as the climate conditions, none of the existing models are applicable to the routes in colder United States (US) climates.

2. DATA SOURCES

Various data sources were analyzed for their potential use in the development of a pothole prediction model. This section describes the original sources and the types of data available.

2.1 Pavement Condition Data

Automated pavement condition data collected by Pathway Services, Inc. (Pathway) was used in this study to assess the road states that lead to pothole formation. Ultimately condition data for the years of 2012, 2013, and 2014 was available for analysis. However, the 2014 dataset was not considered in this section because it was still being collected at the time the analysis was done. Later sections of this report do include the 2014 dataset.

The Pathway data is discretized to 0.1-mile segments and includes the data points shown in Table 2.1.

Data for longitudinal and alligator cracking, spalling and corner breaking were recorded separately for different parts of the road. Figure 2.1 illustrates the longitudinal sections defined in the Pathway 2014 documents prepared for the Indiana Department of Transportation. The letters A to F represent centerline, left wheel path, center lane, right wheel path, edge and shoulder, respectively. The cracking data were further broken into Non-Wheelpath (NWP), Wheelpath (WP), Outside Zone/Edge (EG) and Shoulder (SD) averages. Non-Wheelpath data are the average cracking of a given type in the centerline and center lane, i.e., sections A and C, Wheelpath is the average cracking of a given type in the left and right wheel path, i.e., sections B and D, Outside Zone/Edge is the average cracking of a given type in the edge, i.e., section E, and finally the Shoulder is the average cracking over a given type in the shoulder, i.e., section F.

2.2 Pavement Maintenance Data

There are no INDOT records of direct pothole measurements for most roads in Indiana. As a proxy for this data, details of pothole repair activities were extracted from the INDOT Work Management System (WMS). This data is typically entered into the database by INDOT personal to keep record of equipment and labor hours, material use, and type and location of work done. When available, the quantity of patch material applied to a road segment was used as an estimate of the number potholes requiring patching.

The extracted maintenance data spanned roughly four years of activity on Indiana pavements starting from 2011 and ending midway through 2015. Three types of patching were included in the data set: deep patching, temporary shallow patching and permanent shallow patching. The pavement segment had varying lengths with a mean of about 6 miles and standard deviation of about 4 miles. For some pavement segments, INDOT personal included free-form text comments as metadata. However, since there was no consistency in these comments over the years or similar type of work they were not included in the analysis.

TABLE 2.1
Pavement condition description.

Data	Measurement Details
Start/end reference post	Post numbers are sometimes from the state system and sometimes for the country system
Direction	Increasing or decreasing reference post numbers
Date	
Pavement type	Only the surface layer. Composite roadways are not represented
IRI	Left wheelpath, right wheelpath, average
Rut depth	Left wheelpath, right wheelpath, average
Texture	Left wheelpath, right wheelpath
Faulting	Number, average depth
Longitudinal cracking	Length, average depth and width for each low, medium and high severities
Alligator cracking	Percent of lane area, average depth and width for each low, medium and high severities
Transverse cracking	Number, average depth and width for each low, medium and high severities
Block cracking	Percent of lane area, average depth and width for each low, medium and high severities
Longitudinal spalling	Number, average depth and width for each low, medium and high severities
Transverse spalling	Number, average depth and width for each low, medium and high severities
Corner breaking	Number, average depth and width for each low, medium and high severities
Start/end GPS coordinates	(Not available in 2014 condition data)

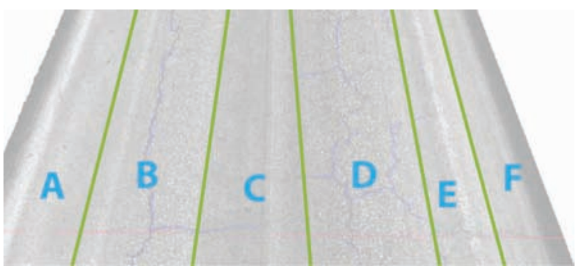


Figure 2.1 Longitudinal sections in a pathway. (Courtesy of Pathway Services Inc.)

2.2.1 Weather Condition and Traffic Data

Past weather conditions including precipitation and average daily and annual air temperatures were retrieved from the Weather Underground website (www.wunderground.com). MATLAB code was developed to calculate the number of freeze-thaw cycles based on the average daily air temperatures. A freeze-thaw cycle was defined to be when the air temperature decreases to less than 0°C (32°F) and then increases to a temperature above 0°C (32°F). Traffic data denoted as Annual Average Daily Traffic (AADT) were extracted from INDOT (n.d.) website.

2.2.2 Pavement Age and Structure

Other important factors affecting pavement deterioration rate and pothole formation are age and structure. Since this kind of data is not recorded in the pavement management system, it needed to be extracted manually for each roadway (see Appendix B). Unfortunately, the data was not available in this phase of the study.

3. PRELIMINARY ANALYSIS OF CONDITION, TRAFFIC AND WEATHER DATA

3.1 Pavement Condition, Traffic and Weather Condition

In this section, three Indiana pavement segments were selected at different geographical locations including SR 49 (MP 31-37) in the northwest, US 27 (MP 100-114) in the northeast and US 41 (MP 0-9) in the southwest. The locations were intentionally varied to enable assessment of its effect on pothole formation. Additionally these routes where selected because of the availability of PPTracker data recorded in 2015. Detailed information about PPTracker device and the recorded data is provided in Appendix A.

The objective of this particular analysis was to assess the quality of the PPTracker and the pavement condition and maintenance dataset. Additionally, it was hoped that possible relationships between the pavement condition data and the areas prone to pothole formation would be uncovered. The following summarizes the collected data for each segment:

- PPTracker device recorded data (2015)
- INDOT repair data (2011–2015)
- Pavement condition data (2012–2013)
- Traffic data (Annual Average Daily Traffic) (2011–2015)

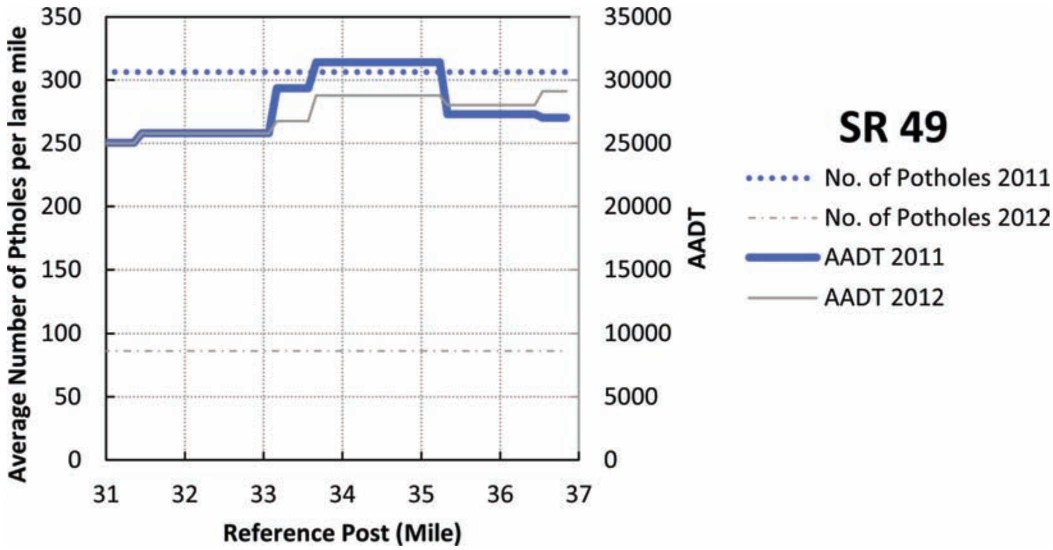


Figure 3.1 Average number of potholes and AADT in 2011 and 2012.

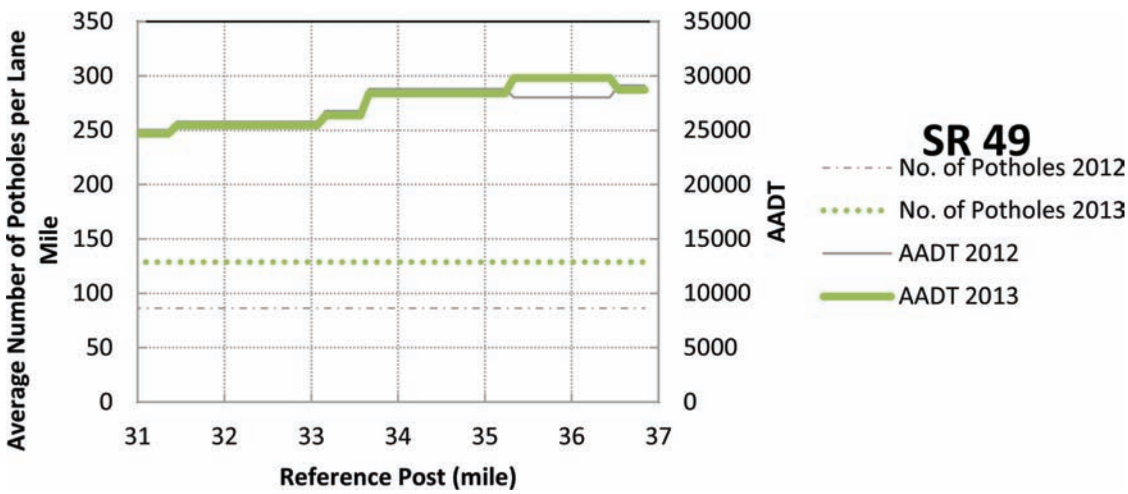


Figure 3.2 Average number of potholes and AADT in 2012 and 2013.

- Weather records (2011–2015) including mean annual precipitation, mean annual temperature and number of freeze-thaw cycles in a year

Assuming the average dimensions of a pothole with depth to diameter ratio of 1 to 6 and an average depth of 80 millimeter (Paterson, 1987), the repair data was used to calculate the average number of potholes requiring patching in each year for each road segment. This conversion is necessary to fairly compare the quality of available repair data with the precise record obtained from the PPTracker device.

The comparison of the average number of potholes per lane mile for each segment in different years indicates that decrease in traffic loads results in decrease in the average number of potholes per lane mile unless the weather conditions worsen. This implies the importance of both traffic loads and weather conditions on the pothole formation. In addition, it indicates that in spite of the low resolution and the low accuracy of the maintenance data, the amount of material per lane mile can reliably estimate the average number of potholes per lane mile in shorter segments. For example, Figure 3.1 shows that overall AADT decreased from 2011 to 2012, even though, the annual average temperature was higher in 2012 compared to 2011 with less precipitation (Figure 3.3). Thus, decrease in the average number of potholes observed on this segment (Figure 3.1). Figure 3.2 shows an overall increase in the segment's traffic load from 2012 to 2013. The average number of potholes per lane mile also increased (Figure 3.2) since there was an increase in precipitation, lower temperature, and a higher number of freeze-thaw cycles for this segment in 2013 compared to 2012 (Figure 3.3).

The pavement distresses were compared to the precise records of PPTracker to further assess the quality of the available pavement condition data and to investigate the possible correlation between pavement distress types and the number of potholes. For example, Figure 3.4 and Figure 3.5 show alligator cracking and block cracking on one pavement segment. Results indicate that

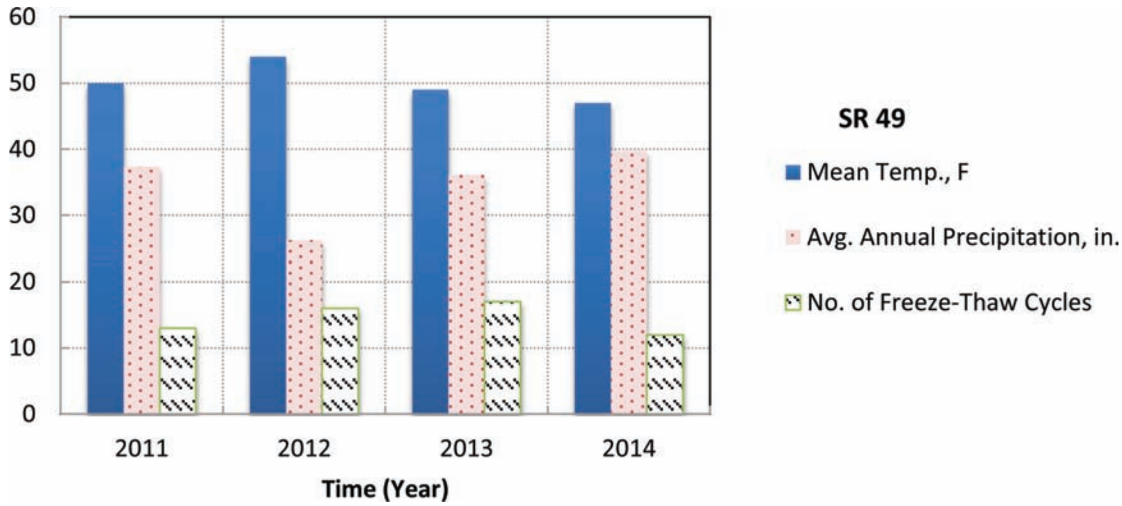


Figure 3.3 Weather conditions in 2011 to 2014.

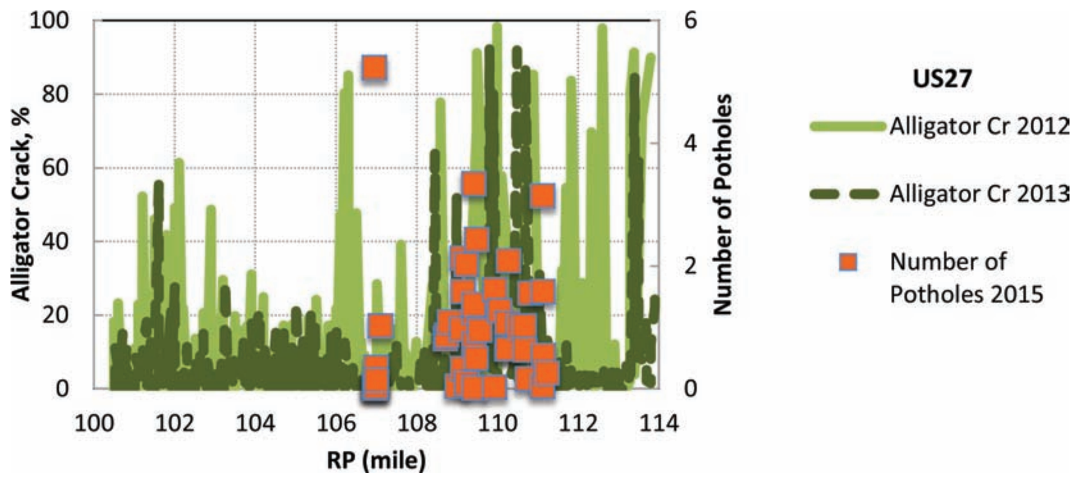


Figure 3.4 Alligator cracks and number of potholes.

potholes occur more frequently where there is higher percentages of alligator and block cracking. Although this correlation can clearly be observed for this segment it is not always so clear for past years data. It is strongly recommended that such an analysis be only carried out on the most recent pavement condition data.

3.2 Preliminary Modeling of Pothole Formation

3.2.1 Modeling of Pothole Formation Due to Traffic and Weather Conditions

The objective of this analysis is to investigate the relation between traffic loads, denoted as annual average daily traffic (AADT), weather data, such as temperature and freeze-thaw cycles, and the number of potholes requiring patching, in an attempt to develop a model to predict pothole formation. Such a model could potentially identify routes more prone to pothole formation before the potholes begin to form. Such information would enable agencies to proactively plan maintenance approaches and schedule activities in the most cost efficient and time effective manner.

To achieve the objective, four years of maintenance data from Indiana highways were collected and statistically analyzed. In particular, the work looked at the significance of AADT and climate conditions on the average number of potholes requiring patching per lane mile for various types of highways. The resulting model indicates the effect of traffic loads, temperature, and number of freeze-thaw cycles on the formation of potholes. Although traffic loads and weather conditions are essential for pothole formation, the effect of pavement condition on the initiation of new potholes cannot be disregarded.

3.2.2 Methodology

Estimates of the temperature variation and number of freeze-thaw cycles that road segments were exposed to were collected from weather data across four Indiana cities with different geographical locations. Those cities

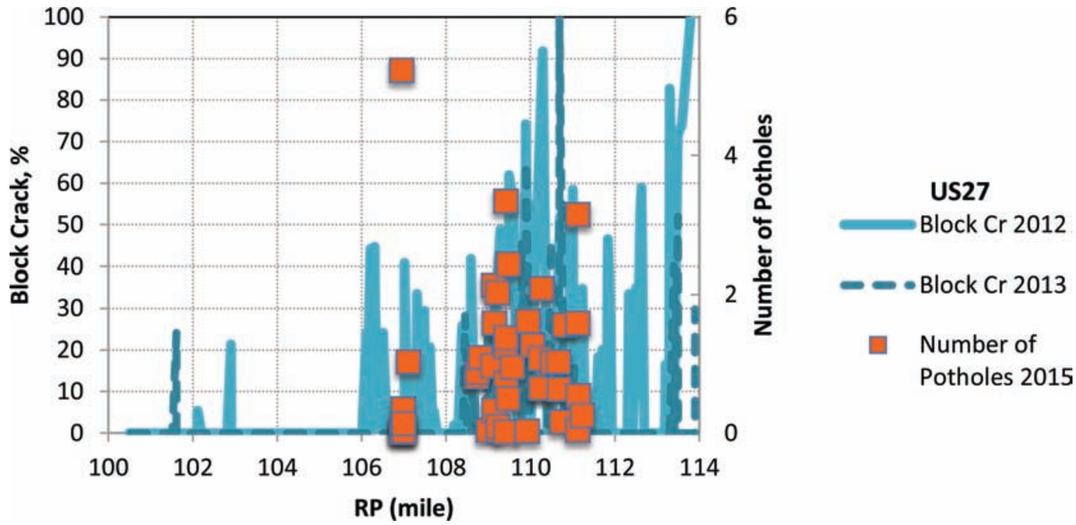


Figure 3.5 Block cracks and number of potholes.

are Gary in the northwest, Fort Wayne in the northeast, Columbus in the southeast, and Evansville in the southwest. Indiana Department of Transportation (INDOT) highways within fifteen miles of these cities were sorted, depending on functional class, into three groups: (i) interstate highways including both rural and urban; (ii) urban US highways and state roads (SR); and (iii) rural SR. While, generally, the first group carries the higher traffic volume compared to the other two groups, no specific range of AADT is assigned for each group since the traffic volume will vary within a group. The maintenance data from 2011 through 2014 were derived from the Work Management System (WMS) database recorded by INDOT. The maintenance data includes the amount of patching material used to patch the potholes on a specific highway length. To eliminate the effects of length and number of lanes for any given highway, the amount of patch material used on a given highway segment was divided by the length of the segment and the total number of lanes in the segment. The amount of patch material per lane mile thereby becomes an indicator of the average number of potholes per lane mile.

Average daily temperatures from 2011 through 2014 for the four cities were extracted from the weather underground website (www.wunderground.com). For purposes of analysis, a freeze-thaw cycle was defined as the event where the average daily air temperature fell to or below 32°F, with a subsequent increase to above 32°F. AADT from 2011 through 2014 for the highway segments were collected from INDOT traffic database.

Once the data were assembled, a probability density function was estimated for each of the three highway groups (Figure 3.6). The functions describe the probability that a particular patch intensity (patch material per lane mile) takes on a specific value. For example, Figure 3.6 indicates there is a 20 percent probability that the patch intensity is approximately 0.2 tons/mile for each of the three highway groups. However, for interstate highways, there is a 45 percent probability the

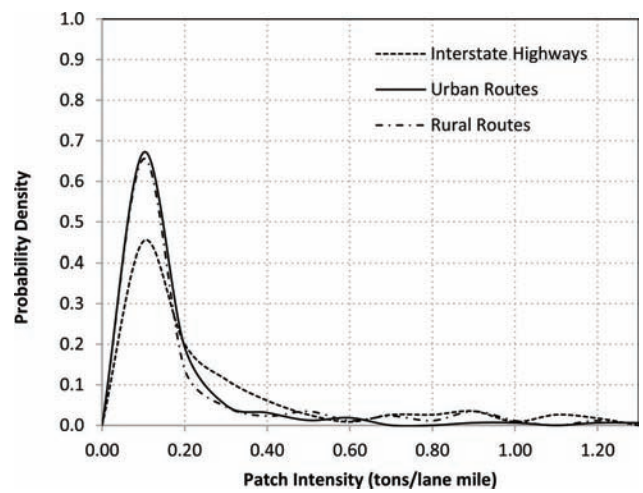


Figure 3.6 Probability density function.

patch intensity is only 0.1 ton/mile, while urban and rural routes have a 65 percent probability of the same patch intensity.

To make the comparison more straightforward, empirical cumulative distribution functions were calculated from the corresponding probability density functions of the three highway groups (Figure 3.7). The cumulative distribution function describes the proportion of the segments within a group that have a patch intensity less than or equal to a specific value. By way of example, Figure 3.7 shows that 65 percent of the interstate pavements have a patch intensity less than or equal to approximately 0.2 tons/mile, while the rural and urban routes have an 80 and 86 percent probability respectively of having that patch intensity. Thus interstate highway segments in the data set need relatively more patch material than do the rural and urban routes.

Four individual cumulative distribution functions were estimated over the variables: patch intensity, AADT, temperature, and the number of freeze-thaw cycles, for each segment in every highway group. From those

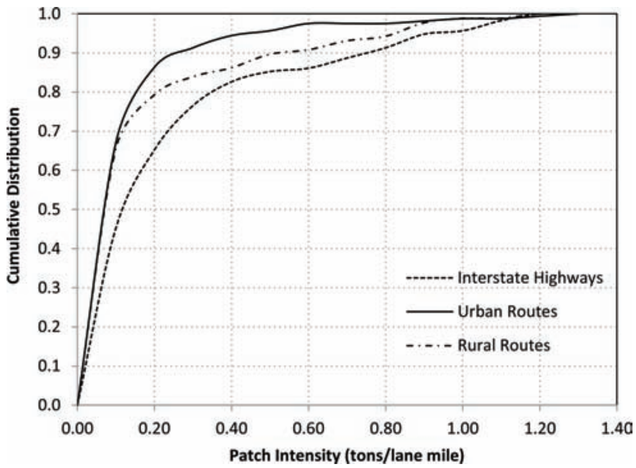


Figure 3.7 Cumulative distribution function.

estimated curves a software package called NLOGIT created an ordinary least-squares regression, resulting in a basic pavement state model. NLOGIT selects the independent variables based on the best correlation between the given model and the corresponding data. Finally, adjusted R-squared values and the Durbin-Watson statistic were computed and used as the criteria to select the best model. The Durbin-Watson statistic is a test designed to look for the presence of serial correlation as well as measure the randomness in the least squares residuals (Kutner, Nachtsheim, Neter, & Li, 2005). The further the Durbin-Watson statistic value is from 2, the less confidence there is in the autocorrelation of the data (Washington, Karlaftis, & Mannering, 2011).

3.2.3 Estimation Results

The models selected for three groups of highways are:

- Interstate highways

$$\ln(\Delta patch) = -6.30 + 0.53\ln(\Delta AADT) + 0.05(FT) + 1.07\ln(CD) \quad (3.1)$$

- Urban Highways

$$\ln(\Delta patch) = -1.46 + 0.50\ln(\Delta AADT) - 0.15(T) + 3.27(CD) \quad (3.2)$$

- Rural Highways

$$\ln(\Delta patch) = 5.61 + 0.0002(\Delta AADT) - 0.26(T) + 5.88(CD) \quad (3.3)$$

where $\Delta patch$ is the increment in the patch intensity (tons/lane mile) cumulative distribution function, $\Delta AADT$ is the increment in the AADT cumulative distribution function, CD is the cumulative distribution, FT is the number of freeze-thaw cycles, and T is temperature (F).

The models indicate that traffic volume is the most significant variable in the formation of potholes on interstate and urban highways as compared to rural routes. As seen in Equation (3.1), (3.2) and (3.3), ln($\Delta patch$) is changed by 0.53% of the change in ln($\Delta AADT$) for interstate highways, 50% for Urban highways, and only 0.02% of $\Delta AADT$ for rural highways. The models also indicate that frequency of freeze-thaw cycles have a more important role in deterioration of the interstate highways than it does on urban and rural routes; in fact, the number of freeze-thaw cycles is so irrelevant to urban and rural roadways that it is not even included in final a model. Conversely, temperature appears more likely to cause potholes in urban and rural routes, while the variable is not used for the interstate highway modeling.

All the variables mentioned in the models are significant at a 99 percent confidence level. The adjusted R^2 values are 0.56, 0.32 and 0.28 for the interstate, urban, and rural highways respectively. The adjusted R-squared indicated a high variability in the data set. However, if other variables that can affect patch intensity were included, e.g., pavement condition, distress severity, etc., the adjusted R-squared values would most likely increase.

For an example of an application, the rural model is illustrated in Table 3.1 and Table 3.2 for a SR route segment near Gary, Indiana, in 2012. The weather condition is the same for the entire segment. In the example, the increment in the cumulative distribution was assumed to be 0.1. While this increment could be any value, it is important that the assumed increment for the cumulative distributions be the same for $\Delta AADT$ and $\Delta patch$.

The estimated cumulative distribution functions, obtained using the data in Table 3.1 and Table 3.2, is plotted in Figure 3.8. The plot indicates that for the rural highways with the traffic loads and weather conditions

TABLE 3.1
Rural highway model example data.

CD (cumulative distribution)	AADT	ΔAADT	Temperature (F)
0.0	2707.4	2707.4	55.00
0.1	3092.7	385.3	55.00
0.2	3308.7	216.0	55.00
0.3	4448.9	1140.2	55.00
0.4	5358.5	909.6	55.00
0.5	6927.1	1568.6	55.00
0.6	8234.2	1307.1	55.00
0.7	9327.3	1093.1	55.00
0.8	10483.0	1155.7	55.00
0.9	11549.0	1066.0	55.00
1.0	14028.0	2479.0	55.00

TABLE 3.2
Rural highway model example results.

ln(Δpatch) (from Equation (3.3))	Δpatch Intensity	Patch Intensity
-8.1485	0.0003	0.0003
-8.0249	0.0003	0.0006
-7.4708	0.0006	0.0012
-6.6980	0.0012	0.0024
-6.1561	0.0021	0.0045
-5.4363	0.0044	0.0089
-4.9006	0.0074	0.0163
-4.3554	0.0128	0.0291
-3.7549	0.0234	0.0525
-3.1848	0.0414	0.0939
-2.3142	0.0988	0.1927

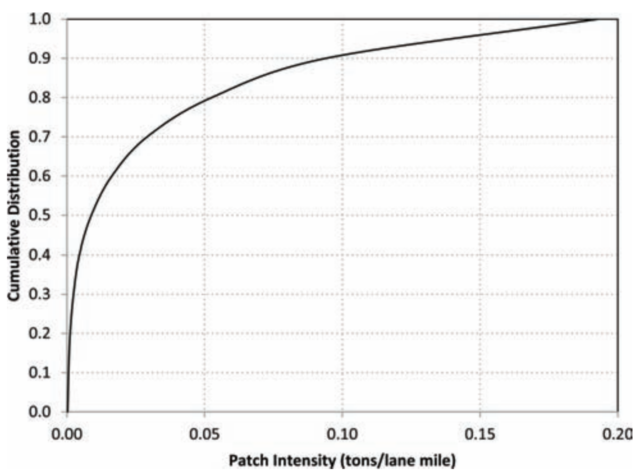


Figure 3.8 Estimated cumulative distribution function.

shown in Table 3.1, approximately 0.20 tons/mile of patch mixture is needed to patch the potholes.

3.2.4 Implications of Findings

Figure 3.9 through Figure 3.11 show the comparisons of actual data and the values estimated by the models using the average traffic loads and weather conditions for each of the three types of highways. In these figures, the solid line displays the cumulative distribution function for the actual data and the dashed line displays the cumulative distribution function for the model-estimated values.

Figure 3.9 shows the interstate highway model can estimate the patch intensity with less than 20 percent difference between estimated and actual data for the cumulative distribution up to 0.9. For the cumulative distribution higher than this, the model underestimates patch intensity. The high variability in the patching methods and severity of potholes on interstate highways could be the reason for this underestimation. In addition, since there are many variables that affect the patch intensity and the current model considers only traffic loads and freeze-thaw cycle. It is likely an improved model can be obtained by including additional variables.

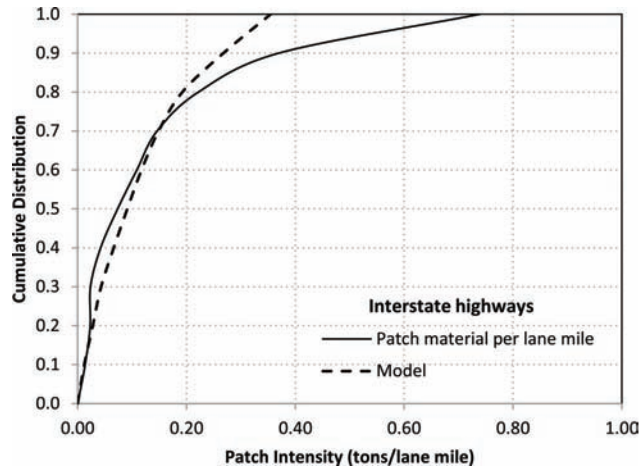


Figure 3.9 Evaluation of the interstate highway model.

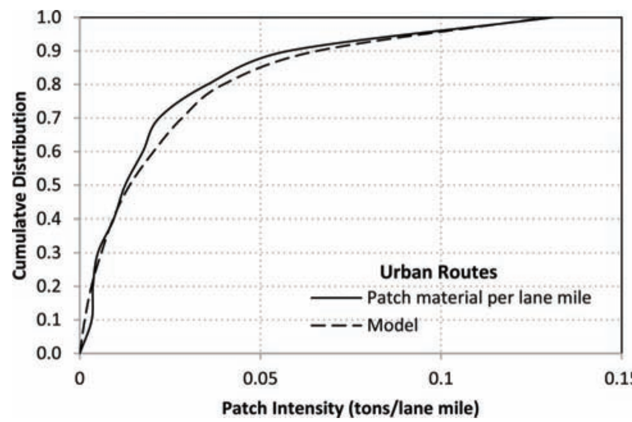


Figure 3.10 Evaluation of the urban highway model.

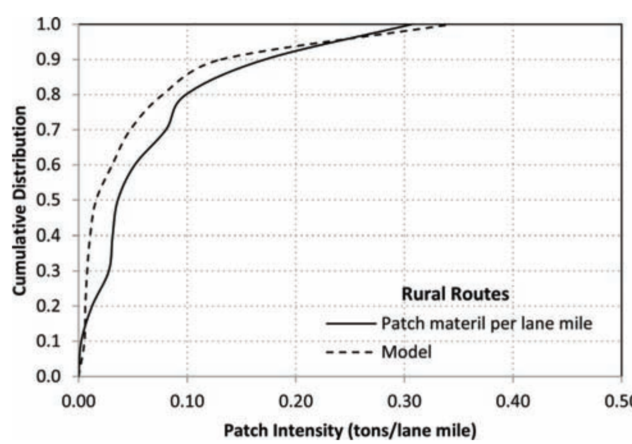


Figure 3.11 Evaluation of the rural highway model.

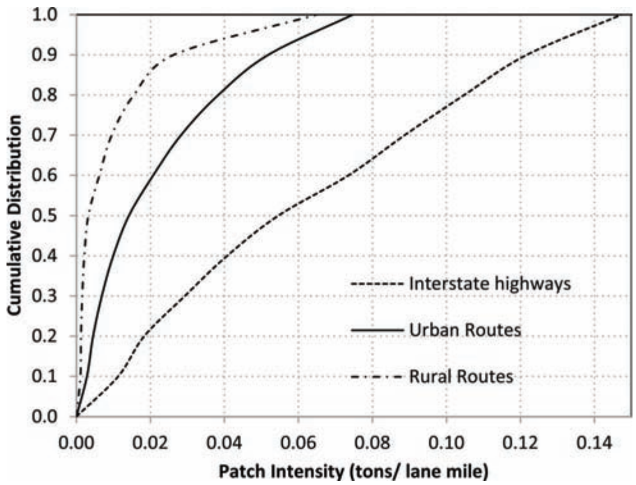


Figure 3.12 Comparison of highway types under the same conditions.

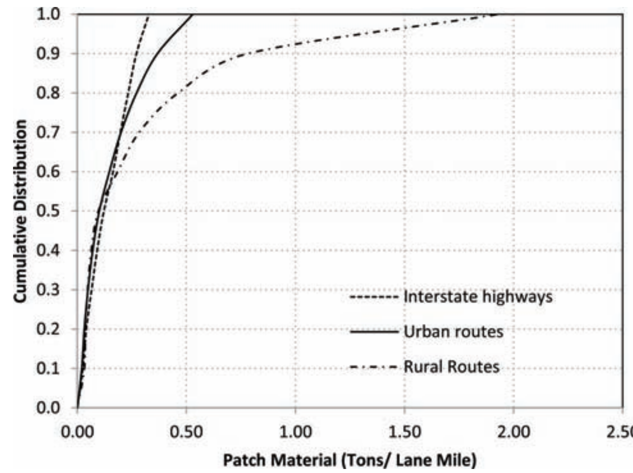


Figure 3.13 Comparison of highway types under the same condition (cold weather).

For urban highways the model is an excellent match to the real data, with differences of less than 10 percent (Figure 3.10). On average, the urban highway model over estimates the patch intensity by 7 percent, a negligible amount in practice.

Unlike the model for interstate highways, the model developed for rural highways (Figure 3.11) estimates with better precision at the higher cumulative distribution values. On average the rural highway model estimates 17 percent variance from the real values.

3.2.5 Implications

The model can also be used to distinguish the routes that are in a more severe condition than expected, i.e., more potholes have developed on the pavement then what the model suggests is normal. These roadways are candidates for more in-depth study and repair.

Figure 3.12 shows the comparison of the three highway groups under the same traffic loads (mean AADT of 11,098 vehicles per day), 60F temperature, and 16 freeze-thaw cycles. This simulates the average weather conditions in Evansville, Indiana for 2012. Overall the results indicate that the amount of material needed to patch the potholes that would have formed is less than 0.15 tons/lane mile. Under this condition, interstate highways need more patch material than US and SR highways do. This is because interstates are more sensitive to the number of freeze-thaw cycles rather than temperature. The average temperature is not terribly low and therefore does not have a critical effect on pothole formation for the US and SR highways.

Figure 3.13 shows another comparison of the three types of routes under the same conditions. Traffic load is assumed to be the same as the previous example (mean AADT of 11,098 vehicles), but the temperature (mean 48F) and number of freeze-thaw cycles (32) simulate colder weather. These conditions are similar to the average weather conditions in Fort Wayne, Indiana for 2014.

Figure 3.13 indicates that overall, more material is needed to patch potholes in colder weather for all three highway types; not totally unexpected result. More potholes are developed under these conditions when compared to the previous warmer environment for all three highway groups. Decreases in temperature have a more serious effect on the rural SR highways and therefore they degrade the worse among the three types.

3.3 Preliminary Analysis of the Patching Data

In this section, the focus is on three selected road segments to investigate the deterioration trends shown by different candidate parameters. The two statistical tools used to reveal the trends are the empirical cumulative density function (CDF) and the Kolmogorov–Smirnov (KS) test. Also computed from the maintenance dataset was the equivalent pothole rate (EPR) per lane-mile; it was used as the reference for road deterioration. By comparing the parameters available in the condition data, alligator cracking (AC) and block cracking (BC) were found most appropriate to be used for indicating pothole formations.

3.3.1 Organization of Data and Analyses

The segments chosen are SR 49 from MP 0 to MP 44, US 27 from MP 0 to MP 118, and US 41 from MP 0 to MP 31. The parameters of interest include Annual Average Daily Traffic (AADT), International Roughness Index (IRI), as well as AC, BC, and EPR. In order to show the trend in time, all available condition and maintenance data from year 2011 to 2015 are manually extracted for these segments.

Figure 3.14 shows the empirical CDF for SR 49 IRI in year 2012 and 2013, along with corresponding histograms in Figure 3.15. As can be seen by comparing the two figures, the empirical distributions are easier to compare year-to-year with the CDF curves than the data's histogram.

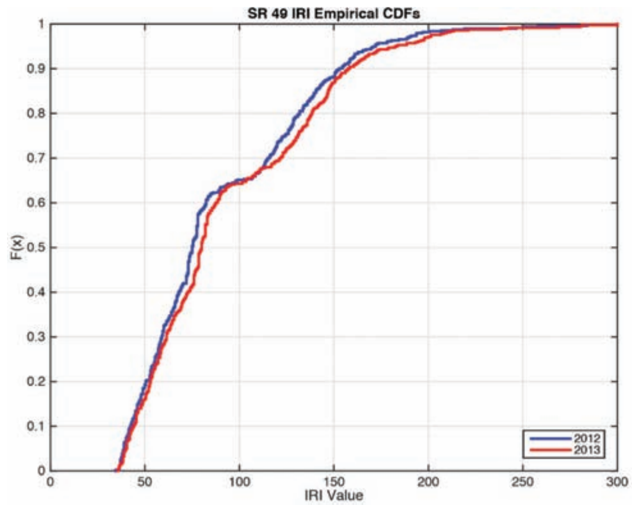


Figure 3.14 Empirical CDF for IRI of SR 49 in 2012 and 2013.

Two-sample KS tests statistically measure the difference between two empirical distributions, the so-called *p* value. For this work, the *p* value provides a simple measure of how the road state has changed from year-to-year. To simplify the metric further the *p* value is compared to a threshold to determine if the distribution has changed significantly or not between in the past year. Figure 3.16 illustrates the test procedure using two sample vectors of observed AC data from 2012 and 2013 for SR 49. The assumption that sample 1 and sample 2 come from the same distribution is called the null hypothesis of the test. The goal of the test is to accept or reject the null hypothesis according to the data. In this example, the null hypothesis says that AC distribution has not changed from 2012 to 2013. Rejecting the null hypothesis is the equivalent to saying that the AC distribution has changed.

The test threshold is referred to as the level of the test, and is normally denoted as α. It is the probability that the corresponding test will reject the null hypothesis, assuming that it is true. Typical values for level are 0.05, 0.01, 0.005, etc. The smaller the level is, the less likely a test will reject the null hypothesis. The *p* value is the probability that the corresponding pair could be observed, assuming the null hypothesis is true. If the *p* value of an observed pair of samples is less than the level of the KS test, then the null hypothesis is rejected.

Applying the KS test to each of the road parameters over the entire road segment may not yield high enough spatial resolution for sufficient predication. So instead, a small windows of observations that slides along with the data was applied to the test,, as shown in Figure 3.16. Here, a window length of 1 mile is selected for the KS test. Currently the highest resolution for the condition data is 0.1 mile. Therefore, a sliding window of 1 mile contains around 10 adjacent points.

Figure 3.17 shows the KS test results versus mile posts for AC on a segment of SR 49, along with a map of its nearby area. Note that the *p* value here equals 0.005 and the window length is 1 mile.

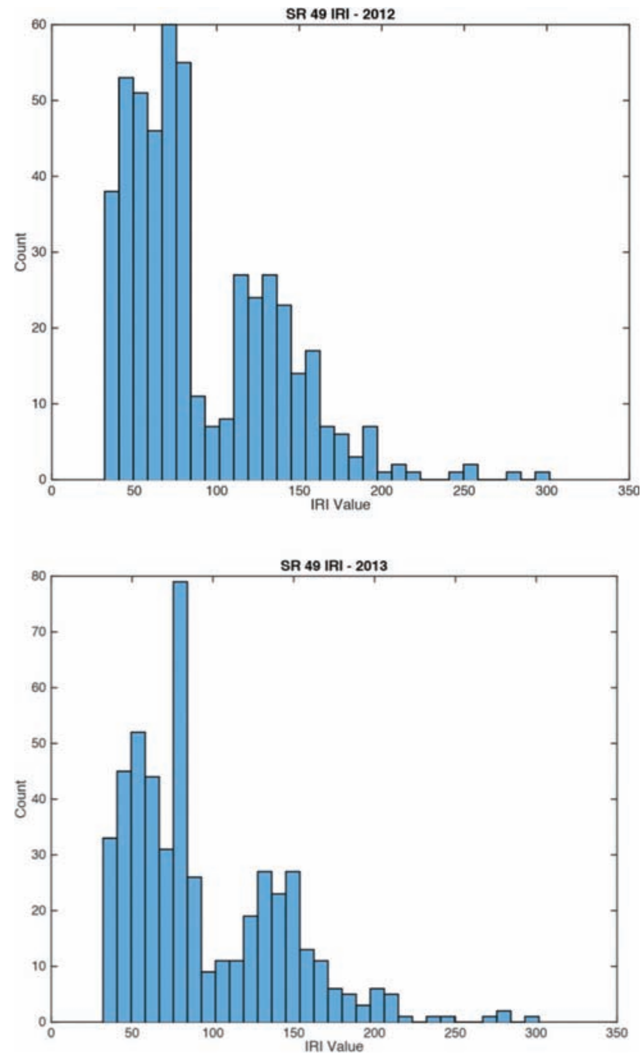


Figure 3.15 Histograms for IRI of SR 49 in 2012 and 2013.

3.3.2 Analysis Results

Other analyses for SR 49 are shown in Figure 3.18 to Figure 3.22. By fitting the data into statistical tools like empirical CDF plots and KS test, it is easier to grasp the indications of the dataset. First, IRI appears to be less affected by repair than does AC or BC, because Figure 3.19 and Figure 3.20 show a moderate improvement in road condition for SR 49, while Figure 3.14 only indicates a slight improvement. So IRI might be less predictive of future road condition. Second, the improvement in AC is much more significant than BC. This could be caused by repaving activities not indicated by the maintenance data, or the change of definition for BC through years. Meanwhile, Figure 3.17 and Figure 3.22 show the locations where AC and BC changed according to the KS test. By comparing them with Figure 3.18, we can see a rough relationship

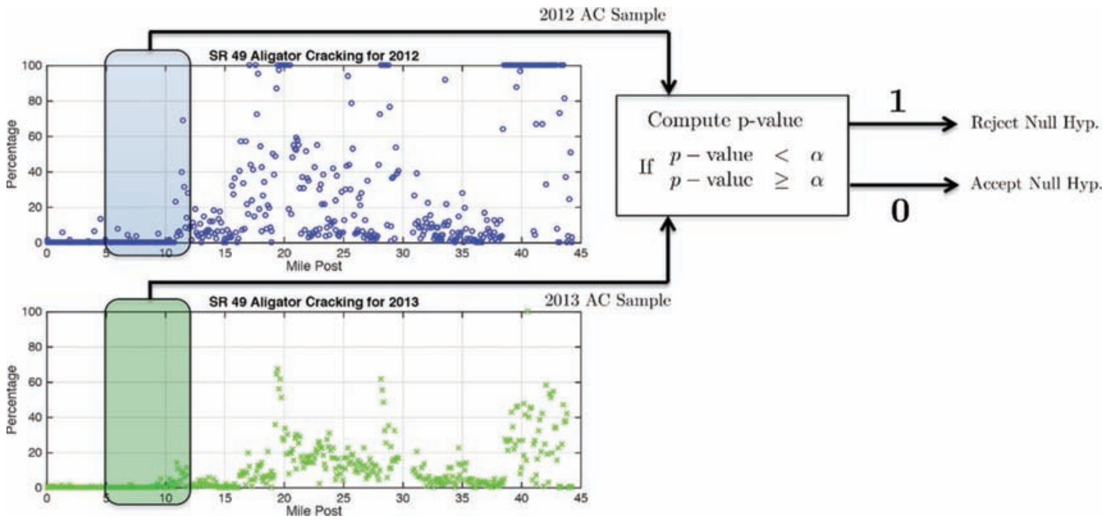


Figure 3.16 Illustration for KS test.

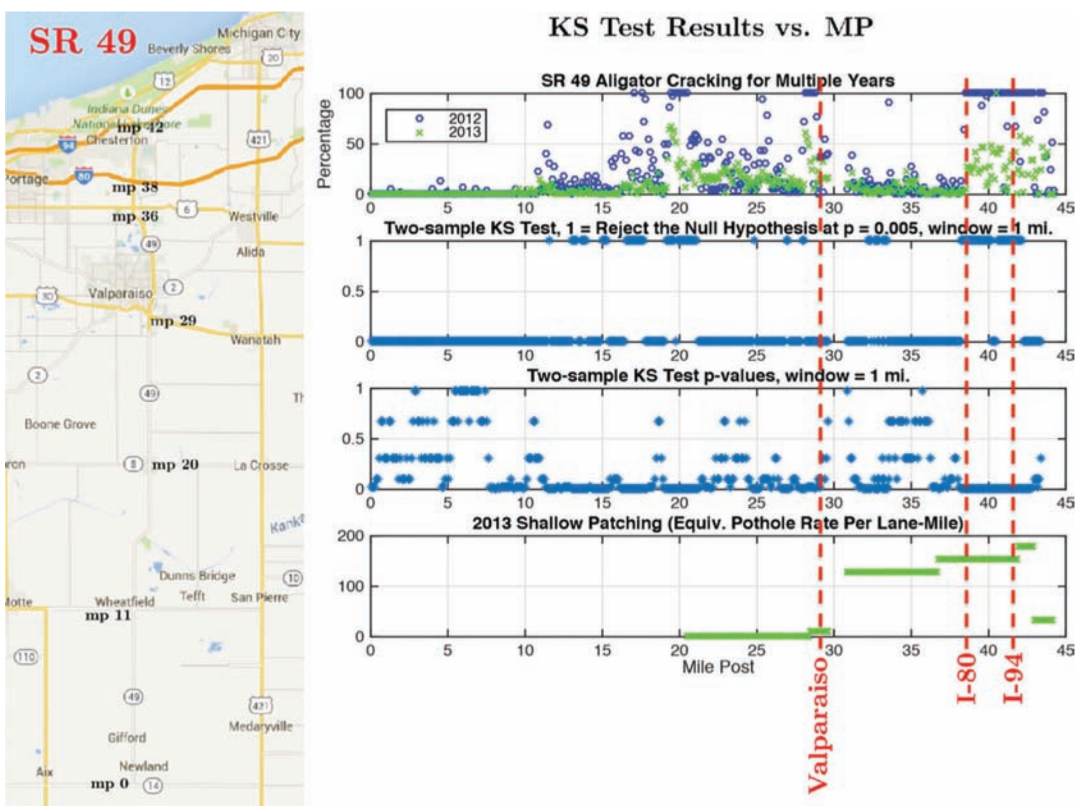


Figure 3.17 SR 49 KS test results for AC.

between traffic load and the change of road condition, which agrees to some degree with the intuition that road segments with more traffic are more likely to crack.

Figure 3.21 shows the amount of patching activity in different years. In some sense this reflects the number of potholes that were formed in each year. Still, the correlation of shallow patching data with AC or BC improvement is not conclusive, partly because the low precision of the patching data. The research team proposes an algorithm that takes advantage of GPS tracking information of the asphalt truck to dramatically improve the precision for shallow patching records. This algorithm has been implemented using MATLAB and tested with the GPS data collected by PPTracker, an Android application written for the project. Please see Appendix 1 for additional information.

Similar results have been obtained for two other road segments. The localized KS testing via sliding windows consistently show the effectiveness in detection of changes in AC and BC percentages.

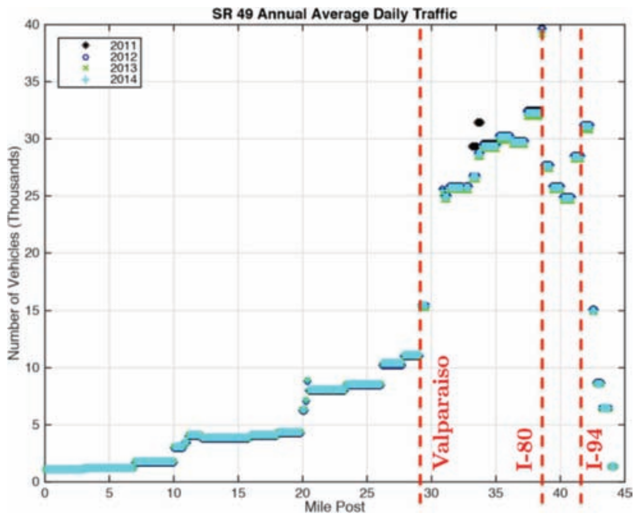


Figure 3.18 AADT over MP for SR 49 from 2011 to 2014.

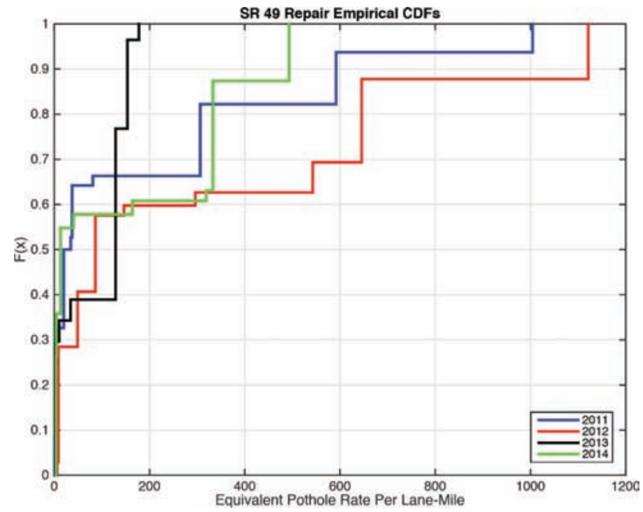


Figure 3.21 Empirical CDF for equivalent pothole rate of SR 49 from 2012 to 2014.

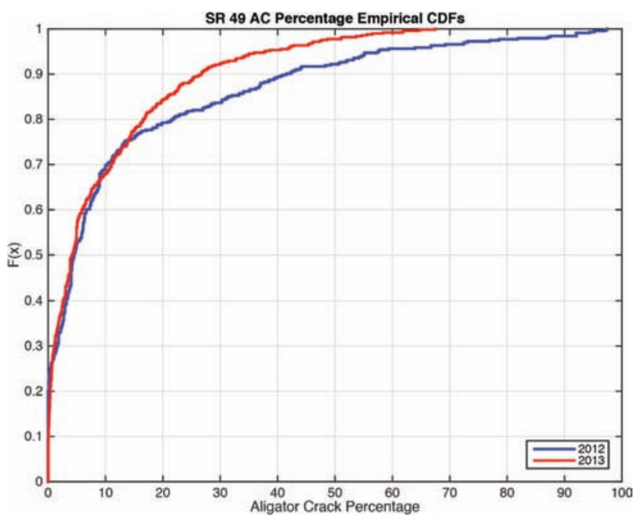


Figure 3.19 Empirical CDF for AC of SR 49 in 2012 and 2013.

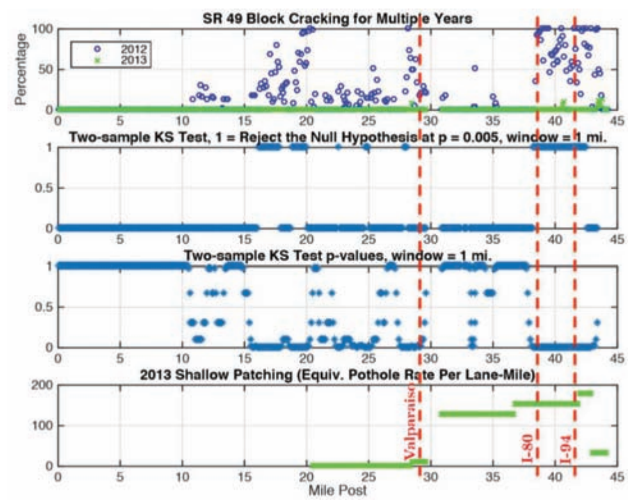


Figure 3.22 SR 49 KS test results for BC.

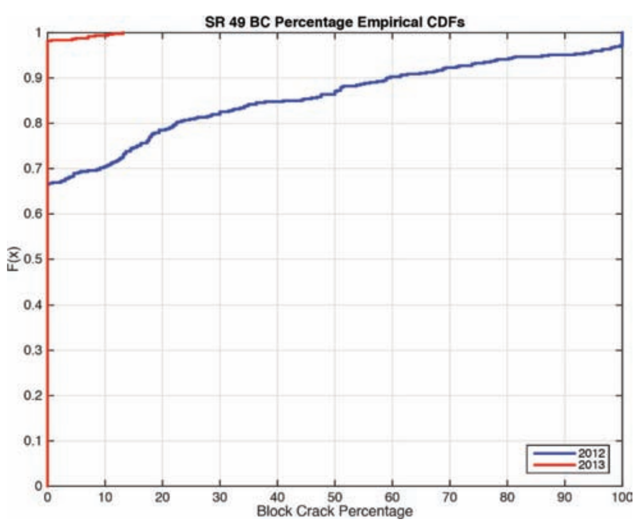


Figure 3.20 Empirical CDF for BC of SR 49 in 2012 and 2013.

3.3.3 Implications

According to these results, statistical analysis methods show promise for identifying roadway segments with emergent distress. We have also discovered a mismatch of record precision for condition data and repair data. Current repair data records are about 100 times lower resolution than the condition data. The results may improve if higher resolution repair records are available. Very high resolution repair records could be recorded with a mobile approach like PPTracker and would require little additional effort.

3.4 Large Scale Analysis of the Patching Data

3.4.1 Raw Data Sets and Assumptions

Both data sets for road condition and pavement repair are utilized. The raw condition data was made available

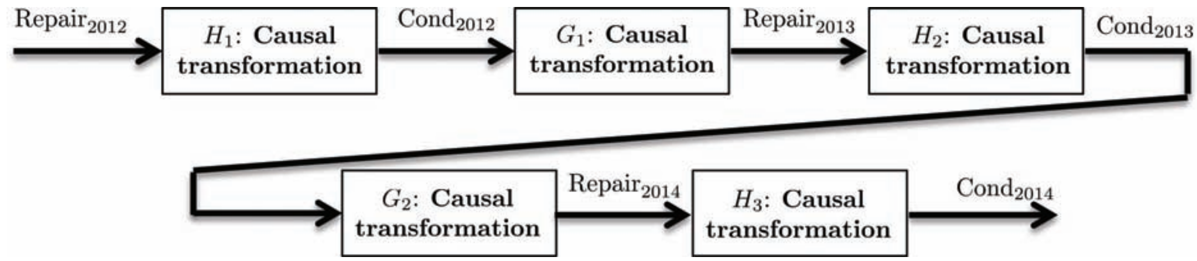


Figure 3.23 Causal model for changes in road condition.

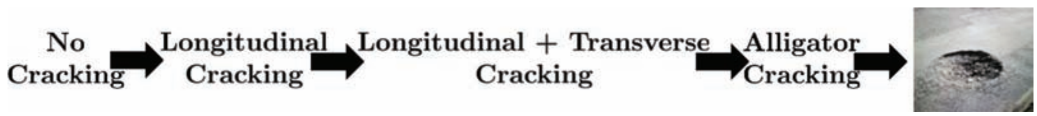


Figure 3.24 Different stages in pothole formation.

for 2012 and 2013 as an Excel spreadsheet and the 2014 data as an Access database. Therefore, a local database that contained the data from all three datasets was created. This allowed for easy queries from the MATLAB software that was used during the algorithm development.

In a particular calendar year, repair usually occurred in months 1 to 6 while condition measurements were usually recorded in months 7 to 12. Therefore, it was assumed the records for repair and road condition are well separated in time and the road only decays between them. Figure 3.23 illustrates this causal model. As shown, repair, road deterioration, and road condition data collection are considered as individual events and they occur one by one in sequence. Furthermore, as a road wears it is assumed the cracking manifests over time in stages, as shown in Figure 3.24.

When working with the data, the wheel path is first examined, followed by the edge, following INDOT instructions. Alligator and longitudinal cracking were the focus, as per information in the Pavement Surface Condition Rating Manual from Washington State DOT (Northwest Pavement Management Systems Users Group & Kay, 1992). This simplifies the algorithm development. This cracking data are indexed by lane section and severity and tied to corresponding parameters such as milepost, direction, and date.

For the maintenance data, three types of repairs are reported: temporary shallow patching, permanent shallow patching and deep patching. Note that another very important maintenance activity, spot paving, is not reflected in the records currently available.

3.4.2 Data Quality Issues

Some issues or anomalies were observed in the available datasets. The most severe one is the mismatch of location references and condition data in different years. More specifically, the data for 2012 had MP linked to GPS information, i.e., latitude and longitude, for each data point, while the data for 2013 had only GPS information and for 2014 only MP was available.

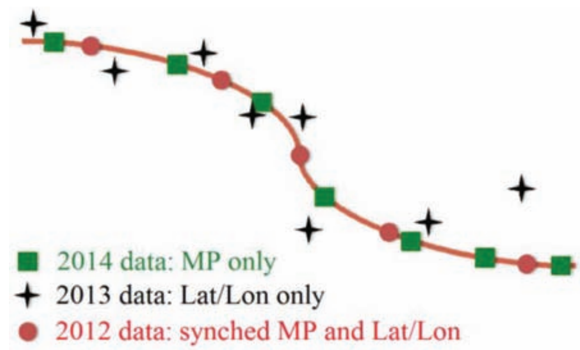


Figure 3.25 Location synchronization problem.

Furthermore, the position of each data point varied slightly from year to year. The problem is illustrated in Figure 3.25. This sets some barriers for comparing the road condition between different years.

To solve this problem, a MATLAB procedure was written that recreated the linear reference for the 2012 data and spatially synchronized the 2013 and 2014 data points to the 2012 linear reference system.

For convenience, 2012 MP and latitudes/longitudes were used as the standard. The synchronization procedure first finds a 2012 location for each 2013/2014 data point, links to that data point, then computes the converted 2013/2014 data location in terms of 2012 locations. A nearest neighborhood approach was employed for the matching. Figure 3.26 illustrates how the algorithm implemented for the 2013 data set works. First, a circle of radius 0.06 miles is centered at each 2012 location. Then, all of the 2013 data points within the corresponding circle were identified. If no 2013 points fall inside the circle, nothing will be assigned. Otherwise, if one or more 2013 latitude/longitude locations fall inside the circle, the average condition data value was assigned to the 2012 location. The procedure is similar but easier for the 2014 data set because in that case the linear reference MP, instead of two-dimensional GPS points, are used for location matching.

The raw repairing records are normally used to track maintenance activities within a day. Route name, start MP, end MP and total amount of material used are reported but exactly where the material was used is not recorded. Therefore, it is hard to compare the data sets for maintenance and road condition and reveal any meaningful relationships between them.

The solution to this problem was to compute the patching intensity in tons per lane-mile, as a way to spread the maintenance work done over the corresponding road segments. Formula (4) shows the procedure to distribute the patching materials. I_{patch} is the patching intensity in tons per lane-mile. M is the total amount of material used. MP_{end} and MP_{start} are the corresponding MP for the end and start locations, respectively. The number of lanes covered in each direction of travel for a segment, N_{lane}, was obtained from the condition data set. In other words, it is assumed that the material was patched uniformly for the whole road segment and for all lanes. This was carried out the computation for each route and for each 2012 standard MP (approximately 0.1-mile grid).

$$I_{patch} = \frac{M}{(MP_{end} - MP_{start})N_{lane}} \tag{3.4}$$

Some additional issues still remain in the dataset.

1. Some roads in the condition data changed between years and it is not clear how to correctly compare points from the two linear reference systems.
2. Alligator cracking percentage exceeded 100 percent in a few places.
3. Patching data has material use outliers. For example, I-65 repair had some entries that were three orders of magnitude too large.
4. The patching metadata sometimes indicates the reported patching occurred on ramps rather than on main roadway, yet the maintenance record indicates the roadway was repaired.

3.4.3 Multiscale Analysis

In order to predict problematic road segments with a high level of confidence and yet simultaneously provide detailed reporting of road state for a particular set of segments a procedure that analyzes the data at various scales is needed. This procedure can be illustrated for 3 different geographical levels: the entire state, individual routes in the state, and individual segments on the routes.

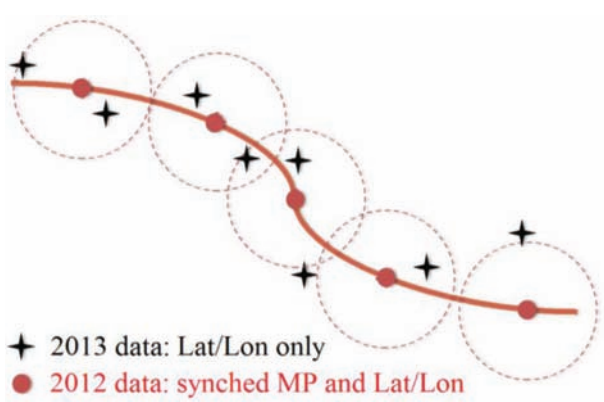


Figure 3.26 Location matching.

The entire state of Indiana has a total of 194 routes, of which 12 are interstates, 163 are state roads and the remaining 19 routes are U.S. highways. As seen before, empirical CDF plots are useful for illustrating the overall trends for multiple years. Here, the sum of LC and AC data from the wheel path with low, medium and high severities is used as the parameter of interest. One of many possible stochastic metrics, such as 90-th percentile, KS test p value, etc., can be used to classify the parameter as "improves" or "degrades" from year-to-year. As seen in Figures 3.27 and 3.28, LC worsened from 2012 to 2014, while AC improved significantly.

Of the 194 routes, 142 appear in the repair database. It is assumed that those not in the repair database had zero patching. Figure 3.29 is the empirical CDF plot (Figure 3.30 is bar graph of the same data) for patch intensity in tons per lane-mile. The figure shows the maintenance activity is similar year-after-year, except for 2015. However, the 2015 data is only partially complete, given that it was received mid-2015.

When investigating individual routes an empirical CDF and a stochastic comparison metric for each roadway and year is first computed. This is computationally efficient enough that all 194 routes can be analyzed on a laptop in a few minutes. The results for SR 49 are shown in Figures 3.31 to 34 and Table 3.3. Figure 3.34 is a summary for the comparison test results over all 194 routes.

From the classification results for each roadway, empirical transition probabilities can be computed. The results are shown in Tables 3.4 and 3.5. These probabilities could be used in prediction models as the true probabilities of corresponding road changes over a year.

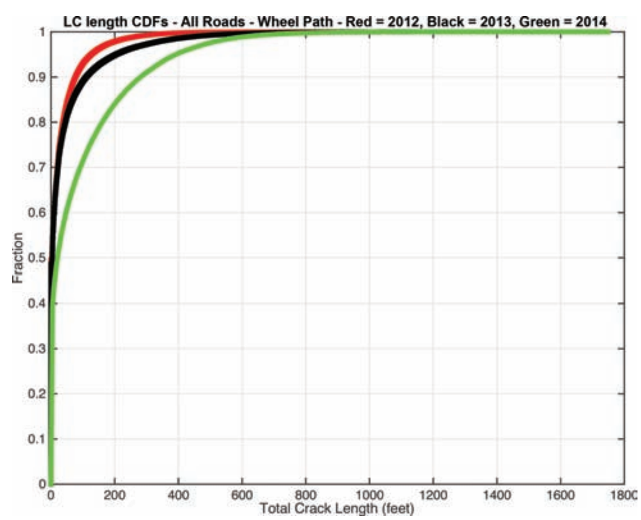


Figure 3.27 Empirical CDF plot for averaged LC of the entire state.

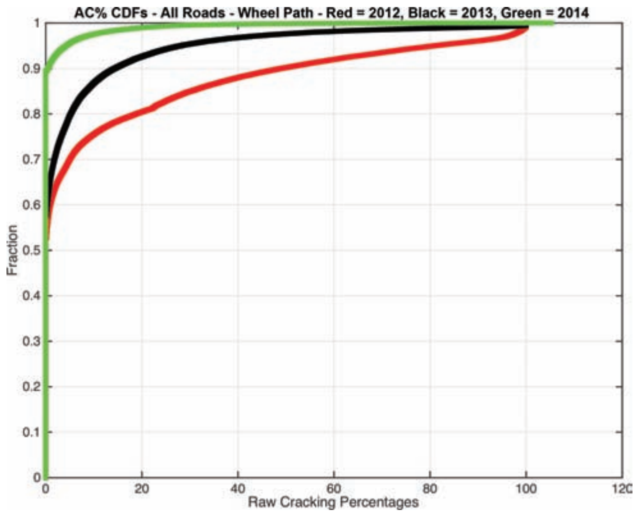


Figure 3.28 Empirical CDF plot for averaged AC of the entire state.

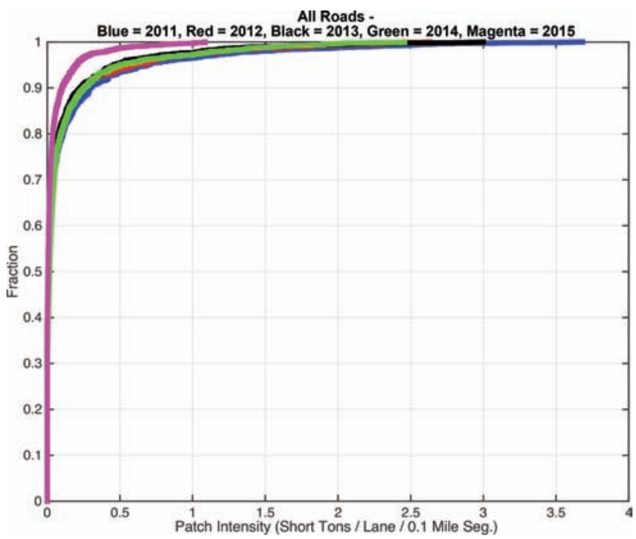


Figure 3.29 Empirical CDF plot for patch intensity of the entire state.

However, as can be seen, the probabilities for different years are consistent, which indicates the model has not uncovered the fundamental probability of degradation and repair. Additional conditioning on the various influences of cracking is necessary.

Finally, as the last demonstration of the multi-scale processing, US 41 is broken into segments of various lengths. The original roadway is 80 miles long and is divided into four parts. Figures 3.35 to 3.42 show the results.

There is a tradeoff that comes with multi-scale processing. Less data is being averaged as the resolution increases, so the statistical significance of the classification is reduced. Figures 3.43 to 3.47 show the LC empirical CDF plots for the first road segment when the segment length changes from 80 miles to 10 miles. As can be seen, shorter segment lengths result in noisier plots, although the spatial resolution is improved.

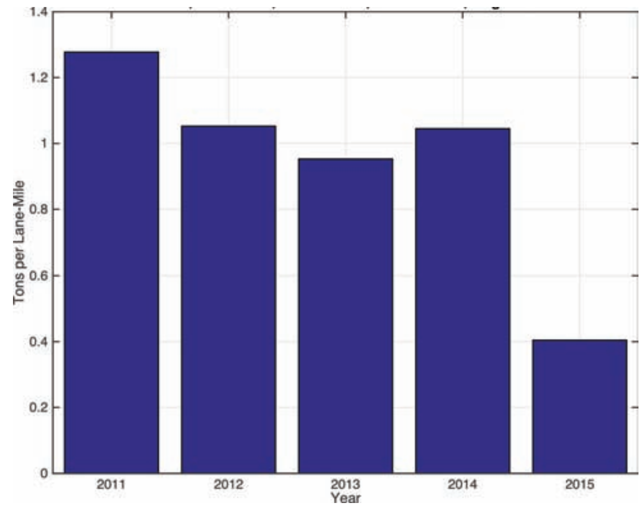


Figure 3.30 Empirical CDF plot for patch intensity of the entire state.

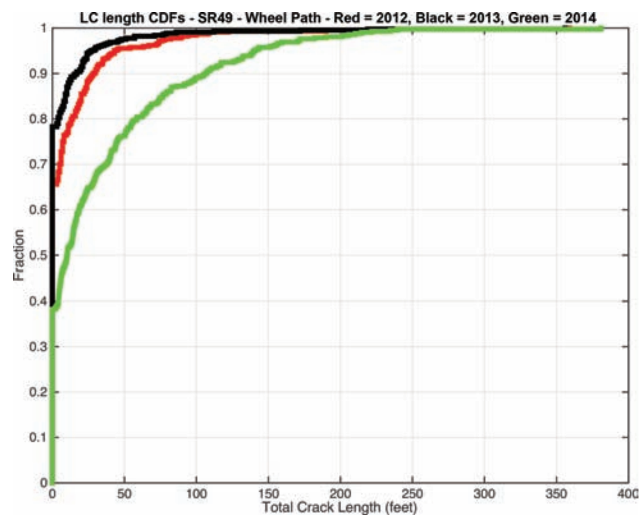


Figure 3.31 Empirical CDF plot for LC of SR 49.

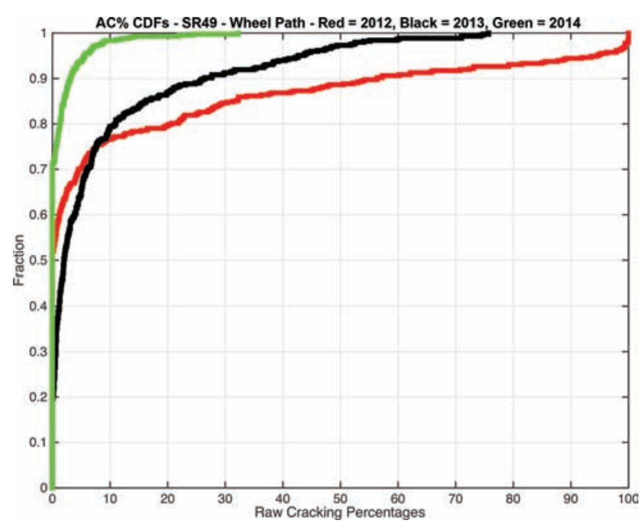


Figure 3.32 Empirical CDF plot for AC of SR 49.

Figure 3.48 demonstrates the above analysis carried out on a select set of Indiana roads. Some patterns of road improvement and degradation that correspond with repair activity (or the lack thereof) can be observed. It is clear that some other outside factors in the model need to be conditioned. However, there is evidence that data may in fact be sufficient to make some observations

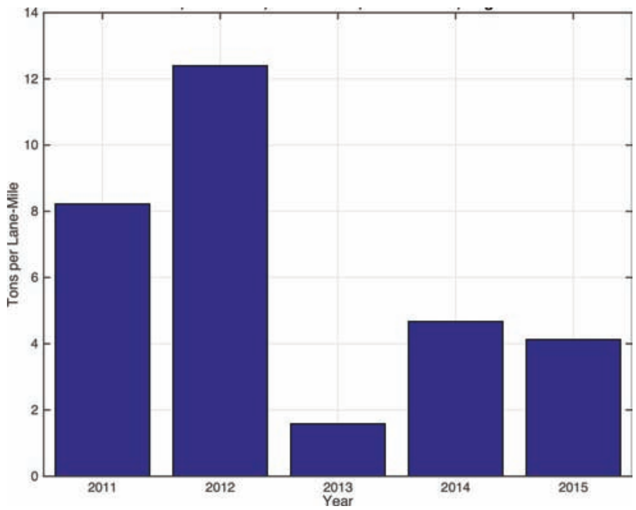


Figure 3.33 Empirical CDF plot for patch intensity of SR 49.

TABLE 3.3
KS test results for SR 49.

	Longitudinal Cracking (LC)	Alligator Cracking (AC)
2012 → 2013	Improves	Same
2013 → 2014	Degrades	Improves

TABLE 3.4
Empirical probabilities of changes in longitudinal cracking (LC).

	Improves	Same	Degrades
2012 → 2013	0.45	0.15	0.40
2013 → 2014	0.20	0.09	0.72

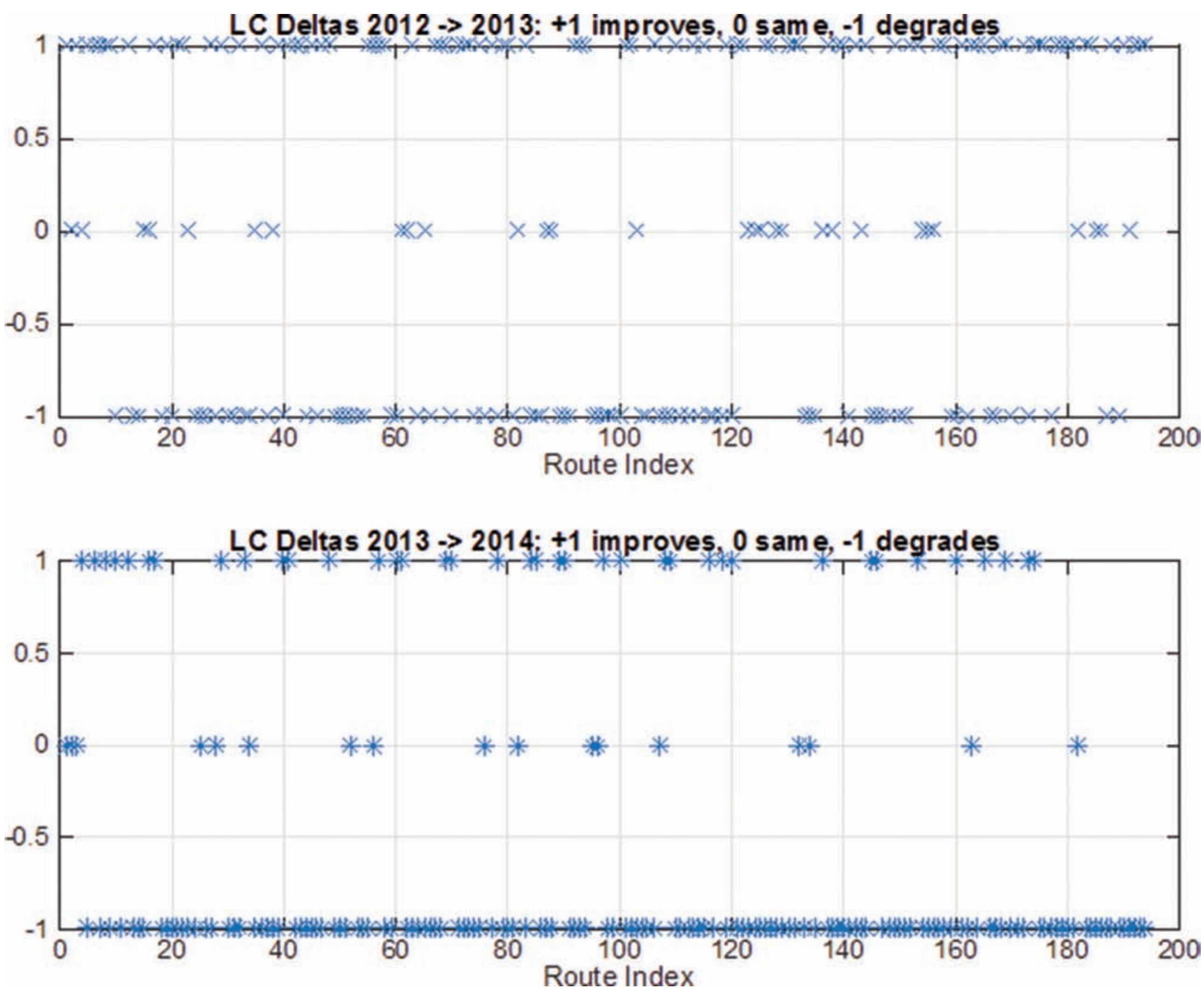


Figure 3.34 Summary for metric comparison over 194 routes.

TABLE 3.5
Empirical probabilities of changes in alligator cracking (AC).

	Improves	Same	Degrades
2012 → 2013	0.57	0.11	0.32
2013 → 2014	0.90	0.05	0.05

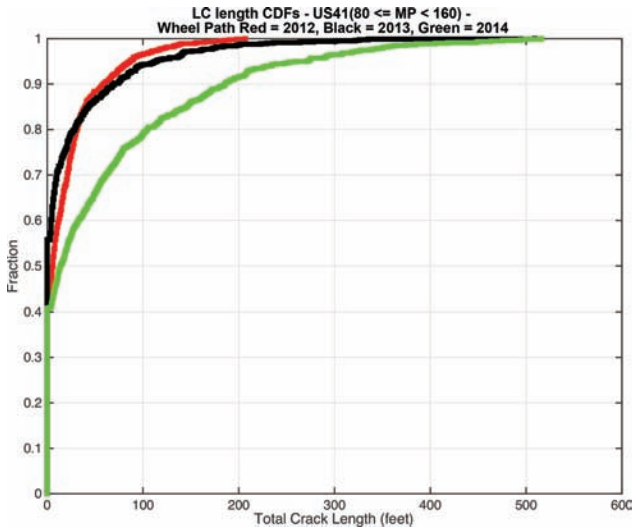


Figure 3.35 Empirical CDF for LC of the 1st 80-mile road segment.

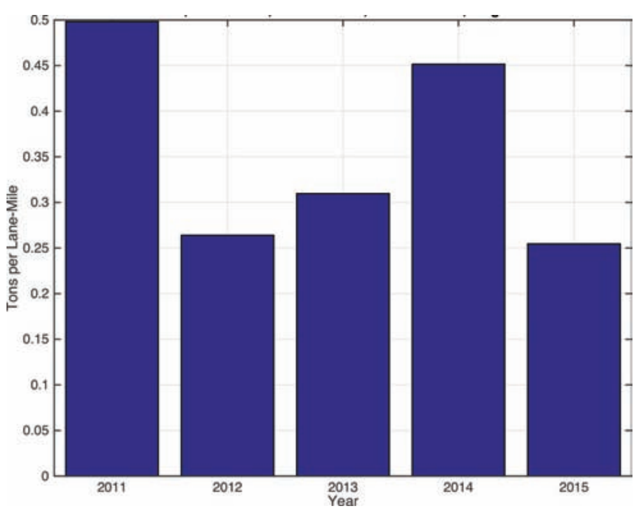


Figure 3.36 Bar chart for patching intensity of the 1st 80-mile road segment.

toward a segment's tendency to have potholes in the future. Figure 3.49 is final probability transition matrix computed across all the measured Indiana roadways.

The analysis procedure is able to generate lists of road segments ranked from worst to best according to some metrics, like total LC/AC, fraction of high severity cracks, or complex combinations of LC, AC, etc., and could be used to prioritize maintenance and/or rehabilitation.

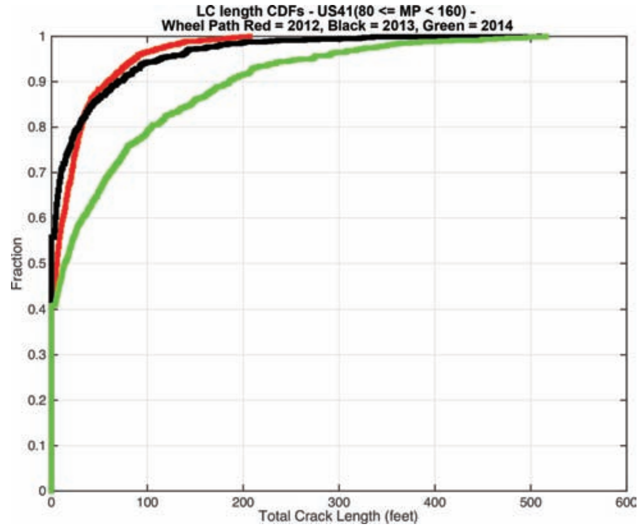


Figure 3.37 Empirical CDF for LC of the 2nd 80-mile road segment.

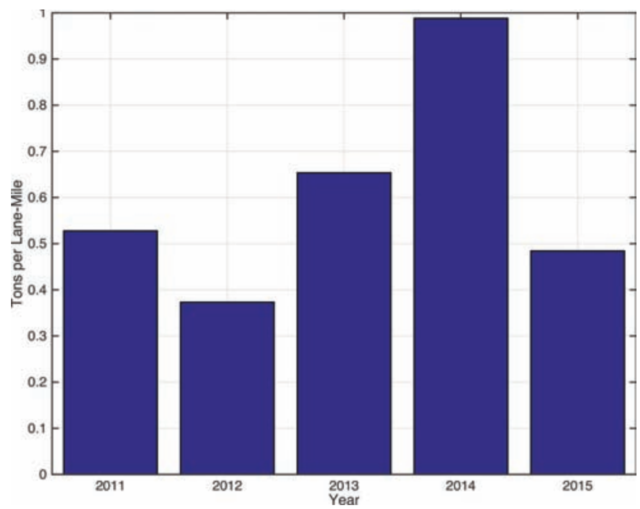


Figure 3.38 Bar chart for patching intensity of the 2nd 80-mile road segment.

3.4.4 Implications

The analysis reveals some interesting facts about the data. First, in general, LC continues to worsen from 2012 to 2014, while AC improved over the same period. Since AC represents a condition state just prior to pothole emergence, this trend may reflect recent effective maintenance statewide. Second, the analysis shows that the resolution of the Pathways condition data may not be sufficient to predict a pothole of only a few feet in size, but may be sufficient to predict road segments of about 1 mile that will, or will not experience distress. However, the repair database is not sufficient to prove or disprove the results accordingly. After all, the repair database resolution is on the order of 10 miles.

It is also worth noting that repair data may not always be the best indicator for where the road condition is

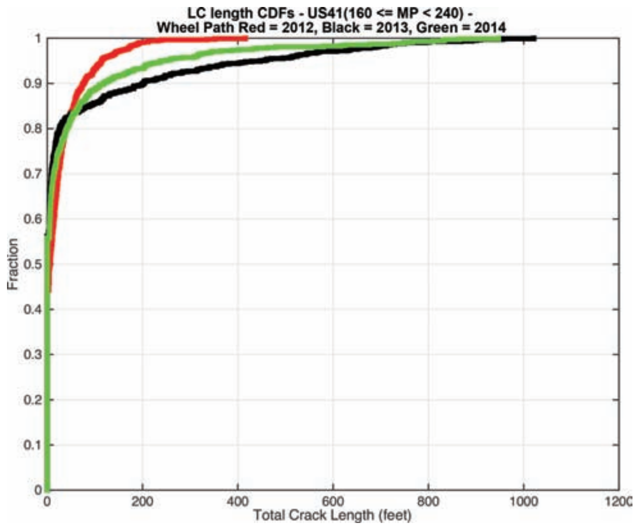


Figure 3.39 Empirical CDF for LC of the 3rd 80-mile road segment.

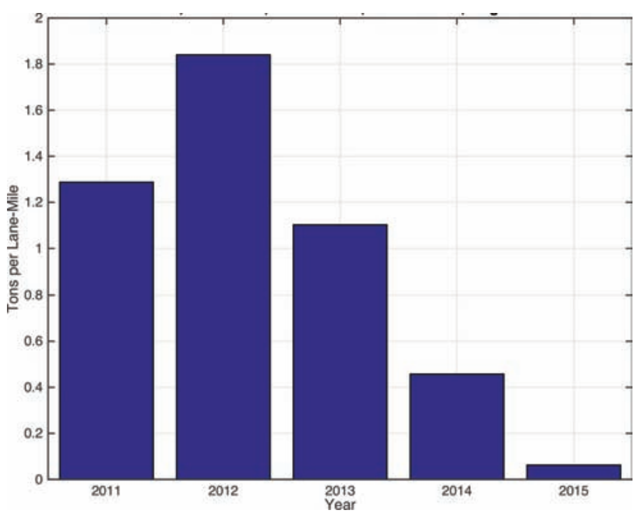


Figure 3.40 Bar chart for patching intensity of the 3rd 80-mile road segment.

severe and maintenance is required. Often repairs are subject to budget and logistical constraints that are unrelated to condition. For example, the lack of repair is not proof of the lack of need for repair.

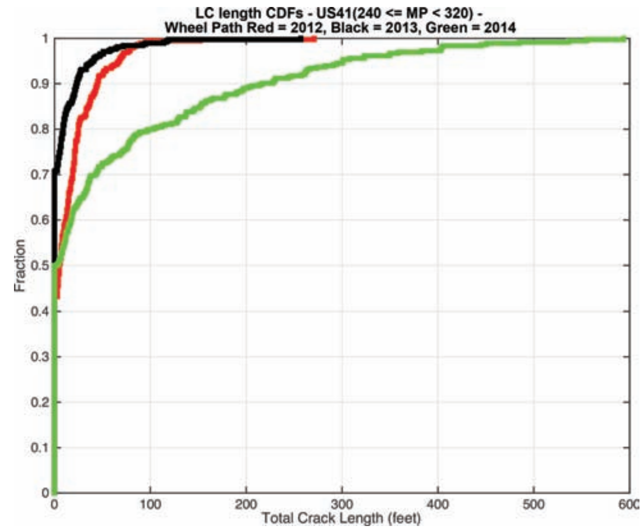


Figure 3.41 Empirical CDF for LC of the 4th 80-mile road segment.

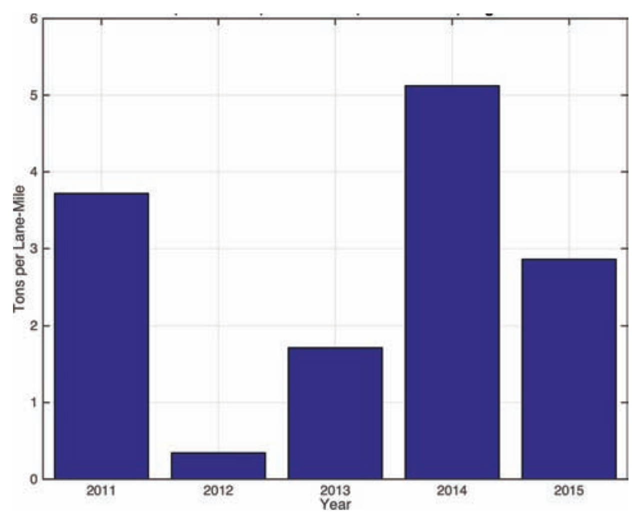


Figure 3.42 Bar chart for patching intensity of the 4th 80-mile road segment.

The preparation stage for the analysis, more specifically, the geographical synchronization process, suggests that the Pathways data could be improved by basing all location data on GPS latitude and longitude rather than MP.

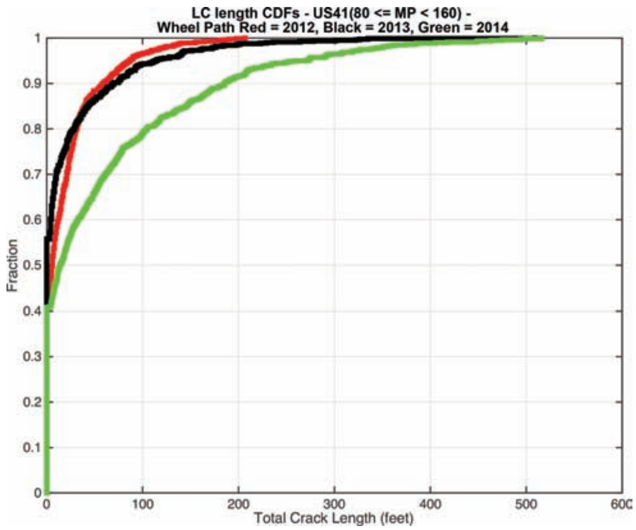


Figure 3.43 Empirical CDF for LC on US 41 with $80 \leq MP < 160$.

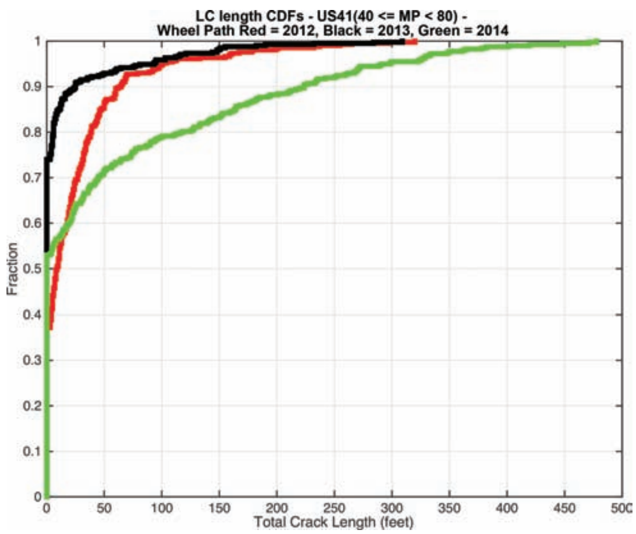


Figure 3.44 Empirical CDF for LC on US 41 with $40 \leq MP < 80$.

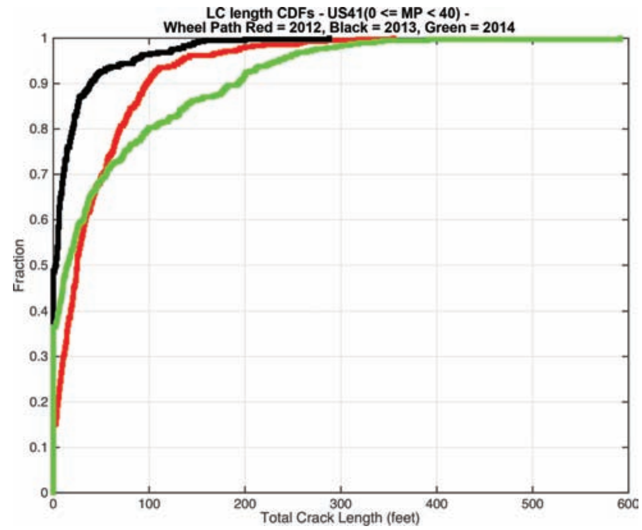


Figure 3.45 Empirical CDF for LC on US 41 with $0 \leq MP < 40$.

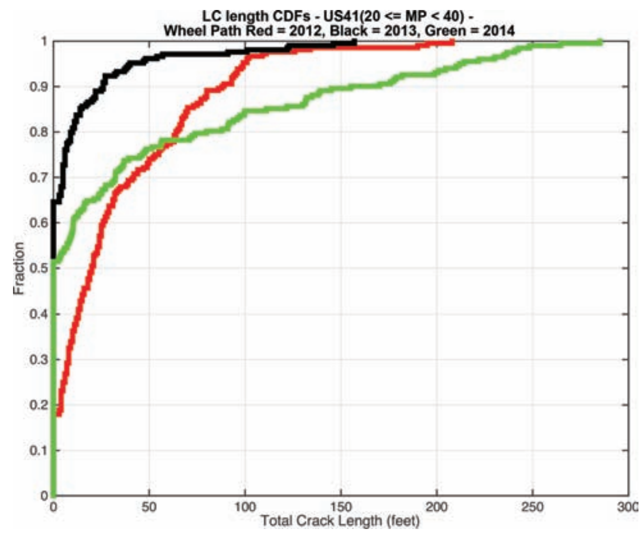


Figure 3.46 Empirical CDF for LC on US 41 with $20 \leq MP < 40$.

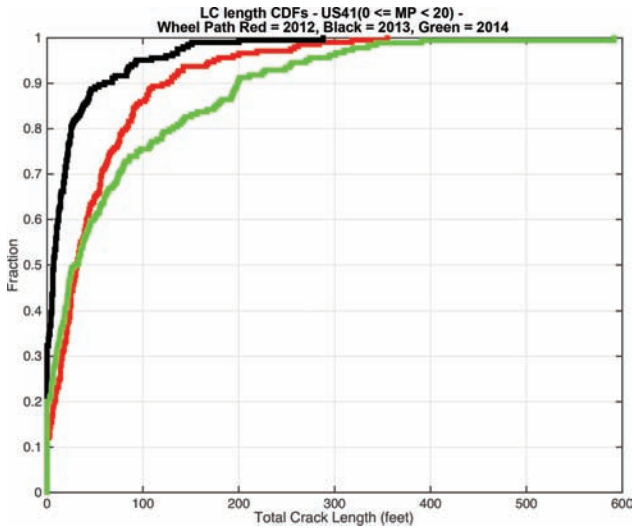


Figure 3.47 Empirical CDF for LC on US 41 with $0 \leq MP < 20$.

Empirical Probabilites of Changes in Longitudinal Cracking			
	improves	same	degrades
2012 → 2013	0.82	0.04	0.14
2013 → 2014	0.29	0.14	0.57

Empirical Probabilites of Changes in Alligator Cracking			
	improves	same	degrades
2012 → 2013	0.70	0.11	0.19
2013 → 2014	0.82	0.09	0.09

Figure 3.49 Road transition probabilities for moderate scale analysis.

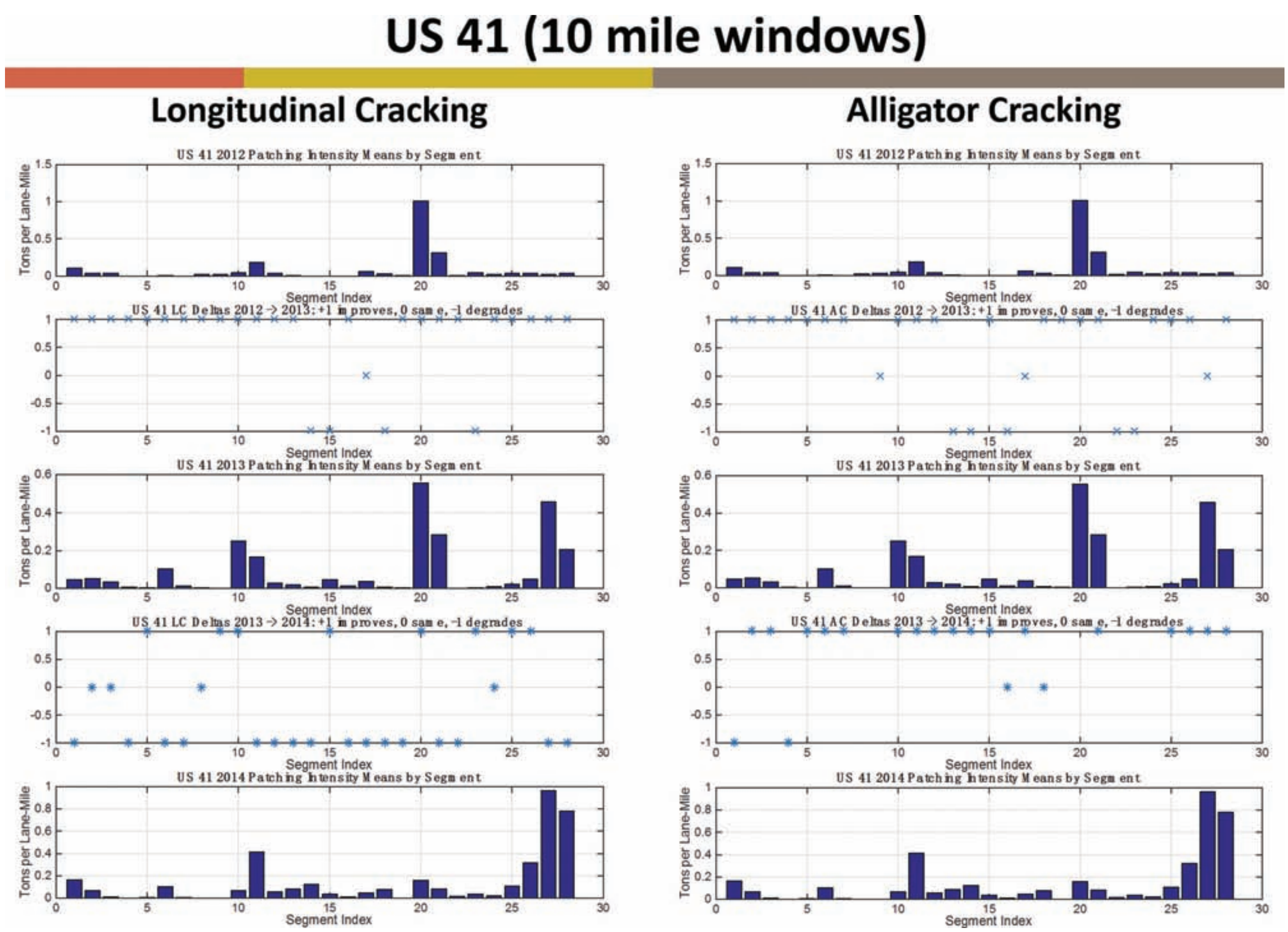


Figure 3.48 Analysis of select roadways throughout Indiana.

4. CONCLUSION

Two different schemes have been presented that ultimately lead to the prediction of pothole formation. The first reveals the relationships between road deterioration and main factors such as weather and traffic load. Three models were developed to estimate the number of potholes requiring patching due to the weather conditions and traffic loads for three different types of roadways. The models indicate that traffic load plays a more important role in formation of the potholes in the urban and rural interstate highways compared to the urban US highways and urban and rural SR routes. The models also reveal that the temperature is more important than the traffic loads in formation of the potholes in rural routes. Applying these models can help agencies assign maintenance priority to highways predicted to develop comparatively more potholes, depending on geographical location, traffic loads, and weather conditions.

A comprehensive collection of higher resolution pavement condition data prior to pothole formation, climatic conditions, and more precise records of existing pothole locations and severities could help more accurately predict the severity and number of potholes with greater likelihood and more precision.

The second analysis scheme was to predict the state of a road segment in the near future according to the past trends that have shown in its condition and maintenance data. More specifically, statistical tools such as the empirical CDF and KS tests were employed. According to the results, these statistical analyses methods show promise for identifying roadway segments with emergent distress. The analyses suggest that the resolution of condition data from Pathways may not be sufficient to predict a pothole of only a few feet in size, but it may very well be sufficient to predict that road segments of length about 1 mile will, or will not experience distress. However, the repair database is not sufficient to prove or disprove the results accordingly because of its low resolution. It would help if higher resolution repair records were available. Very high resolution repair records could recorded with a mobile approach like PPTracker and would require little additional effort.

Ultimately there is more work that could be done on the topic. For example, schemes to automatically flag and fix quality control issues. Similarly, the analysis algorithms could be made fully automatic and user friendly to sort and search through. Other INDOT databases could be more directly integrated into the algorithms, such as: pavement age, pavement type, traffic volumes, complete weather data, etc. Finally, when increased resolution data becomes available, such as INDOT's ultra-high resolution condition mapping van, a true pothole formation prediction algorithm could be developed.

REFERENCES

Indiana Department of Transportation (INDOT). (n.d.). Traffic data. Retrieved from www.in.gov/indot/2469.htm (access date May 4, 2015).

Jimoh, Y. A. (2012). Model and application for a pothole in a flexible pavement maintenance due to axle loads. *Epistemics in Science, Engineering and Technology*, *2*(1), 1–7.

Kuennen, T. (2004). The pothole patching playbook. *Better Roads*, *74*(2), 30–41.

Kutner, M. H., Nachtsheim, C., Neter, J. & Li, W. (2005). *Applied linear statistical models* (5th ed.). Boston, MA: McGraw Hill Irwin.

McDaniel, R. S., Olek, J., Behnood, A., Magee, B., & Pollock, R. (2014). *Pavement patching practices* (NCHRP Synthesis 463). Washington, DC: Transportation Research Board of the National Academies. Retrieved from http://www.trb.org/Main/Blurbs/171155.aspx

Miller, J. S., & Bellinger, W. Y. (2003). *Distress identification manual for the long term pavement performance program.* Washington, DC: Federal Highway Administration.

Morosiuk, G., Riley, M. J., & Odoki, J. B. (2004). *Modeling road deterioration and works effects.* Birmingham, England: University of Birmingham, UK Highway Development and Management.

Mubaraki, M., & Thom, N. (2012). Sigmoid distress prediction models at project level for main urban flexible pavements based on historical data. In *Transportation Research Board 92nd Annual Meeting compendium of papers* (Paper #13-1803). Washington, DC: Transportation Research Board.

Northwest Pavement Management Systems Users Group & Kay, R. K. (1992). *Pavement surface condition rating manual.* Seattle, WA: Washington State Transportation Center, University of Washington. Retrieved from https://www.wsdot.wa.gov/NR/rdonlyres/1AB0E29D-72D7-466A-9547-C9F631B4CE6C/0/PavementSurfaceCondition RatingManual.pdf

Paterson, W. D. (1987). *Road deterioration and maintenance effects: Models for planning and management.* The highway design and maintenance standards series. Baltimore, MD: The Johns Hopkins University Press.

Washington, S. P., Karlaftis, M. G., & Mannering, F. L. (2011). *Statistical and econometric methods for transportation data analysis* (2nd ed.). Boca Raton, FL: Chapman and Hall/CRC.

APPENDIX A: PPTRACKER

A1. Introduction

During the pothole prediction algorithm development, a set of inputs for the prediction model with expected output information, is essential not only for the algorithm developers to better understand the situation they are dealing with, but also for the training process of the supervised learning model to set parameters which assure a satisfactory performance (Hagan, Demuth, & Beale, 1996). In this case, considering all the currently available records, it is intuitive to use road conditioning data from previous years as the input to the prediction model and confirm that the predicted trouble spots indeed required extra attention and if they have been patched according to the maintenance records. Therefore, the maintenance records are crucial to the pothole prediction algorithm development. They will serve as both the source of training data generation, to correctly determine the expected outputs for the input data in the training set, and the key to evaluate how well a prediction model works, by comparing the outputs of the model for the testing set with the corresponding maintenance data.

However, after closely examining the data that can be extracted from the maintenance records, it was found that the best available precision in terms of location, which is in the order of miles, may still be much too long for fulfilling the roles mentioned above. There are two major obstacles. First, potholes, or the deteriorated road areas that can be fixed by pavement patching procedure, are in the length of several meters to tens of meters. While the precision for the condition data may be good enough to catch their characteristics, the precision for the maintenance data does not match them. Second, the maintenance records only show rough mile post ranges in which the pavement patching occurs, without specifying which road segments are actually patched. This makes it very hard to get correct classification results for the condition data input, because more often than not, only a small subset of the whole range in the record has been fixed. We will show this in a case study discussed in Section IV.

In order to overcome these obstacles, an algorithm for automatically exacting records with high precision for the patched road segments from GPS tracks was proposed (Zhang, Balmos, Krogmeier, & Buckmaster, 2015). This was implemented using Matlab and tested with GPS data collected during a shallow patching trip on Indiana State Road 49. The result shows that the proposed system is able to increase the record precision on patched locations from kilometers to the order of several meters, without creating a burden to the patching crew. This not only makes it possible to have better management and performance evaluation, but also provide a solid foundation for future research on road deterioration observation and prediction.

A2. Patching Zone Identification

Through activity recognition for vehicles in the patching fleet, it is possible to extract patched areas from GPS tracks. As an illustration, an algorithm using MATLAB to recognize GPS sample points that are in the patched area for shallow patching activities was implemented.

A3. Background on Shallow Patching

Pavement patching plays an important role in road maintenance (McDaniel, Olek, Behnood, Magee, & Pollock, 2014). Here we focus on one specific patching procedure, shallow patching, or so-called "cold patching." It is used to temporarily fix cracks and potholes, when the conditions like weather, temperature and traffic amount, are not suitable for better treatment procedures.

A patching fleet for shallow patching may have multiple vehicles. Normally, melted asphalt, carried by a distributor truck, will be used to fill the crack first, and then a layer of sand, carried by one or several small sand trucks, will be put on top to provide temporary protection against the traffic. The patching work, for both asphalt and sand, is normally done manually using shovels. Besides those vehicles, trucks mounted with crash impact attenuators may be used to guide the traffic and keep the working area safe. Depending on how severely the road deteriorates, the time required for fixing one location may vary, but the procedure is simple and quick enough to be carried out without completely shutting down a road.

We describe here a typical case as an example. The asphalt distributor truck, followed by the sand trucks, goes beyond but stops near the next spot to patch. Staff from the sand trucks patched the area with asphalt unloaded from the asphalt distributor truck and sand from the small sand trucks. Meanwhile, the truck with attenuator and signs is guiding the traffic to avoid the lane that is been patched at the end of the fleet.

A4. Patching Zone Extraction Algorithm

To simplify the algorithm design, we have made several key assumptions. First, we assume the fleet is not patching if it is moving at a speed higher than 5 m/s. And based on the fact that the asphalt truck normally stops when unloading the patching material, we assume it stops right at the patched spot, instead of near it. Third, it is assumed that the patching procedure for one zone will take 5 to 20 minutes.

Accordingly, the algorithm is developed as a 3-step procedure: speed test first, then patching zone extraction, and at last exception treatment. Figure A.1 shows the flowchart for the algorithm. In the speed test, all GPS sample points with speed higher than 5 m/s will be classified as out of the patching zone. Among the

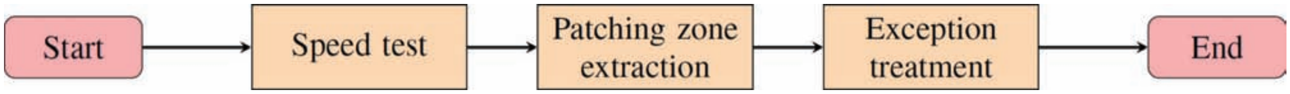


Figure A.1 The flowchart for the patching zone extraction algorithm.

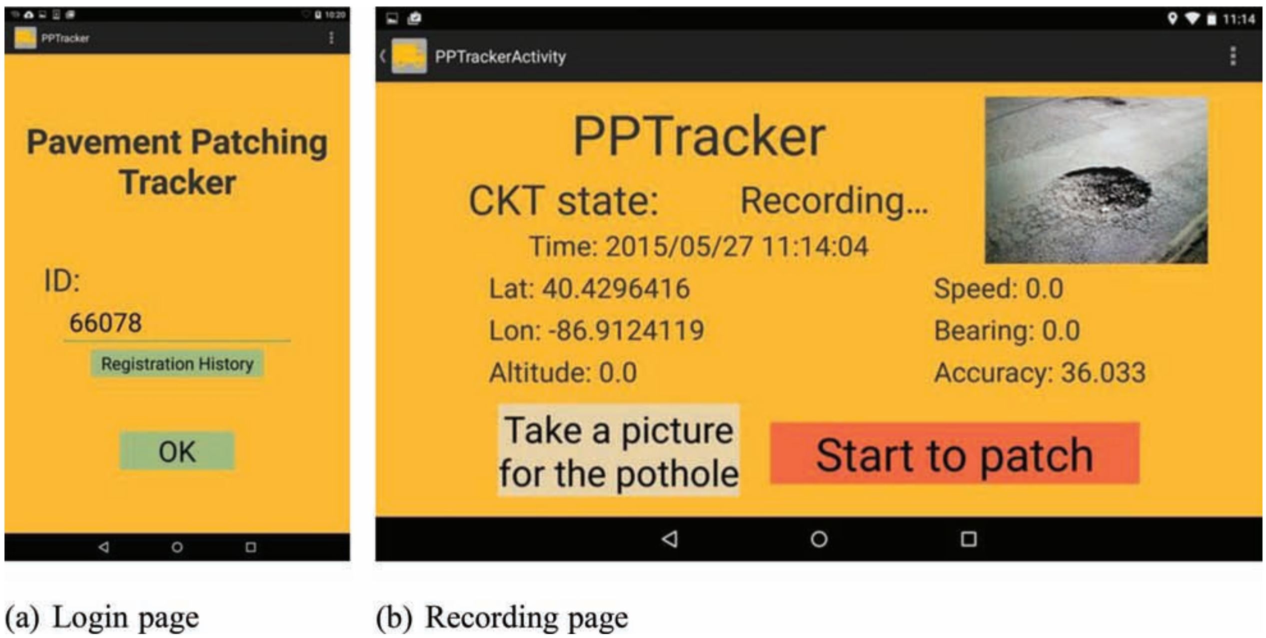


Figure A.2 Screenshots for PPTracker.

remaining data, zero speed data sequences will be found and treated as candidate patching zones. The sequences spanning shorter than 5 minutes (e.g., when the fleet stops for a red light) or longer than 20 minutes (e.g., for a lunch break) will be discarded. The remaining sequences are classified as the patching zones and their locations are where the fixed spots are. We have implemented the algorithm using MATLAB.

A5. GPS Data Collection

In order to test the algorithm, we've created an Android app called PPTracker, which stands for Pavement Patching Tracker, to collect GPS data during the pavement patching process. Much research has been conducted on the use of ubiquitous sensor-equipped mobile devices in Intelligent Transportation Systems (Alvarez, García, Naranjo, Anaya, & Jiménez, 2014; Handel et al., 2014; Simroth, & Zähle, 2011). This approach is cheap, efficient, and simple to implement.

PPTracker is essentially a GPS track logger app specialized for pavement patching. Figure A.2 illustrates two screenshots of PPTracker, for user login and GPS data collection respectively. It is able to automatically record GPS information including time, latitude, longitude, speed, etc., and organize log files in the device storage according to the vehicle information collected in the login page. With a little human interaction, PPTracker is also able to capture images and even label the data according to whether it is collected within patching zones. What's more, with the help of a free Android app called Autosync for Google Drive (https://play.google.com/store/apps/details?id=com.ttxapps.drivesync), the files generated can also be uploaded to a Google Drive account automatically, to save the trouble of exporting log files for the MATLAB program.

PPTracker was developed on a Nexus 7 tablet, but it should work on any commercial Android devices equipped with GPS sensors. Many of these devices, including the tablets we used, can be easily mounted in vehicles of the patching fleet to collect location information.

It is worth noting that the algorithm itself is designed for general GPS data of the patching activities, so if there are other location information sources available for the patching fleet, the algorithm can be easily adapted for the data structure of any specific source. And we expect the algorithm to work best if the location data is for the asphalt distributor truck in the fleet.

The GPS data collection has been carried out for one pavement patching trip on Indiana State Road 49. Besides the GPS samples collected for one truck within the patching fleet, the patching activities are also manually recorded using the app for performance evaluation. Figure A-6(a) plots the full track with mile markers on a Google map.

A6. Results and Discussion

The results of the algorithm are shown in Figure A.3. The speed test is illustrated in Figure A.3 (b). We can

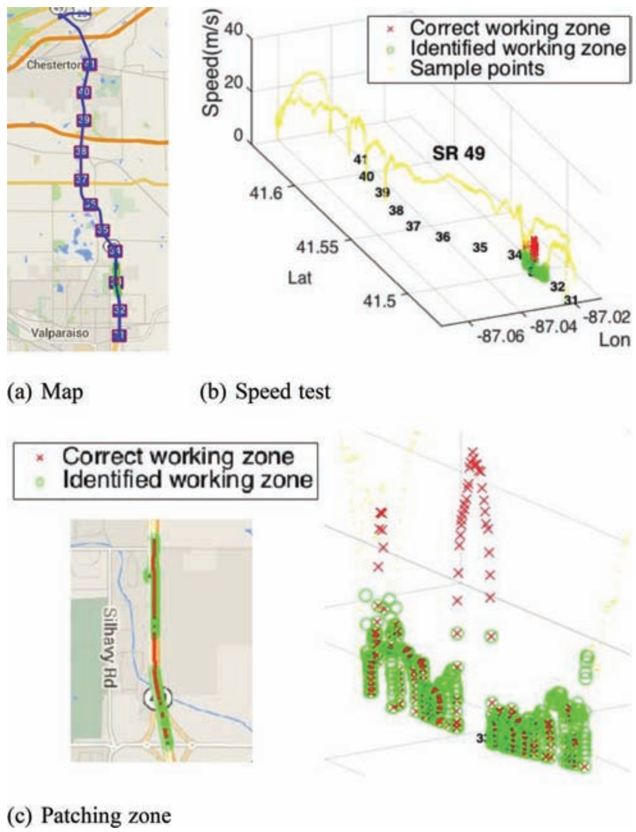


Figure A.3 Patching area identification results.

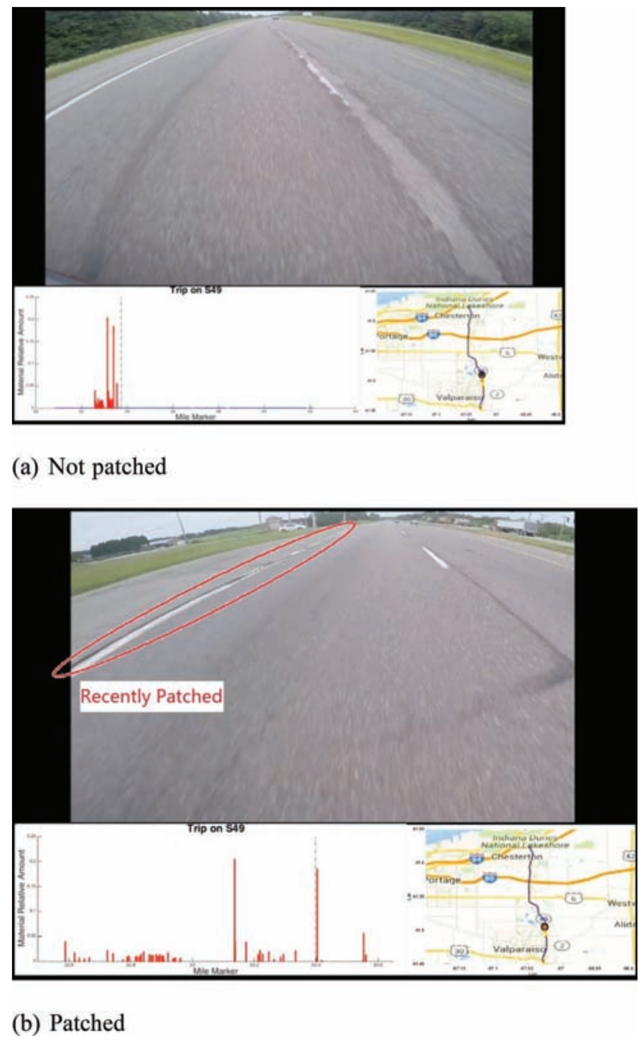


Figure A.4 Video frames for road surface condition.

have a closer look at the patching zone in Figure A.3 (c) and see that the identified sample points inside the patching zone match well with those in the correct patching areas. According to these results, the patching zone extracted is from mile marker 32.57 to 33.58 with the true working zone from 32.62 to 33.57, while the corresponding patched area in the maintenance record is represented as a road segment between mile markers 30.83 and 36.73 with no direction information. So the overall improvement for the record precision in this case is dramatic, from kilometers to the order of meters.

We also revisited the same road segment and recorded the road condition using a GoPro camera, as another way to evaluate the performance of our algorithm. Figure A.4 shows two frames of the video. The first one is a frame for the area classified by our algorithm as not patched while the second one is for the area classified as patched. The plots also illustrate where the spot is on a map and how severely the algorithm thinks the patched spot was damaged. As we can see, the road captured in the first image is in good condition and there are no recent patched spots, while there is clearly a recently patched pothole, which is darker because of the color of asphalt, along the road on the left side of the second image.

Maintenance records with higher precision for shallow patching can provide many benefits. For logistics, it will be easier to temporarily patch a pothole, accurately track where the patched areas are, and come back for better maintenance procedures later. For research, as mentioned in Section I, better maintenance records can be of great help for developing pothole prediction algorithms. What's more, there are other possible applications of the high-precision records provided by the algorithm. For example, they can be used to track potholes, possibly fixed by different patching methods, and evaluate their performance in the long term.

A7. Conclusion

An algorithm to automatically extract patched areas from GPS tracks of shallow patching activities has been proposed and tested using data collected during a shallow patching trip on SR 49. The results show the algorithm is able to dramatically improve the precision of maintenance records for shallow patching, without burdensome efforts.

APPENDIX A REFERENCES

Alvarez, A. D., García, F. S., Naranjo, J. E., Anaya, J. J., & Jiménez, F. (2014). Modeling the driving behavior of electric vehicles using smartphones and neural networks. *Intelligent Transportation Systems Magazine*, *6*(3), 44–53.

Hagan, M. T., Demuth, H. B., & Beale, M. H. (1996). *Neural network design*. Boston, MA: PWS Publishing.

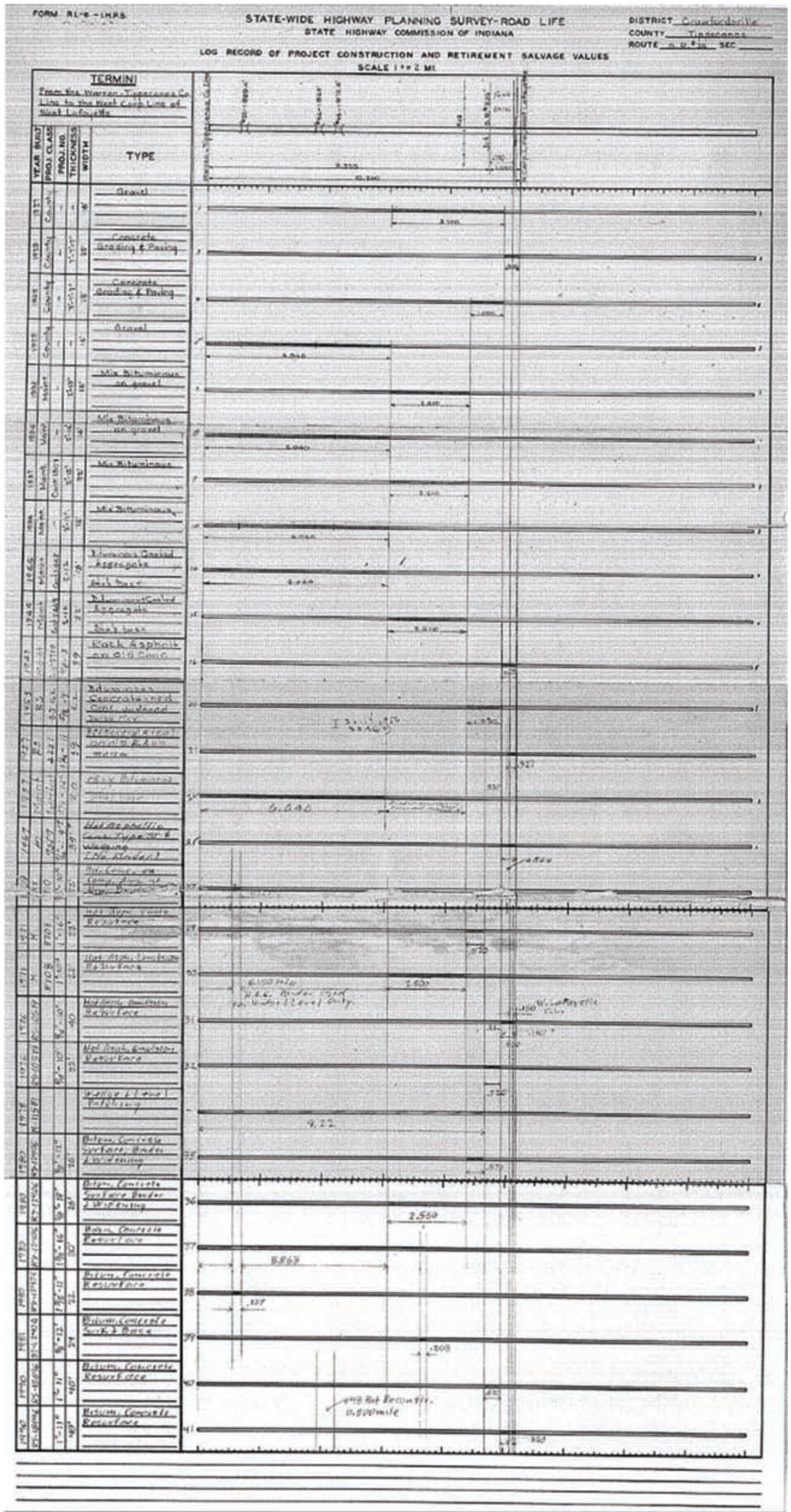
Handel, P., Skog, I., Wahlstrom, J., Bonawiede, F., Welch, R., Ohlsson, J., & Ohlsson, M. (2014). Insurance telematics: Opportunities and challenges with the smartphone solution. *IEEE Intelligent Transportation Systems Magazine*, *6*(4), 57–70.

McDaniel, R. S., Olek, J., Behnood, A., Magee, B., & Pollock, R. (2014). *Pavement patching practices* (NCHRP Synthesis 463). Washington, DC: Transportation Research Board of the National Academies. Retrieved from http://www.trb.org/Main/Blurbs/171155.aspx.

Simroth, A., & Zähle, H. (2011). Travel time prediction using floating car data applied to logistics planning. *IEEE Transactions on Intelligent Transportation Systems*, *12*(1), 243–253. Retrieved from http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=5654589.

Zhang, Y., Balmos, A., Krogmeier, J. V., & Buckmaster, D. (2015). Working zone identification for specialized micro transportation systems using GPS tracks. In *2015 IEEE 18th international conference on intelligent transportation systems (ITSC)* (pp. 1779–1784). http://dx.doi.org/10.1109/ITSC.2015.289.

APPENDIX B: PAVEMENT STRUCTURE



About the Joint Transportation Research Program (JTRP)

On March 11, 1937, the Indiana Legislature passed an act which authorized the Indiana State Highway Commission to cooperate with and assist Purdue University in developing the best methods of improving and maintaining the highways of the state and the respective counties thereof. That collaborative effort was called the Joint Highway Research Project (JHRP). In 1997 the collaborative venture was renamed as the Joint Transportation Research Program (JTRP) to reflect the state and national efforts to integrate the management and operation of various transportation modes.

The first studies of JHRP were concerned with Test Road No. 1—evaluation of the weathering characteristics of stabilized materials. After World War II, the JHRP program grew substantially and was regularly producing technical reports. Over 1,500 technical reports are now available, published as part of the JHRP and subsequently JTRP collaborative venture between Purdue University and what is now the Indiana Department of Transportation.

Free online access to all reports is provided through a unique collaboration between JTRP and Purdue Libraries. These are available at: http://docs.lib.purdue.edu/jtrp

Further information about JTRP and its current research program is available at: http://www.purdue.edu/jtrp

About This Report

An open access version of this publication is available online. This can be most easily located using the Digital Object Identifier (doi) listed below. Pre-2011 publications that include color illustrations are available online in color but are printed only in grayscale.

The recommended citation for this publication is:

Sadeghi, L., Zhang, Y., Balmos, A., Krogmeier, J. V., & Haddock, J. E. (2016). *Algorithm and software for proactive pothole repair* (Joint Transportation Research Program Publication No. FHWA/IN/JTRP-2016/14). West Lafayette, IN: Purdue University. http://dx.doi.org/10.5703/1288284316337