Leadership in Sustainable Infrastructure
Leadership en Infrastructures Durables

Vancouver, Canada
May 31 – June 3, 2017/ *Mai 31 – Juin 3, 2017*

AN ENHANCED FRAMEWORK FOR DYNAMIC SEGMENTATION OF PAVEMENT SECTIONS

Abdelaty, Ahmed[1,3] and Jeong, H. David[2]
[1] Graduate Research Assistant, Dept. of Civil, Construction & Environmental Engineering, Iowa State University, Ames, IA 50011, USA
[2] Associate Professor, Dept. of Civil, Construction & Environmental Engineering, Iowa State University, Ames, IA 50011, USA
[3] abdelaty@iastate.edu

Abstract: Over the past decades, highway agencies have used automated and semi-automated data collection methods such as laser scanning and ultrasonic waves, resulting in the collection of an enormous amount of high-density pavement condition data. The agencies are now able to quantify the level of extent and severity of different distresses even for extremely short length of pavement segments. A scientific method to aggregate those small pavement sections into a reasonable size of segments plays an important role in accurately representing the overall pavement network performance and making practical maintenance and rehabilitation decisions. This paper reviews the current methodologies for segmenting pavement condition data and summarizes their limitations. Then, the study proposes a new segmentation framework for pavement sections that finds homogenous segments by considering multiple pavement distresses for performance evaluation and treatment selection purposes using the affinity propagation clustering technique and heuristic rules. The affinity propagation clustering technique finds the similarity between pavement sections based on the distress data. However, the clustering technique does not consider the spatial nature of pavement features. As such, heuristic rules are formulated to overcome this limitation and identify homogenous pavement segments. A case study is conducted to illustrate the capabilities and applications of the proposed segmentation framework. The proposed segmentation framework will improve the a) representation of pavement condition data, b) formulation of pavement maintenance and rehabilitation strategies, and c) pavement performance evaluation.

1 Introduction

State Highway Agencies (SHAs) are now able to collect a huge amount of data because of the technological advancements in data collection methods. The majority of SHAs collect pavement performance data including international roughness index (IRI) and rutting using electronic sensing devices that utilize laser, acoustic, or infrared technologies (McGhee 2004). Those agencies also use imaging technologies and automated image processing techniques to estimate the severity and extent levels of surface distresses (McGhee 2004). For instance, Iowa Department of Transportation (DOT) collects pavement distress data every 52 feet for half of its network (5,630 miles) annually which results in more than half a million records. Similarly, Oklahoma DOT collects 800,000 pavement data records annually for approximately 8,000 miles (Calvarese 2007) while Florida DOT collects 62 data points per kilometer (Ping et al. 1999).

This high-density raw data can be used to determine pavement maintenance and rehabilitation strategies. However, first, the raw data must be processed into determining reasonable lengths of pavement segments that consists of small segments in homogeneous or at least similar conditions Each segment of relatively homogeneous conditions can then considered individually to select the most appropriate maintenance or rehabilitation strategy. This will allow pavement managers make logical and sound maintenance and rehabilitation decisions. There are many important pavement management processes and decisions that can significantly benefit from an effective method of segmenting pavement condition data. They include maintenance and rehabilitation performance evaluation, pavement deterioration model development, and maintenance and rehabilitation programming. For instance, the performance of a maintenance or rehabilitation treatment depends on the pavement condition before the application of the treatment. Thus, finding homogenous segments that share relatively uniform pavement condition will help agencies evaluate the effectiveness of different treatments accurately. Similarly, the same concept applies to the process of developing pavement deterioration models. Data management is another reason that creates a need to reduce the data from its high-density format to form longer segments of uniform pavement sections. Managing large databases is becoming more challenging because of the myriad size of the road network and the variety of the pavement data collected. Agencies may face challenges regarding the data storage of historical data (McGhee 2004). The main objective of this study is to develop a methodology that detects the longest homogenous pavement segments and accurately represents the pavement performance. The proposed methodology also overcomes the limitations of the existing segmentation methods. Data was collected from the Iowa DOT and a case study was conducted to illustrate the utilization of the proposed methodology and possible future implications.

2 Background

The cumulative difference approach (CDA), developed by the American Association of State Highway and Transportation Officials (AASHTO), is one of the earliest methods used for delineating pavement condition data (AASHTO 1993). The CDA finds statistically homogeneous segments based on the pavement condition/distress data such as deflection, skid resistance, severities of various pavement distresses, etc. Figure 1 shows the CDA approach based on an idealized assumption of a continuous and constant distress value (r_i). The segment length is represented on the x-axis using different intervals specified at (x_i). Figure 1(a) illustrates the actual pavement distress value over pavement length. Figure 1(b) represents the cumulative area, which is determined by integrating each individual pavement response rate over the interval limits. Finally, Figure 1(c) shows the difference in cumulative area values between the actual and the average area, which represents the fundamental concept used to determine uniform and homogeneous segments (AASHTO 1993).

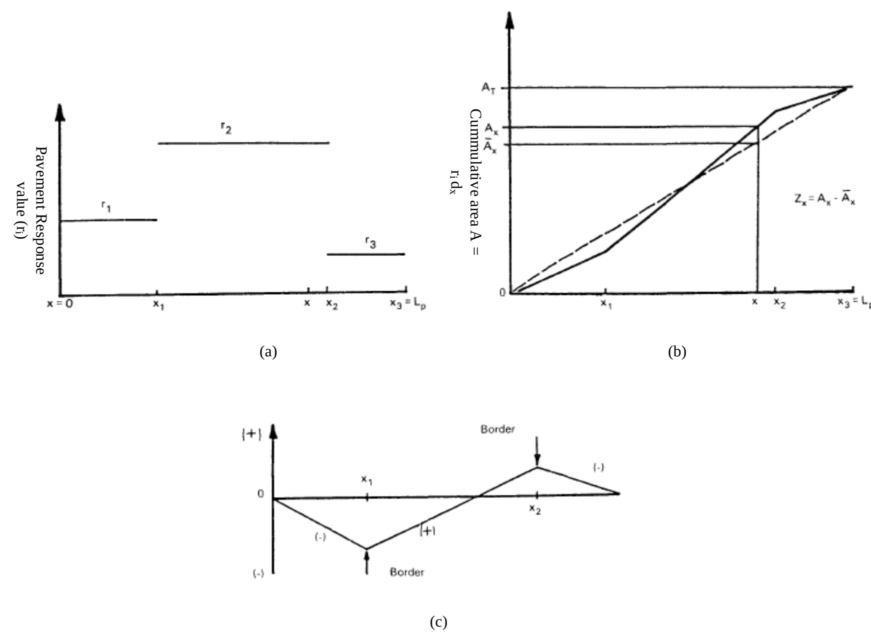

Figure 1: CDA approach (AASHTO 1993)

Since the pavement condition data are collected as point measurement, the numerical difference between pavement responses (i.e., conditions) is calculated using equation 1:

$$Z_x = \sum_{(i=1)}^{n} a_i - \frac{\sum_{i=1}^{n} a_i}{L_p} \sum_{i=1}^{n} x_i \qquad (1)$$

Where a_i is the actual interval area and calculated using equation 2:

$$a_i = \frac{(r_{i-1} + r_i) \times x_i}{2} \qquad (2)$$

Where n is the n^{th} pavement response measurement, r_i is the pavement response value of the i^{th} measurement, and L_p is the total project length.

Many studies identified the limitations of the CDA approach and developed new delineation algorithms to overcome the limitations. For instance, Divinsky et al. (1997), Misra and Das (2004), and El Gendy and Shalaby (2004) identified some limitations associated with the CDA method in determining homogeneous segments. First, the CDA fails to identify more than one homogeneous segment with different average response levels because the method only considered the absolute slope change in magnitudes. Second, the CDA method fails to identify the same homogenous segments when elevating the distress value by a fixed value. Third, the CDA method fails to provide the decision maker with the flexibility to choose the number of homogenous segments. Finally, the delineated segments are significantly influenced by the overall average of the pavement distress values. Based on the limitations of the CDA approach, several studies proposed new delineating methods. Table 1 summarizes those methods and performance indicators used to identify homogeneous pavement segments based on the pavement response data.

Table 1: Methods and performance indicators used to identify homogenous pavement segments

Study	Methods	Response variable
Divinsky et al. (1997)	CDA	Roughness
Ping et al. (1999)	CDA and significance testing	Rut depth
Kenedy et al. (2000)	CDA and significance testing	Roughness
Cuhadar et al. (2002)	Wavelet transform	Generic
Misra and Das (2004)	Classification analysis and regression trees (CART)	Roughness
El Gendy and Shalaby (2004)	Statistical quality control charts-Absolute difference approach	Generic
Thomas (2005)	Bayesian Algorithm	Roughness
Tejeda et al. (2008)	Accumulated sum (CUSUM)	Skid resistance
Yang et al. (2009)	Fuzzy c-mean clustering	Pavement condition rating
D'Apuzzo and Nicolosi (2012)	Cumulative sum or difference, Bayesian algorithm, LCPC (Labortoire Central des Ponts et Chasussees)	Skid resistance

Divinsky et al (1997) modified the CDA approach to delineate pavement condition data using the IRI as a response variable and performance indicator. The authors noted that the delineation method should also consider the scatter characteristics of the response values such as the standard deviation or range value. As such, Divinsky et al. (1997) modified the CDA approach to include the calculation of the response value standard deviation to delineate pavement condition data. Although the modified CDA approach overcame one of the major weaknesses of the CDA approach, it failed to account for considering the variability of other response values. Additionally, the modified CDA approach failed to provide the decision maker with the flexibility to choose the number of homogenous segments.

Ping et al. (1999) combined the CDA method with statistical significance testing (t-test) to identify homogenous pavement segments based on the rut depth using data gathered by Florida DOT. The proposed methodology uses two constraints to delineate rutting values. The first constraint considers achieving a minimum segment length for practicality reasons. The second constraint addresses joining adjacent segments by minimizing the mean rut depth. The study found out that the sum of squared error values increased when the minimum segment length was higher than 0.06 mile or when the rut depth measurements between adjacent segments were very disperse. The approach implemented by Ping et al. (1999) did not consider extreme values of rut depth in segmentation. Similarly, Kennedy et al. (2000) combined the CDA approach and paired t-test significance testing to identify homogenous segments using IRI as response values.

Several studies used complicated algorithms to delineate pavement condition data. Cuhadar et al. (2002) argue that the CDA approach is not effective because of the “noise-like ripples” of the observed data which are averaged. Thus, they developed a wavelet transform waveform that was well concentrated in time and frequency, to automatically delineate the pavement condition data. The wavelet transform approach was found to have high performance for automatic segmentation. The algorithm detects singularities of the smooth waveform and mark them as border points.

Misra and Das (2004) developed a segmentation approach that attempted to overcome the limitations of the CDA using Classification analysis and regression trees (CART). The first step of the approach is to divide the dataset into several subsets by minimizing the value of the mean and the mean squared error of the pavement IRI values. The algorithm keeps dividing the dataset to subsets until the minimum segment length was achieved. The proposed methodology also joined adjacent segments based on statistical similarity. However, this approach did not consider the variations in response data in terms of the existing distresses and their effect on the segmentation process.

El Gendy and Shalaby (2004) also developed a segmentation approach using the quality control charts which considered the variance of the pavement response instead of the average to reduce the response variability within the aggregated pavement sections. The control chart approach works by defining upper and lower limits based on a sample standard deviation. The lower and upper limits are used to determine homogenous segments by keeping the pavement distress value between the limits. When the pavement distress values are beyond the limits, segment border has to set and new segment is determined. However, the lower and upper limit can change based on the response variable.

Thomas (2005) developed an automated road segmentation model using Bayesian algorithm. The algorithm uses a long series of transformed measurements and returns the series into partitioned homogenous segments. Box-Cox transformation is utilized to transform the IRI data before analysis to bring the returned segments to the model assumption of normally distributed observations in each segment. The posterior mode is used to determine the change point in the long series of data. However, the proposed algorithm fails to practically determine additional segment break points after the first break point is determined. As a result, a heuristic part is integrated with the algorithm to place initial change points in a sequential way based on the road construction history.

Tejeda et al. (2008) developed a procedure for specifically delineating skid resistance data to potentially facilitate road safety management. The procedure uses the leverage method to find outlier skid resistance data. Then, the CUSUM method is used to delineate the skid resistance data. The CUSUM method is used to find a point that divides two segments with different means. Finally, the procedure groups adjacent segments using the Student’s t-test of mean equities at 95% confidence level. An algorithm developed by Yang et al. (2009) spatially clusters pavement segments to determine pavement preservation project boundaries. The algorithm uses fuzzy c-mean clustering method in order to minimize the variation in each cluster of pavement segments. Pavement condition rating, an overall pavement condition measure that considers several surface distresses, is used as the response variable in the propose algorithm. The algorithm also considers hard natural boundaries such as bridges, roadway characteristics and so forth. The study also recommended using the detailed segment-level distress to increase the accuracy of the segmentation process.

3 Segmentation Framework

The proposed segmentation is divided into three stages as shown in Figure 2. The first stage of the segmentation algorithm aims at detecting hard boundaries such as pavement characteristics of traffic data. These attributes are usually stored and identified by the agency officials. The output of the first stage is a list of pavement segments that share the traffic and pavement design attributes. Additionally, the start and end points of these segments consider the existing natural barriers such as bridges.

The output of the first stage is then used to find condition-based homogenous segments. First, similar clusters are determined using the affinity propagation algorithm, which is developed by Frey and Dueck (2007). The affinity propagation is a clustering method that finds representative examples “exemplars” and their clusters by exchanging real-valued messages between data points (Frey and Dueck 2007). This clustering algorithm uses the Euclidean distance to measure the similarity between potential exemplars and data points. Unlike other clustering techniques that randomly choose initial subset of data points to find a good solution, affinity propagation considers all data points as potential exemplars (Frey and Dueck 2007). Affinity propagation also clusters the data points in N dimensional space where N represents the number of performance indicators. The affinity propagation also outperforms other techniques such as the k-means clustering and the expectation minimization algorithm since they rely on random sampling to identify the initial clusters which may result in poor solutions. Additionally, the affinity propagation clustering technique has advantage over the hierarchical agglomerative clustering and spectral clustering since these methods do not require all points within a cluster to be similar to a single center as they rely on pairwise grouping (Frey and Dueck 2007).

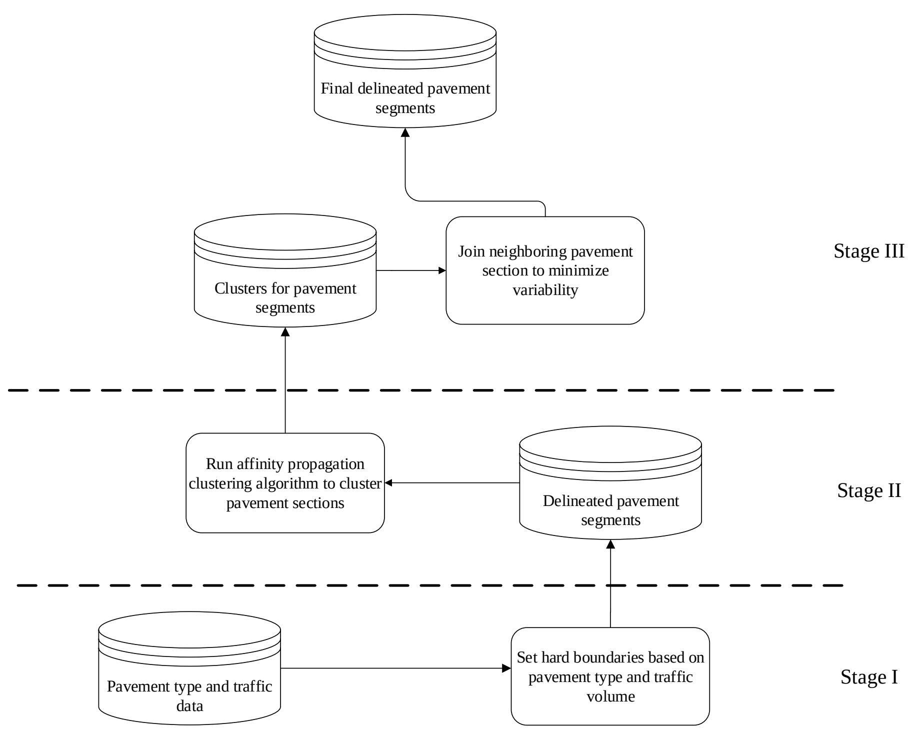

Figure 2: Dynamic segmentation framework

The output of the stage II is a list of clusters that share similar condition data. However, the method used to cluster the pavement sections ignores the geometric continuity feature of pavements. As such, heuristic rules in the third layer are used to detect the most homogenous segments. There are two possible cases that could occur after using affinity propagation for initial clustering as shown in Figure 2. Assuming that there are three clusters A, B and C that are initially identified, the first case represents the existence of a one data point that belongs to cluster B between two clusters A. The second case represents the existence of three different consecutive clusters.

In order to form longer segments and by considering the geometric continuity feature of pavements, the proposed framework allows some variability within the aggregated segments by joining adjacent short segments. The adjacent join of pavement segments should be done to achieve the minimum segment length, which the agency can reasonably use to determine pavement maintenance or rehabilitation treatments. However, the adjacent join of pavement segments should be implemented in a way that reduces the overall condition variability. In this framework, joining one cluster to one of its adjacent clusters is implemented by minimizing the standard deviation of the newly formed cluster. Case (b) in Figure 3 could be used as an illustrative example for the adjacent join constraint. An adjacent join is necessary in this case assuming that the length of the segment in the middle, which belongs to cluster C, is shorter than the specified minimum segment length. Thus, two standard deviations for the pavement response are calculated. The first standard deviation is calculated for the segment AC that represents the join of cluster A and cluster C. Similarly, the second standard deviation is calculated for the segment CB. Finally, the adjacent join is made based on the minimum standard deviation and a list of condition-based homogeneous segments is determined.

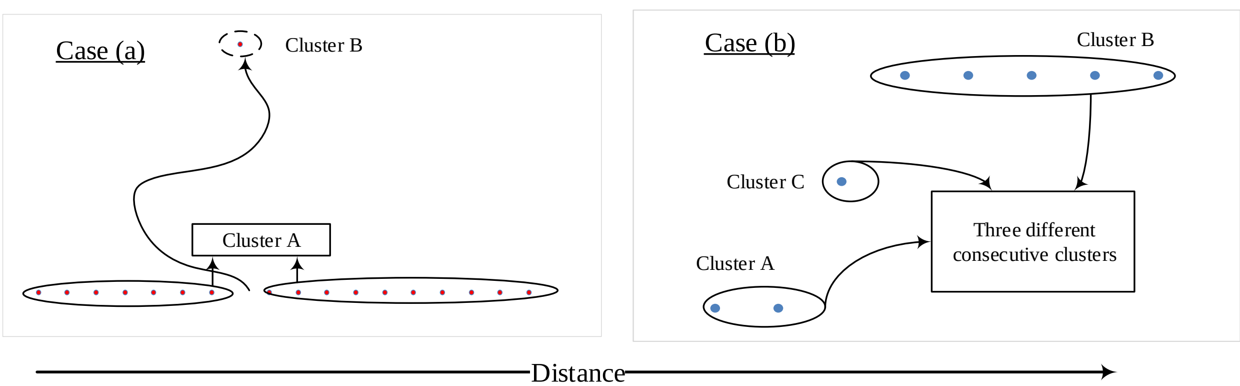

Figure 3: Possible cases of adjacent clusters join to ensure geometric continuity

4 Implementation

A case study was conducted to illustrate the implementation of the proposed framework using raw pavement condition data for a 3.95-mile asphalt concrete (AC) highway in Iowa. Since the Iowa DOT collects pavement condition data every 52 feet and hence the number of data points collected was 396. Each record contains International Roughness Index (IRI), rutting, alligator cracking, transverse cracking, longitudinal cracking, and longitudinal cracking on wheelpath data. The DOT collects rutting and roughness through a vendor that uses a sensor technology. The vendor also takes images and uses image processing algorithms to estimate the cracking data. Afterward, the DOT engineers inspect the image processing results and accept or override the results. Additionally, the Iowa DOT determines the severity level for several distresses including low, moderate and high.

The selected data shared the same traffic and pavement design attributes. The affinity propagation clustering method was first applied to determine the initial clusters among the data points. The ride quality index, rutting index, and crack index were used as the pavement responses for each data point. Each index ranges from 0 to 100 where 0 represents a failure score and 100 represents an excellent condition. The calculation formulas for each index and the failure threshold values were used according to the Iowa DOT practice (Bektas et al. 2015). Table 3 summarizes the threshold values for ride quality, rutting and cracking indexes for AC pavements.

Table 3: Condition Indexes threshold values (Bektas et al. 2015)

Index	Failure Threshold	Remarks
Ride quality	254 in/mile	32 in/mile corresponds a score of 100.
Rutting	0.5 in	
Transverse cracking	483 count/km	
Longitudinal cracking	2640 m/km	Cracks are aggregated by using the coefficients of 1, 1.5, and 2 for low, moderate and high severities.
Longitudinal cracking on wheelpath	2640 m/km	
Alligator cracking	6236 m2/km	

In addition to the individual index for each crack type, an overall cracking index was calculated using the weighted sum of the individual cracking indexes. These weights are 0.2, 0.1, 0.3, and 0.4 for transverse cracking, longitudinal cracking, wheelpath cracking, and alligator cracking respectively (Bektas et al. 2015). Similarly, the pavement condition index (PCI) was also calculated using the weighted sum of cracking index, ride quality index and rutting index. The weights used to calculate the PCI are 0.4, 0.4, and 0.2 for cracking index, ride quality index and rutting index respectively.

In order to reduce the dimensionality of the clustering problem, only three indexes that represent cracking, ride quality and rutting were used to determine the most similar clusters. Thus, the affinity propagation clustering method was conducted to determine similar data points based on the Euclidean distance of cracking, ride quality and rutting indexes. As a result, a total of three clusters were identified and the summary of each cluster's attributes is presented in Table 4.

Table 4: Statistical Summary of Generated Clusters

	Number of data points	Crack index*	Ride quality index*	Rut index*	PCI index*
Cluster 1	152	84.4(8.09)	46.29(12.62)	50.37(10.59)	62.25(5.67)
Cluster 2	119	76.12(11.80)	10.71 (8.75)	31.65 (13.70)	41.06 (7.33)
Cluster 3	125	78.50 (8.74)	40.80 (9.74)	27.83 (8.46)	53.29 (5.82)
Overall	396	80.05 (10.18)	33.87 (18.75)	37.63 (15.02)	53.09 (10.75)

*Average and standard deviation

The first cluster is the best performing cluster since it has better crack index, ride quality and rut index. The second cluster is the worst performer since it has the lowest overall condition with poor ride quality. The third cluster performs better than the second cluster however, it has deeper rut depth than the other two clusters. In summary, there is less variation in terms of cracking but there are considerable variations in terms of ride quality and rutting.

The next step is to use the initially generated clusters as seeds to create longer segments through several iterations to reduce the overall variability. The first iteration starts by joining single different data points to any of its suitable adjacent cluster. The preference of joining the data point to one of its adjacent clusters is constrained by reducing the overall PCI variability of the newly created segment. The next iterations involve joining short segments to its adjacent segments until achieving the minimum segment length. In this study, the minimum segment length is initially set to 0.2 miles. A total of three iterations were performed in order to find the final segments, total of seven segments, shown in Figure 4 and Figure 5.

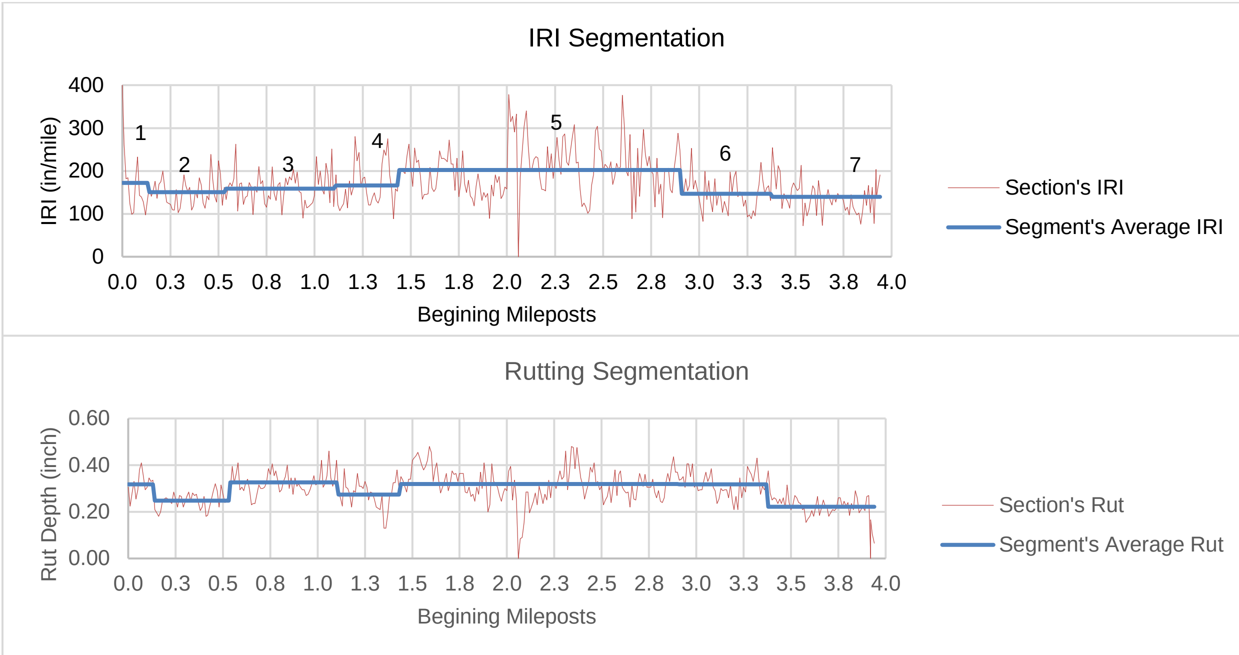

Figure 4: Detected homogenous segments versus original response values (IRI and rutting)

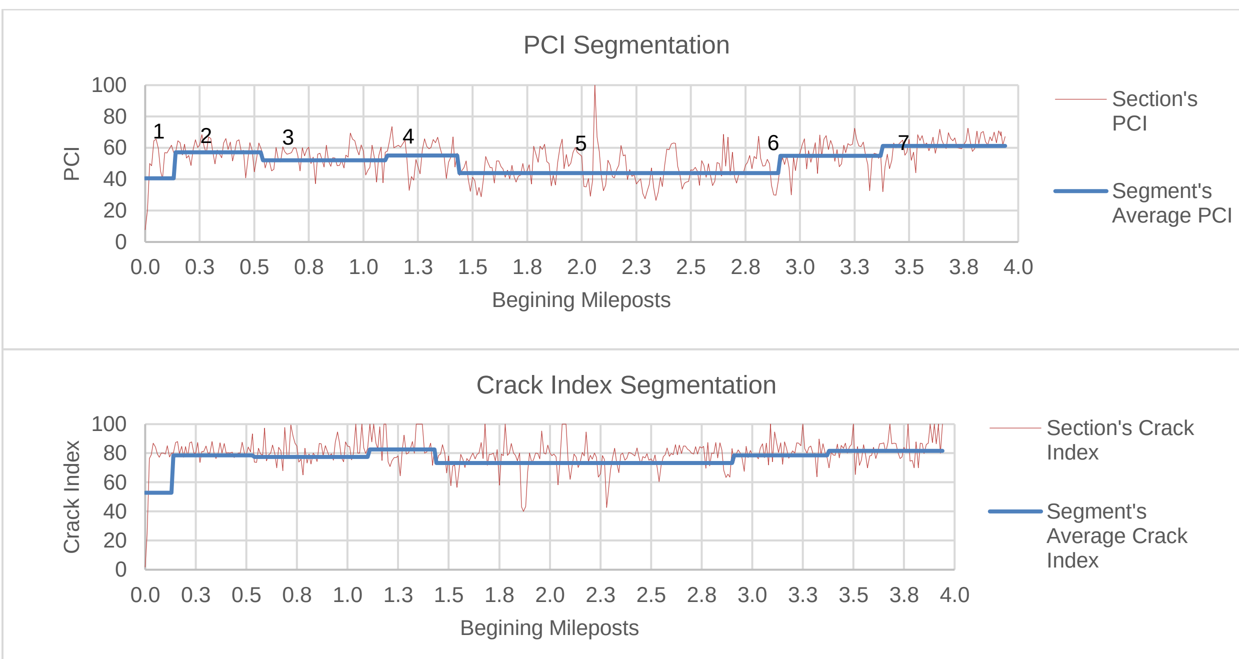

Figure 5: Detected homogenous segments versus original response values (crack index and PCI)

Some of the adjacent segments may have insignificant difference and hence it would be more practical to join them together. For example, the first four identified segments, from left to right, have insignificant differences in terms of IRI. However, there is a significant difference between those segments in terms of rutting. It is worth noting that a t-test was conducted to verify the significance of difference between adjacent segments. This shows the importance of using multi-attributes in detecting homogenous segments.

5 Conclusion and Future Work

This study presented a pavement segmentation framework that considers multiple pavement responses or performance indicators concurrently. The algorithm uses the affinity propagation clustering method, which is a powerful method that identifies similar clusters efficiently. The framework also considers the geometric continuity of pavement features and uses the standard deviation as a measure to join adjacent segments which are shorter than the minimum segment length. The study implemented that proposed segmentation on a 3.95-mile long asphalt concrete highway in Iowa.

The results of the proposed segmentation framework reveal a major limitation associated with the other segmentation algorithms. This limitation is mainly associated with using one performance indicator to delineate the pavement condition data. This study reveals that the difference between adjacent segments is not significant for one performance indicator however, there is a significant difference when considering other performance indicators. The proposed segmentation framework is able to consider multiple performance indicators at the same time which increases the efficiency of the segmentation process.

The proposed segmentation framework is expected to enhance the accuracy and reliability of project level maintenance and rehabilitation decision-making. Accordingly, agencies can predict future funding more accurately. Additionally, the segmentation process would help agencies identify localized deteriorated segments. Thus, agencies can accurately assess the pavement condition and determine the boundaries of pavement segments that exhibits poor condition. As for future work, the proposed algorithm will be fully automated in order to implement the segmentation algorithm at the network level. Additionally, case studies will be developed to quantify the benefits of the proposed method in terms of budget savings and accuracy of condition representation.

References

AASHTO Guide for Design of Pavement Structures. AASHTO, Washington, D.C. 20001, 1993

Bektas, F., Smadi, O., & Nlenanya, I. (2015). Pavement Condition: New Approach for Iowa Department of Transportation. Transportation Research Record: Journal of the Transportation Research Board, (2523), 40-46.

Bennett, C. R. (2004, October). Sectioning of road data for pavement management. In 6th International Conference on Managing Pavements: The Lessons, The Challenges, The Way Ahead.

Calvarese Justin (2007), “QA/QC Oklahoma DOT”, National Conference on Pavement Management, Norfolk, Va.

D'Apuzzo, M. and Nicolosi, V., Detecting Homogeneous Pavement Section using Econometric Test for Structural Changes in Linear Model, in Transportation Research Board 91st Annual Meeting 2012.

Divinsky, M., Livneh, M., & Nesichi, S. (1997). Development of a road roughness profile delineation procedure. Journal of testing and evaluation,25(4), 445-450.

El Gendy, A., and Shalaby, A., Using quality control charts to segment road surface condition data, International conference on managing paving assets, ICMPA 2008, Calgary, AB, June 2008.

Frey, B. J., & Dueck, D. (2007). Clustering by passing messages between data points. science, 315(5814), 972-976.

McGhee, K. H. 2004. Automated Pavement Distress Collection Techniques. NCHRP Synthesis 334. Transportation Research Board, Washington, DC.

Misra, R., and A. Das. Identification of Homogeneous Sections from Road Data, Inb International Journal of Pavement Engineering Vol. 4, Taylor & Francis Ltd., December, 2003, pp. 229 –233.

Ping, W., Yang, Z., Gan, L., & Dietrich, B. (1999). Development of procedure for automated segmentation of pavement rut data. Transportation Research Record: Journal of the Transportation Research Board, (1655), 65-73.

Tejeda, S. V., de Solminihac Tampier, H., & Navarro, T. E. (2008). Proposal of a segmentation procedure for skid resistance data. Arabian Journal for Science and Engineering, 33(1), 89-104.

Thomas, F. (2005). Automated road segmentation using a Bayesian algorithm. Journal of transportation engineering, 131(8), 591-598.

Yang, C., Tsai, Y. J., & Wang, Z. (2009). Algorithm for spatial clustering of pavement segments. Computer-Aided Civil and Infrastructure Engineering,24(2), 93-108.