PAVEMENT CONDITION MODEL BASED ON AUTOMATED PAVEMENT DISTRESS SURVEYS

ALDOT Project 930-758

Final Report

Prepared By:

Dr. David H. Timm
Brasfield & Gorrie Professor of Civil Engineering
238 Harbert Engineering Center
Auburn University
Auburn, AL 36849

Dr. Rod E. Turochy
Associate Professor of Civil Engineering
238 Harbert Engineering Center
Auburn University
Auburn, AL 36849

Sponsoring Agency:
Alabama Department of Transportation
Bureau of Materials and Tests
Pavement Management Section
1409 Coliseum Blvd
Montgomery, AL 36110

May 23, 2014

ACKNOWLEDGEMENT

This project was sponsored by the Alabama Department of Transportation as ALDOT Project 930-758.

DISCLAIMER

The contents of this report reflect the views of the authors who are responsible for the facts and accuracy of the data presented herein. The contents do not necessarily reflect the official views or policies of Alabama DOT or the National Center for Asphalt Technology at Auburn University. This report does not constitute a standard, specification, or regulation. Comments contained in this paper related to specific testing equipment and materials should not be considered an endorsement of any commercial product or service; no such endorsement is intended or implied.

TABLE OF CONTENTS

1 INTRODUCTION

Many government agencies and private consulting companies must deal with the issue of pavement maintenance. The ability to maintain an in-service pavement structure in acceptable condition from structural and functional points of view is related to many factors which are often not explicit and change with time. Although the maintenance strategy may also depend on human experience, data interpretation and agency policies, it is largely founded on the implementation of a pavement management system (PMS).

A PMS is established on the complete inventory of the pavement network and includes basic pavement information such as section type (i.e., layer thickness, type of subbase and subgrade, etc.), location, size, number of traffic lanes, route designations, functional classification and section conditions (i.e., type of distresses, roughness, deflections). In addition, this system incorporates tools and methods to facilitate agencies in the decision-making process to maintain pavements in serviceable and functional conditions throughout their lives.

A tool capable of accurately describing pavement condition represents an important asset for agencies for planning at the network level. Network level planning takes into the account agency's short and long term budget needs to identify and prioritize potential projects. Pavement condition evaluation is one of the most important and difficult processes in pavement management and is strictly connected to the rating of the distresses affecting the infrastructure. Added complexity is found in the variety of measurement procedures employed to establish distress levels on pavement structures.

A pavement evaluation program focuses on collecting pavement condition data to determine the current conditions of the road network. Typically, a pavement evaluation program includes agency- or contracted-manual, semi-automated or automated surveys to assess the type, severity and extent of the deterioration occurring at the pavement surface. These surveys may follow the *Distress Identification Manual for the Long-Term Pavement Performance Program* (FHWA 2003) that was developed by the National Research Council Strategic Highway Research Program (SHRP) in 1999 and later updated in 2003. This distress manual classifies the signs of pavement deterioration into several types of distresses based on specific visual characteristics.

Although manual and semi-automated distress evaluations are done according to well-defined guidelines or criteria, a certain amount of subjectivity and the experience of the raters are expected to have an influence on the ratings. In addition, differences in the distress evaluation of two raters may occur due to a difference in the appearance of the pavement surface depending on the direction and angle of the sunlight, pavement temperature and moisture and the direction from which the raters view the pavement surface (Smith et al. 1996). The subjectivity of the distress measurements affects the final assessment of the pavement condition.

As examples of differences between automated and manually-collected distress measurements, consider the data presented in Figures 1.1 through 1.3 studied in a previous investigation (Timm and McQueen, 2004). The manually-collected data were obtained by Alabama Department of Transportation (ALDOT) personnel while the automated data were generated by an external vendor (Timm and McQueen, 2004). Figure 1.1 shows that the automated rut depth measurements were, on average, 0.18 inches greater than those collected manually with a large degree of scatter contributing to a relatively low R^2 value when the manual measurements are used to estimate the automated counterpart data. Figure 1.2 shows very little correlation between manually and automatically-collected fatigue cracking data ($R^2<0.03$) with much more cracking measured manually than automatically. Finally, Figure 3 shows relatively good agreement between roughness measurements as quantified by the international roughness index (IRI), though the scatter yields an R^2 equal to 0.64. It should be noted that the data sets in Figure 1.3 were both collected automatically. The "Automated" data were collected by Roadware while the "Manual" data were collected by an ALDOT profiler.

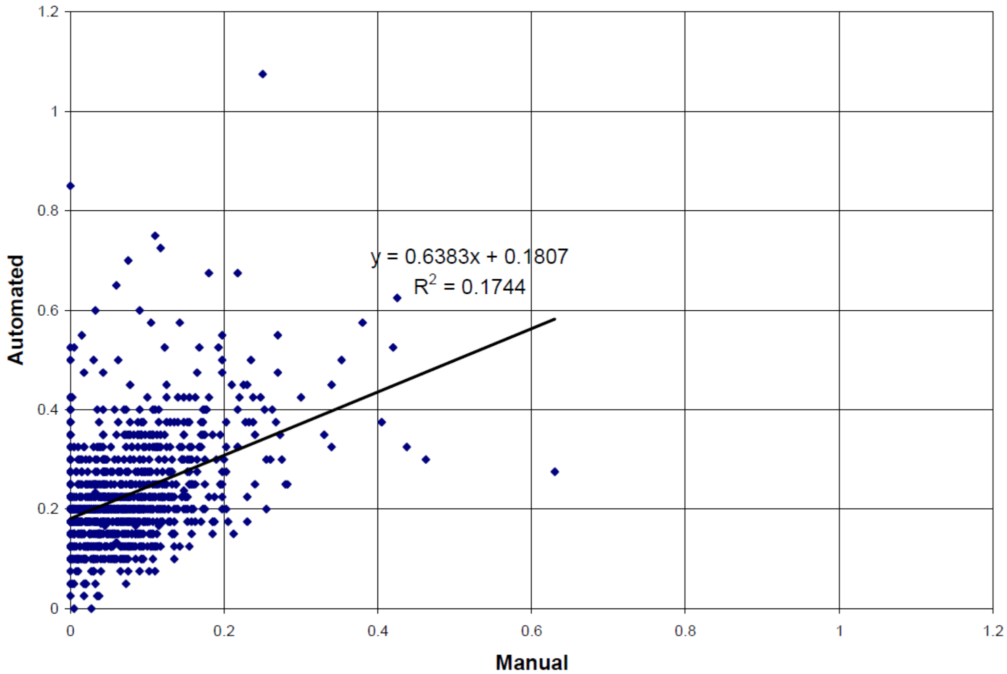

FIGURE 1.1 Automated vs. Manual Rut Depths (inches) (Timm and McQueen, 2004).

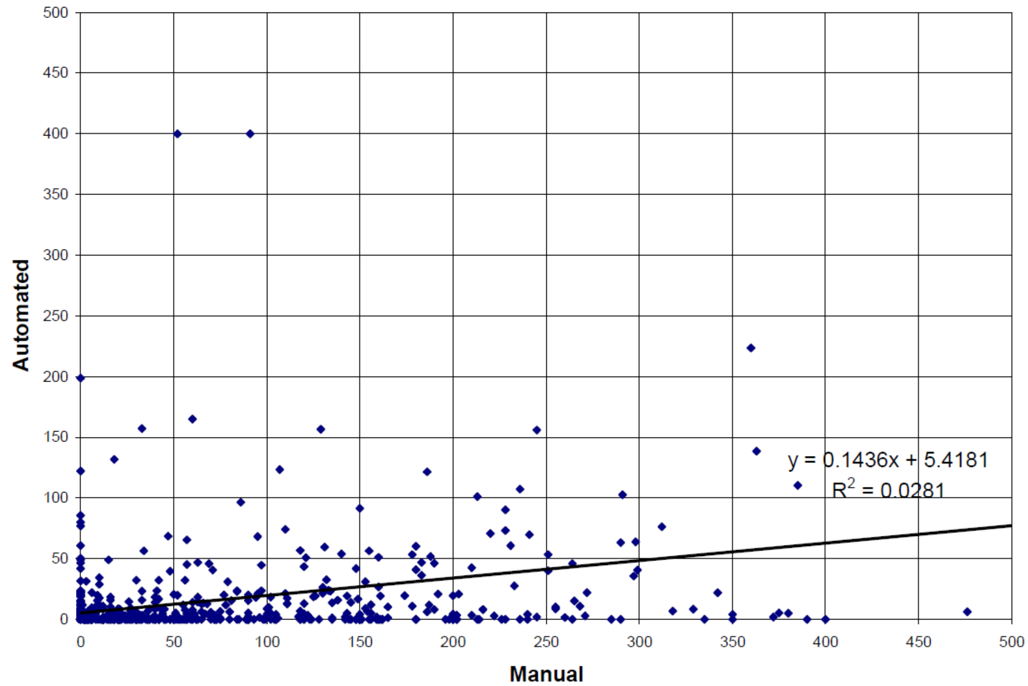

FIGURE 1.2 Automated vs. Manual Automated Alligator Cracking (ft^2) (Timm and McQueen, 2004).

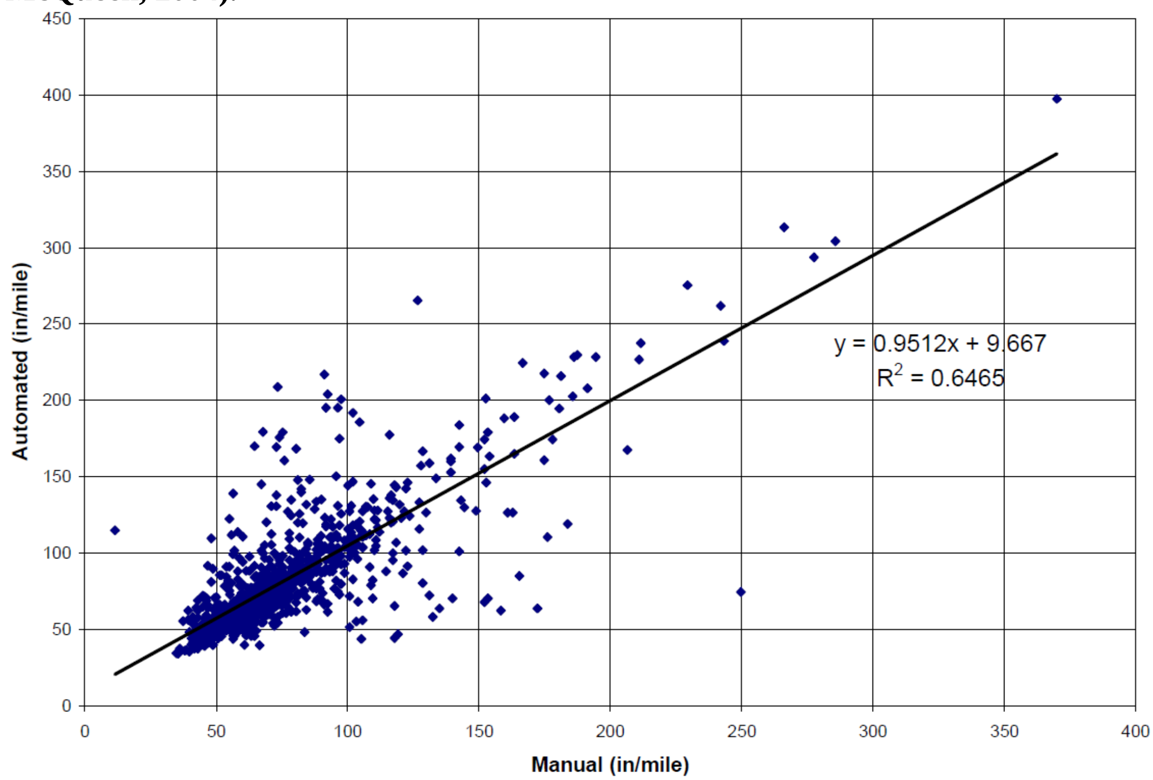

FIGURE 1.3 Automated vs. ALDOT (manual) IRI (Timm and McQueen, 2004).

The data sets shown in Figures 1.1 through 1.3 are certainly not unique to Alabama but are rather symptomatic of differences in data collection and analysis methods. As another example, consider Table 1.1 which shows the Pennsylvania DOT (PENNDOT) acceptable tolerance levels for their automatically-collected distress data. PENNDOT contracts out their automated data collection but performs routine quality assurance, using a PENNDOT automated test vehicle, on 5% of the data with the acceptable tolerance levels listed in Table 1.1. PENNDOT allows up to 5% of the vendor IRI data to exceed ±25% of their own measurements. They allow up to 10% of the other distresses to exceed either ±20% or ±30%, depending on the distress type. Clearly, a low degree of precision is expected when dealing with multiple distress measurements over the same segments of roadway.

TABLE 1.1 PENNDOT's Tolerances for Discrepancies in Data (Timm and McQueen, 2004)

Reported Value	Initial Criteria	Percent within Limits	Recommended Action if Criteria Not Met
IRI	±25%	95	Reject deliverable
Individual Distress Severity Combination	±30%	90	Feedback on potential bias or drift in ratings, retrain on definitions.
Total fatigue	±20%	90	Reject deliverable
Total Non-Fatigue Cracking	±20%	90	Reject deliverable
Total Joint Spalling	±20%	90	Reject deliverable
Transverse Cracking, JCP	±20%	90	Reject deliverable

Regardless of how the data are collected, many agencies develop a composite index based upon distress measurements for pavement management purposes (e.g., WSDOT, 1999; GDOT, 1996). Sometimes called a pavement condition rating (PCR) or pavement index (PI), these indices are often used to prioritize roadway segments for maintenance and rehabilitation. The subjectivity inherent in assessing pavement condition means that previously-developed equations must be reassessed whenever the measurement system changes. This reality is regardless of whether the system is manual or automated. In fact, even when switching between similar automated systems, there are likely to be changes in measured pavement distresses that would necessitate changes in the pavement condition rating.

One such equation, employed by the Alabama Department of Transportation (ALDOT), includes a variety of pavement performance parameters. The equation, shown below, was developed at the University of Alabama and was based upon expert opinion and manually-collected data (Turner et al., 1985):

$$PCR = 95.5727 - 5.5085\,(5.0\text{-}ROUGH) - 1.5964\,LNALL1 - 1.9629\,LNALL2 - 2.9795\,LNALL3 - .01630\,PAT2RD - .07262\,BLK2RD - .2220\,AVGOUT - 3.4948\,RAVL31 - 7.5269\,RAVL32 - 11.2297\,RAVL33 - .03032\,LONG12 - .05484\,LONG34 - .53050\,TRAN12 - .69736\,TRAN34 \quad (1.1)$$

where:
PCR = pavement condition rating
ROUGH = roughness or present serviceability index (0-5)
LNALL1 = ln (level 1 alligator cracking + 1.0)
LNALL2 = ln (level 2 alligator cracking + 1.0)
LNALL3 = ln (level 3 alligator cracking + 1.0)
PAT2RD = patching (level 2 + level 3), $\leq 400\ ft^2$
BLK2RD = block cracking (all levels summed), $\leq 400\ ft^2$
AVGOUT = outer wheel path rutting (all locations averaged), 10^{-2} inches
RAVL31 = severe localized raveling (Code: 0 = none, 1 = present)
RAVL32 = severe wheel path raveling (Code: 0 = none, 1 = present)
RAVL33 = severe entire lane raveling (Code: 0 = none, 1 = present)
LONG12 = longitudinal cracking (level 1 + level 2), ft
LONG34 = longitudinal cracking (level 3 + level 4), ft
TRAN12 = transverse cracking (level 1 + level 2), number of cracks
TRAN34 = transverse cracking (level 3 + level 4), number of cracks

A previous study conducted at Auburn University examined manually-collected data versus automated data to determine their relative impact on computed pavement condition ratings (Timm and McQueen, 2004). The study determined that automated and manually-collected data sets were fundamentally different and therefore resulted in different pavement condition scores using the above equation. Though the study pointed out the differences and the need to improve the correlation between manual and automated data, it did not suggest developing a new equation. A subsequent study at the University of Alabama and discussions with ALDOT pavement management engineers have brought to light the need to revise this equation to lead to an effective pavement management system.

The fact that the data sets do not agree to an acceptable degree is a difficult, if not impossible, problem to resolve. Improvements in automated technology may help in the future, but differences in ratings made by individual raters highlight the challenge of objectively quantifying some forms of pavement distress and the overall pavement condition. Therefore, the focus should be on generating consistent pavement condition ratings rather than the distresses leading to those ratings. The benefit of this approach overcomes reliance on a particular technology but still results in pavement condition ratings consistent with the "ground-truth." Two such approaches, described below and used in this investigation, are the application of artificial neural networks and re-calibration of the original PCR equation.

The application of the artificial neural network (ANN) approach in the field of pavement management presented encouraging results and suggested further research ideas over these new methodologies. The literature reports many studies on ANNs specifically applied to pavement performance and structural assessment (e.g., Eldin and Senouci, 1995; Roberts and Attoh-Okine, 1998; Saltan et al., 2002; Loizos and Karlaftis, 2006). ANNs represent an excellent tool for dealing with the complexity of the pavement structure and the non-linearity of the measured data. Expressing a complex system through neural networks proved to overcome the limitations of classical modeling techniques such as finite element or statistical methods. ANNs are characterized by their extreme flexibility and adaptability to the system and to the information available in describing it. Thus, it may be possible to train an ANN to overcome the discrepancies between different pavement distress measurement techniques to arrive at comparable pavement condition ratings.

Another approach to handling discrepancies in the measured data is to simply re-calibrate the existing PCR equation. Theoretically, this could be accomplished using existing ALDOT pavement management data sets collected through manual and automated means. The goal of this recalibration would be to arrive at the same pavement condition rating using both sets of data. The original ALDOT equation presented above would be used to establish the "ground-truth" PCR value for individual segments in the recalibration exercise. Data collected through automated means, on the same segments, would be used as inputs to recalibrate the original equation to generate a new equation having new calibration coefficients. This would be accomplished through non-linear least-squares regression.

In either case, there is a clear need for a method of updating PCR predictions that accounts for differences in measured pavement distresses. The method should rely upon a known "ground-truth" set of measurements to provide calibration or training points for the new model. This will allow ALDOT pavement managers to have consistently generated pavement condition ratings from which to quantify current condition and predict future pavement performance.

2 OBJECTIVE

Given the background information and needs described above, the primary objective of this research was to develop a methodology for updating ALDOT pavement condition ratings to reflect past experience in pavement management using new means of distress data collection.

3 SCOPE OF WORK

This research explored two main approaches. The first was to develop ANNs while the second was to recalibrate the existing ALDOT PCR model through regression analysis. The ANN training and PCR recalibration relied on a sample of pavement performance data collected statewide in 2009 and 2010 through automated means paired to manually-collected ALDOT "ground-truth" distress measurements. Further evaluation was conducted using all automated data collected in ALDOT 2nd Division in 2009. Finally, independent validation of the revised model was conducted using automated and manually-collected distress measurements collected in 2011 from 10 quality control segments identified by ALDOT.

4 DEVELOPMENT OF REVISED MODEL

The development of the revised PCR model was divided into two main approaches. The first was accomplished through ANN modeling while the second was to perform non-linear least-squares regression. ANN modeling, generally-speaking, can be very powerful in that the model simply recognizes patterns in the data and "learns" how to make predictions based on past experience. The disadvantage of ANN modeling, however, is that ANN models tend to be complicated black boxes that do not provide a fundamental understanding of the relative influence of the various input parameters on the output predictions. ANN models also require relatively large amounts of training data to arrive at stable solutions as will be demonstrated below.

Alternatively, statistically-based regression models are generally well-understood and the relative influence and importance of the input parameters can be easily ascertained and quantified statistically. However, they may suffer from over simplicity and may lack the ability to capture all the trends between inputs and outputs.

Regardless of the method of model development, the effectiveness can be generally judged by how well each is able to match its PCR predictions to the "ground-truth" PCR. As will be described in the following subsections, there are a number of statistical and practical means of judging how well distress measurements and PCR predictions made through automated means match the ground-truth measurements and PCRs.

4.1 Revised Model Data Set

The data used for model revision was collected in 2009 and 2010. Table 4.1 lists the year of collection, the ALDOT route designation and the beginning and ending milepoints for each segment. The ground-truth data were collected manually by the ALDOT Pavement Management Division within the Bureau of Materials and Tests. The automated data were collected through Pathway, a vendor providing statewide distress data to ALDOT. The data were compiled on a 0.01 mile basis and the number of paired data points for each segment is shown in Table 4.1. For this investigation, there were 570 total paired data points.

On each segment in Table 4.1, the pavement distress data listed in Table 4.2 were collected. It should be noted that the International Roughness Index (IRI) and rut depths were only measured by Pathway and were taken as ground-truth measurements in this investigation. Past studies had shown generally good agreement between manual and automated measurements for these distresses (Timm and McQueen, 2004), so they were deemed acceptable as ground-truth measurements. The ALDOT-measured cracking was determined through walking surveys on-site while the Pathway cracking measurements were made off-site from high resolution images of the pavement surface using their normal image processing procedures.

TABLE 4.1 Revised Model Data Set

YEAR	Route	Starting Milepoint	Ending Milepoint	Paired Data Points
2009	AL0003	209	209.29	30
	AL0008	70.71	71	30
	AL0014	191	191.29	30
	AL0021	166	166.29	30
	AL0022	4	4.29	30
		37	37.29	30
	AL0053	90	90.29	30
		95	95.29	30
	AL0063	2	2.29	30
		6	6.29	30
	AL0110	10	10.29	30
	AL0271	1	1.29	30
2010	AL0001	120	120.29	30
		120.71	121	30
	AL0022	130	130.29	30
	AL0077	35	35.29	30
	AL0148	18	18.29	30
	AL0169	12	12.29	30
	AL0259	2	2.29	30
Total				570

TABLE 4.2 Manual and Automated Distress Measurements

Measurement	Manual-ALDOT	Automated-Pathway
Low severity transverse cracking	✓	✓
Medium severity transverse cracking	✓	✓
High severity transverse cracking	✓	✓
Low severity wheelpath cracking	✓	✓
Medium severity wheelpath cracking	✓	✓
High severity wheelpath cracking	✓	✓
Low severity nonwheelpath cracking	✓	✓
Medium severity nonwheelpath cracking	✓	✓
High severity nonwheelpath cracking	✓	✓
Outside wheelpath rutting		✓
Inside wheelpath rutting		✓
Outside wheelpath IRI		✓
Inside wheelpath IRI		✓

4.2 Comparison of Manual and Automated Data Sets

Prior to developing a new ANN or recalibrating the ALDOT PCR equation, the cracking data set collected by ALDOT and Pathway were compared to identify potential discrepancies between the data sets. The comparisons were grouped according to types of cracking (i.e., transverse cracking, wheelpath cracking and non-wheelpath cracking).

Figure 4.1 illustrates the transverse cracking comparison. Visual inspection clearly shows large discrepancies between the two sources of data and very little, if any correlation. Paired t-testing (α=0.05) was conducted for each transverse cracking severity level from which the Pearson correlation coefficient (r) and two-tailed p-value was determined. Table 4.3 summarizes the t-test results. The low and moderate severity cracking clearly have poor Pearson correlation ($r<0.57$) and are statistically different ($p < 0.025$). The high severity cracking was somewhat a special case. Zero transverse cracking was reported for every segment in the automated data set while there were two manual non-zero readings. The result from t-testing these data sets was an undefined Pearson correlation coefficient, but the p-value indicates a lack of a statistically significant difference between the data sets. This is a non-sensical result and it can be generally stated that the transverse cracking readings had generally poor agreement.

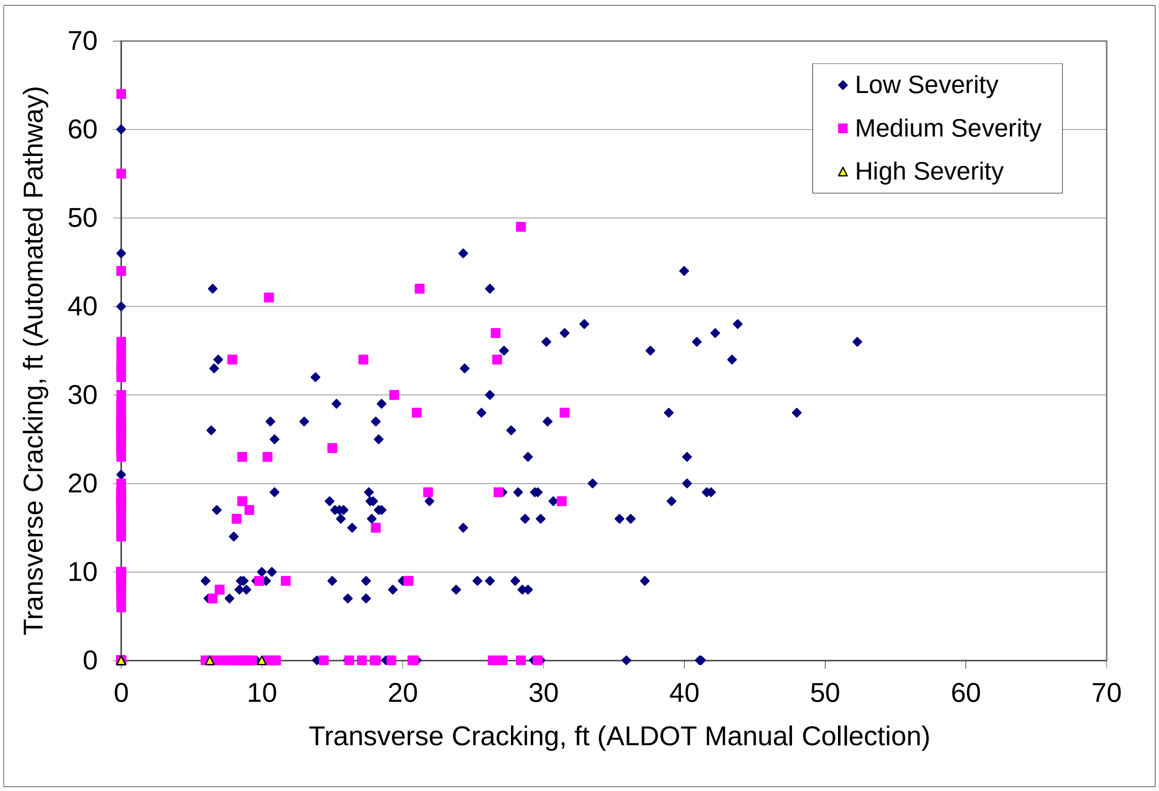

FIGURE 4.1 Transverse Cracking Comparison.

TABLE 4.3 Transverse Cracking Paired t-Test Summary

Transverse Cracking Severity	Pearson Correlation Coefficient (r)	Two-tailed p-value
Low	0.568	0.018
Medium	0.277	1.809E-9
High	Undefined	0.168

Figure 4.2 shows the wheelpath cracking comparison. Wheelpath cracking can also be called alligator or fatigue cracking and is sometimes reported as an area. For consistency, the automated and manual data were both collected and reported as lineal feet of cracking over 1/100 mile or 52.8 ft. Therefore, the maximum possible value was 52.8 ft with many segments having this value.

Visual inspection of Figure 4.2 clearly shows very poor correlation of the data sets. Paired t-testing (α=0.05) was executed for each wheelpath cracking severity level from which the Pearson correlation coefficient and two-tailed p-value was determined. Table 4.4 summarizes the t-test results. All cracking levels had very low correlation ($r < 0.26$) and were statistically different. What is perhaps most striking is the large number of instances in Figure 4.2 where large amounts of cracking was reported by one method but not the other, as evidences by the high degree of scatter in that data and proportion of data points that fall on one of the four boundary axes of the chart. This clearly would pose a problem for future model development.

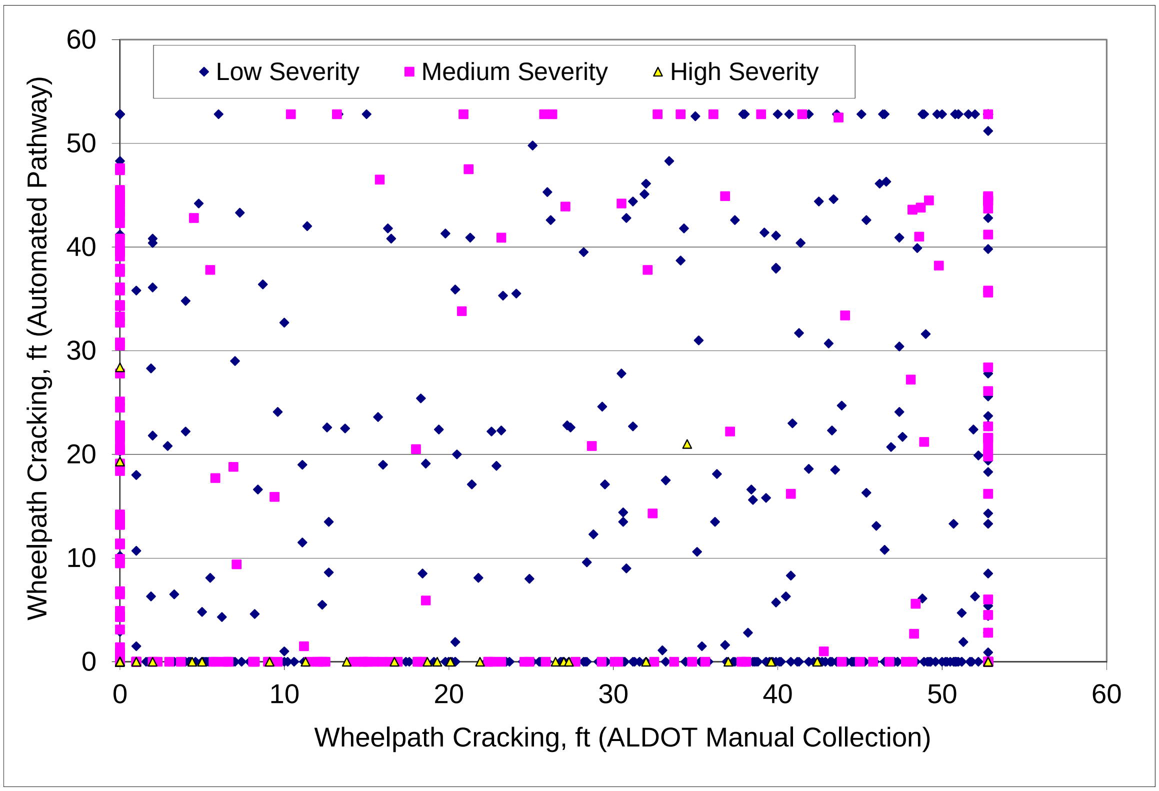

FIGURE 4.2 Wheelpath Cracking Comparison.

TABLE 4.4 Wheelpath Cracking Paired t-Test Summary

Transverse Cracking Severity	Pearson Correlation Coefficient	Two-tailed p-value
Low	0.248	1.942E-39
Medium	0.250	6.923E-4
High	0.102	1.400E-4

The last comparison was made upon the non-wheelpath cracking. Measured in a similar manner to the wheelpath cracking, it was again reported in terms of lineal feet per 52.8 ft segment. Figure 4.3 shows generally poor agreement between automated and manual data sets. Table 4.5 summarizes the paired t-testing with low Pearson correlation coefficients and very low p-values indicating large differences between these data sets. The high severity cracking again had the problem of many pairs with zero cracking reported and a few manually-reported cracking that yielded an undefined Pearson coefficient.

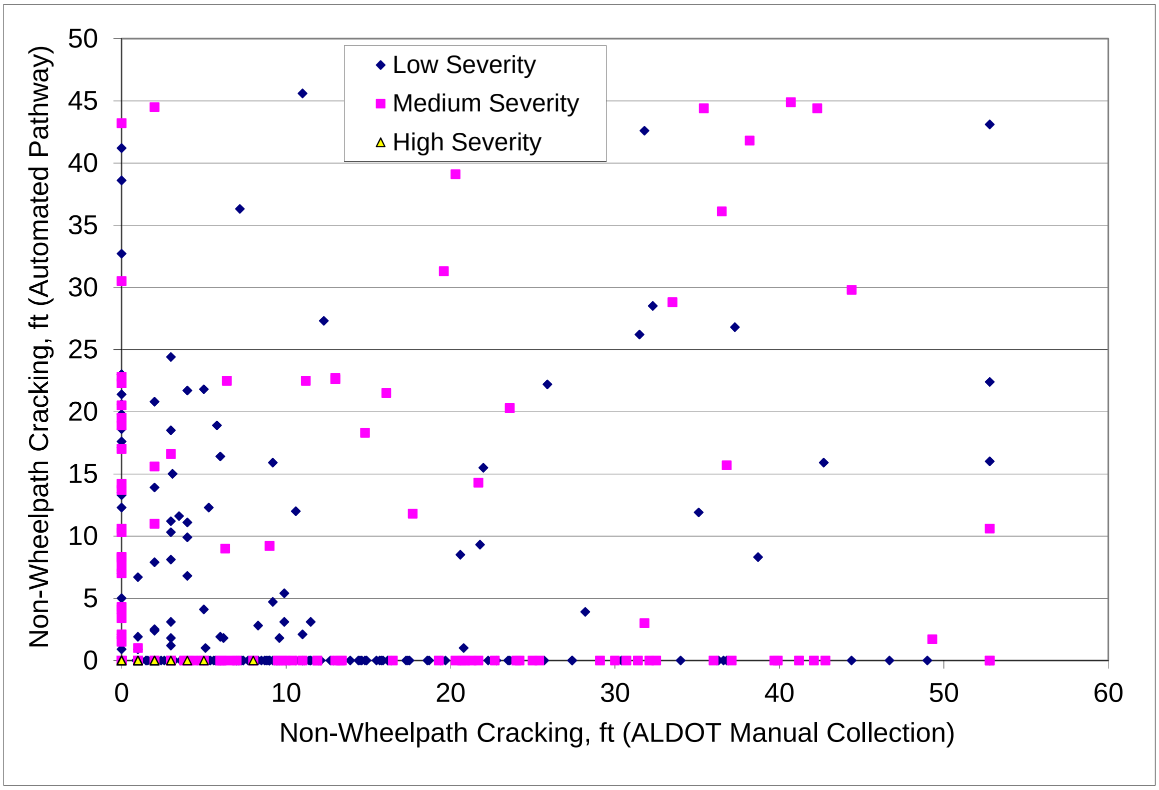

FIGURE 4.3 Non-Wheelpath Cracking Comparison.

TABLE 4.5 Non-Wheelpath Cracking Paired t-Test Summary

Non-Wheelpath Cracking Severity	Pearson Correlation Coefficient	Two-tailed p-value
Low	0.241	2.236E-15
Medium	0.391	2.072E-4
High	Undefined	1.049E-3

The three data sets presented above showed generally poor correlations between the two methods of data collection. While this finding is important from a data collection standpoint, the larger question was how much these discrepancies would affect predictions of pavement condition rating through neural network modeling and regression analysis as explored in the following sections.

4.3 Artificial Neural Network Modeling

ANN modeling relies upon building a model from known data by recognizing patterns in the data. More technically, as described by Leiva-Villacorta (2012), ANNs consist of a number of very simple and highly interconnected processors called neurons, which are the analogs of biological neurons in the brain. The neurons are connected by a large number of weighted links, over which signals can pass. Figure 4.4 illustrates the concept of a neural network where the input layer includes the measured data such as IRI, rutting and cracking. The output layer represents the pavement condition rating. In between are the so-called hidden layers comprised of neurons that connect the input and output layers.

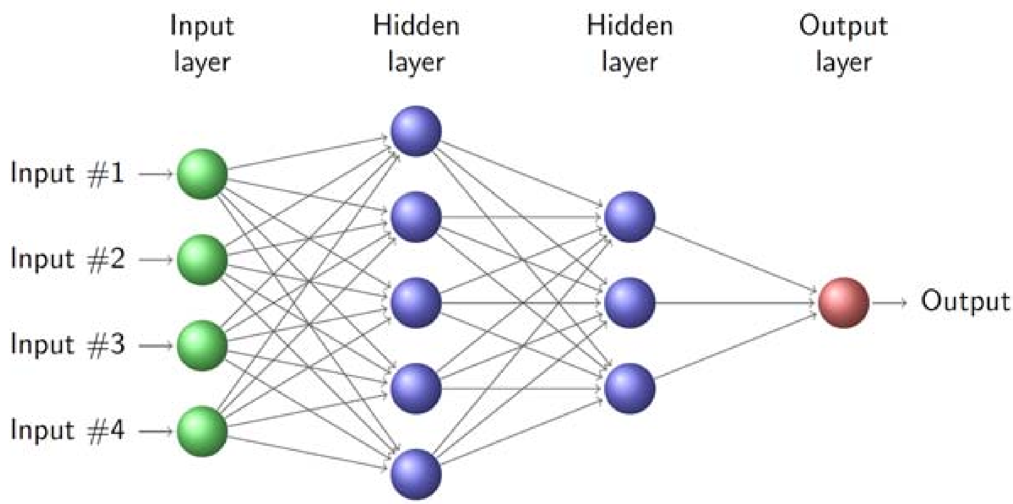

FIGURE 4.4 Artificial Neural Network Schematic (BW Mining, 2013).

ANNs rely upon a training process using known data sets. Typically, the process includes using a large percentage of the known data set for actual training and smaller proportions for validation and testing. For this investigation, the training was conducted in MATLAB and 70% of the data were used for training while 15% were used for validation and 15% for testing. MATLAB randomly selected the data for each subset (training, validation and testing) from the 570 records described in Section 4.1.

To develop the ANN data set, PCRs were computed for the 570 records using a modified version of the original ALDOT PCR equation (equation 1.1) where patching (PAT2RD), block cracking (BLK2RD) and raveling (RAVL31, RAVL32 and RAVL33) had been removed at the direction of ALDOT. Rating the patching severity is highly subjective and depends heavily on the method of collection (walking survey vs. downward-fired imaging and post-processing). Therefore, the vendor was only asked to report the presence of patching (not severity level) and it was removed from the equation. Furthermore, the automated system reported non-wheelpath cracking which could technically include longitudinal cracking and block cracking. Since block cracking is more rarely observed on state routes than longitudinal cracking, ALDOT decided that the better option for network-level data was to assume that the non-wheelpath cracking was longitudinal rather than block cracking. For this reason, block cracking was removed from the equation. Finally, in the original equation, raveling required classification according to severity level. However, raveling severity is also highly subjective and the vendor was asked to only report high severity raveling. Therefore, raveling was also removed from the equation. The modified equation was:

$$\begin{aligned} PCR = {} & 95.5727 - 5.5085\,(5.0\text{-}ROUGH) - 1.5964{*}LNALL1 \\ & - 1.9629{*}LNALL2 - 2.9795{*}LNALL3 - 0.2220\ AVGOUT \\ & - .03032{*}LONG12 - .05484{*}LONG34 - 0.53050{*}TRAN12 - .69736{*}TRAN34 \end{aligned} \tag{4.1}$$

The roughness term (ROUGH) in equation 4.1 was computed from the measured IRI according to an equation previously developed by ALDOT (Holman, 1995):

$$ROUGH = 5e^{-0.0051118*IRI-0.0016027} \tag{4.2}$$

where:
ROUGH = roughness or present serviceability index (0-5)
IRI = international roughness index, in./mile

Equations 4.1 and 4.2 were used with the ALDOT-measured data to determine a PCR for each of the 570 records. These ground-truth PCR values were used as the output layer for training the ANN. The input layer utilized the Pathway-measured data and the ANN was trained to map the Pathway inputs onto the ground-truth PCR values.

Once all the data had been organized and loaded into MATLAB, it was a relatively simple matter to run the training algorithm. Multiple training cycles were executed to evaluate whether consistent results were achievable. Table 4.6 and Figures 4.4 through 4.6 illustrate the results of three training cycles output from MATLAB. Each figure is comprised of four plots with common elements. The top-left plot shows the training data set with the Pathway-computed PCR on the y-axis and the ALDOT PCR on the x-axis. Ideally, the data should cluster along the dashed line of equality (Y=T). MATLAB also graphs a best-fit (Fit) solid linear trendline as an indication of the departure from equality. The y-axis label expresses the best-fit parameters where “Output” is the Pathway PCR and “Target” is the ALDOT PCR with the corresponding slope and intercept. The correlation coefficient between the data sets (R) is shown above each plot. Again, under ideal conditions, the slope should approach 1 while the intercept approaches zero and the correlation coefficient approach 1. The remaining plots in each figure represent the validation (top-right), testing (bottom-left) and all (bottom right) data, respectively.

Visual inspection of Figures 4.4 through 4.6 generally show poor results with a great deal of scatter. No ALDOT PCRs were computed below 11.3 while the ANN model, in an effort to provide acceptable matches to the ALDOT data, computed PCRs down to zero. This explains the large amount of data stacked vertically at 11.3 on the x-axes.

Table 4.6 summarizes the results of the ANN training. It is clear from the table that the training was non-unique in that various results were achieved for each training cycle depending upon on how the data were randomly drawn for each subset. The intercept ranged from 8.4 to 20 while the slope varied between 0.51 to 0.68. Table 4.6 also contains the R^2 for each linear trendline computed from the reported correlation coefficient which ranged from 0.36 to 0.63. While one could argue that an R^2 of 0.62 indicated reasonable accuracy, the other lower values were deemed unacceptable. In terms of predicting PCR from automated data, these results were deemed unacceptable and it was decided to proceed with statistical regression as described in the next section.

TABLE 4.6 Summary Results of ANN Training

Training Cycle	Component	Training	Validation	Test	All
1	Pathway =	0.51*ALDOT+ 11	0.52*ALDOT+ 12	0.51*ALDOT+ 11	0.52*ALDOT+ 11
	R^2	0.55	0.50	0.61	0.55
2	Pathway =	0.68*ALDOT+ 10	0.68*ALDOT+ 8.4	0.53*ALDOT+ 20	0.65*ALDOT+ 11
	R^2	0.62	0.74	0.36	0.58
3	Pathway =	0.61*ALDOT+ 9.9	0.53*ALDOT+ 13	0.62*ALDOT+ 11	0.60*ALDOT+ 10
	R^2	0.63	0.44	0.58	0.59

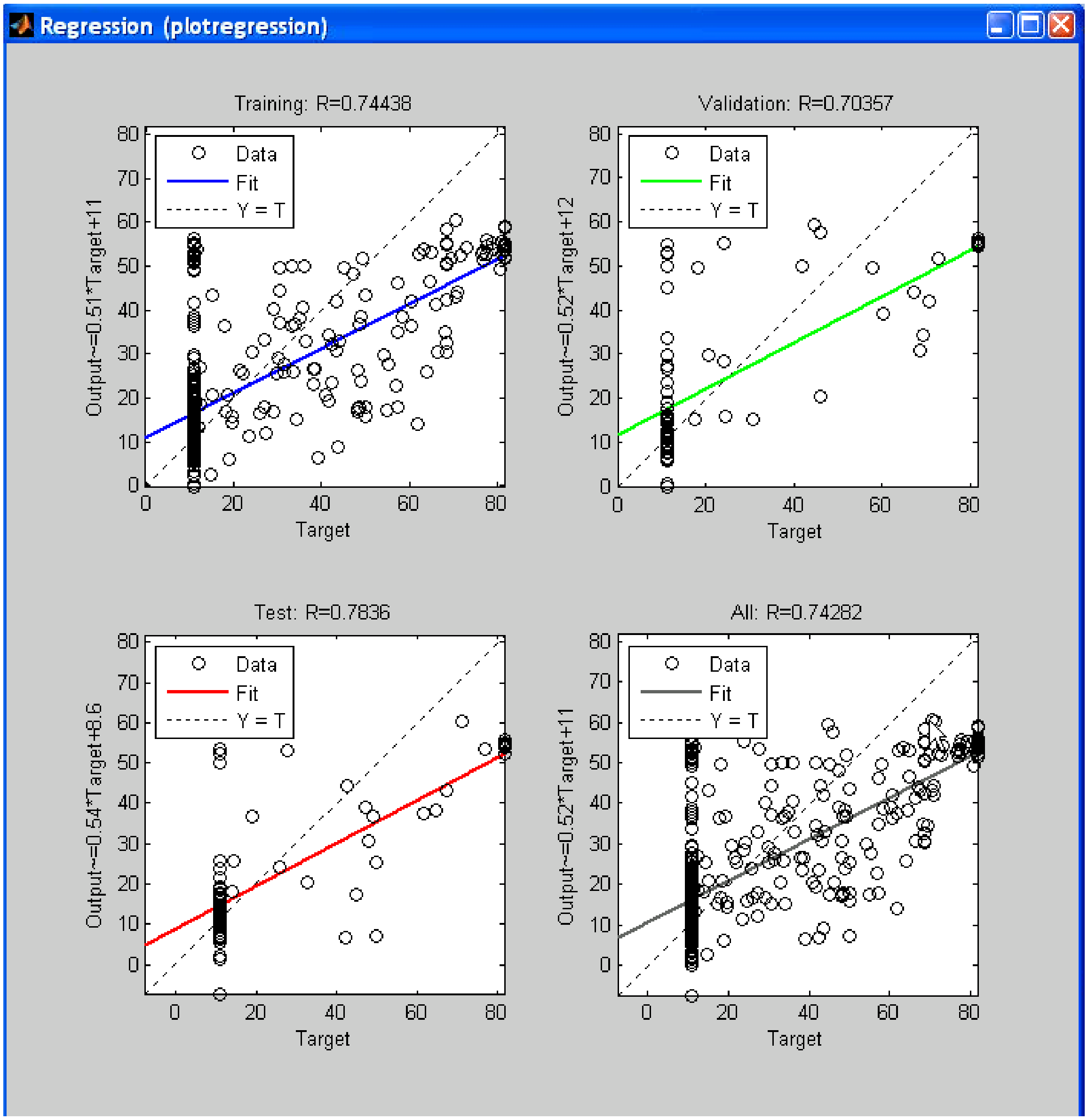

FIGURE 4.4 ANN Training Cycle One.

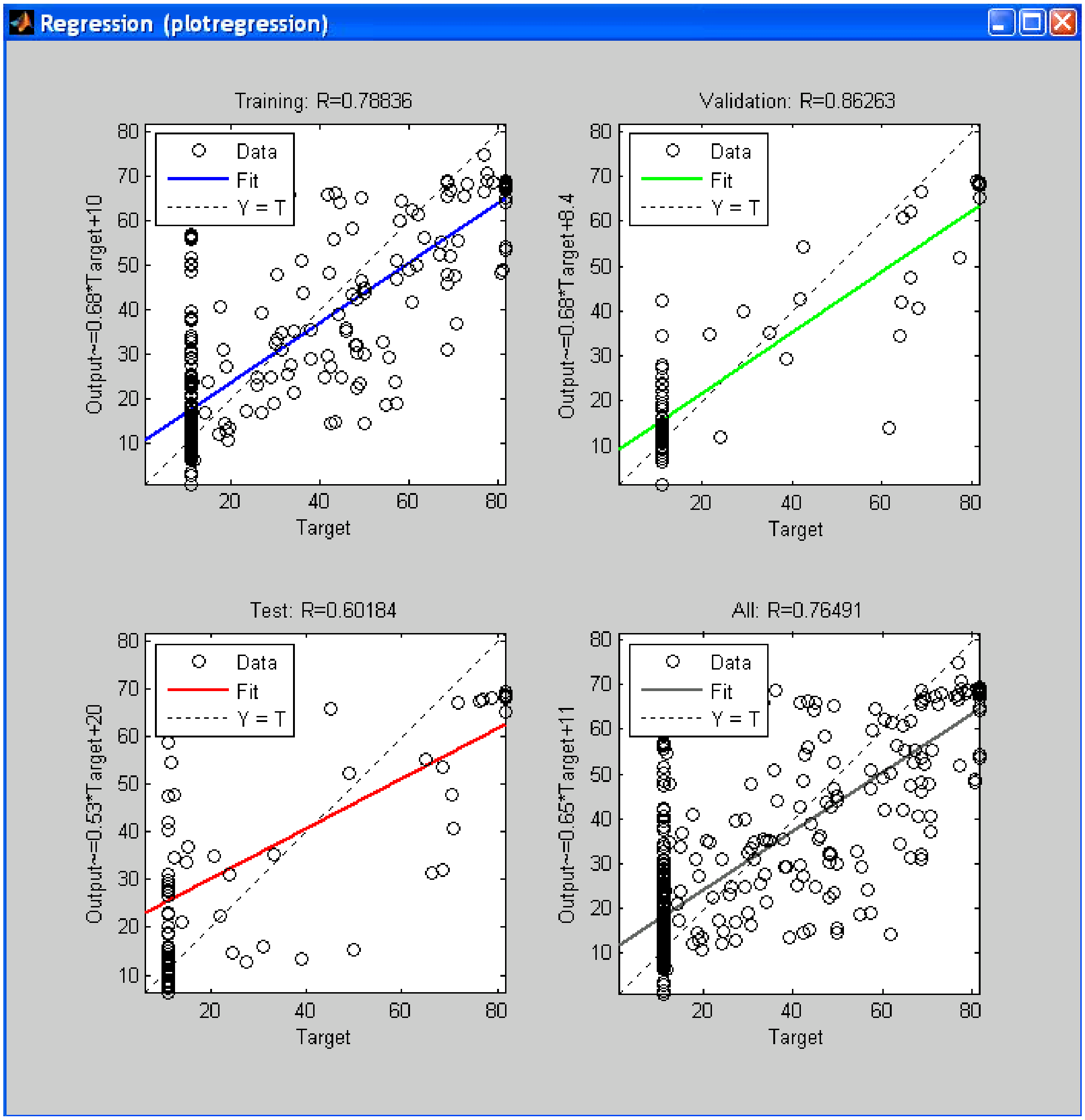

FIGURE 4.5 ANN Training Cycle Two.

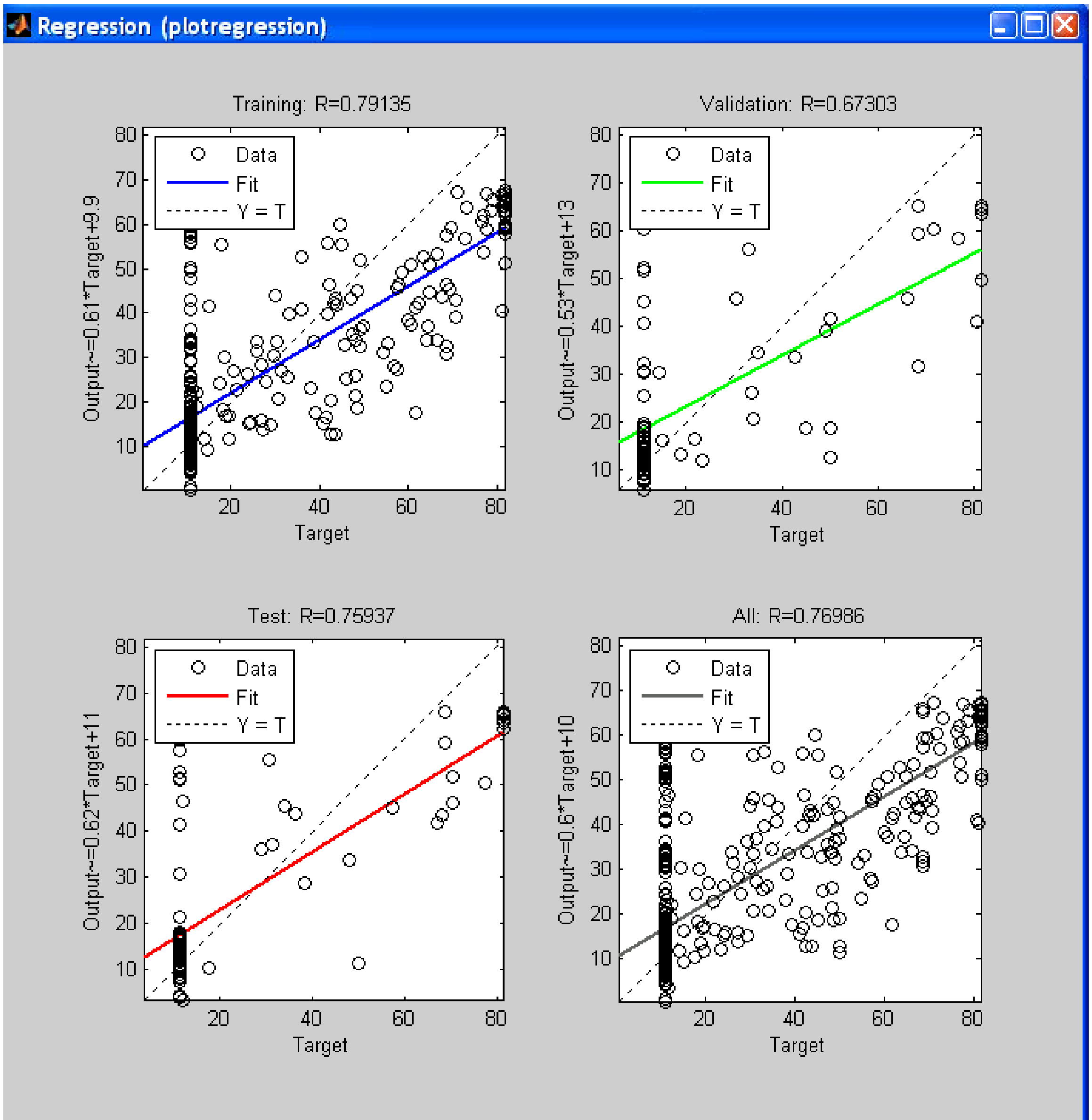

FIGURE 4.6 ANN Training Cycle Three.

4.4 Regression Modeling

As noted in the previous section, the ANN model proved relatively inconsistent and highly dependent upon the random sample drawn for training purposes. Furthermore, even if the model had been reliable, ANNs suffer from being black boxes that can be difficult to use in practice. The alternative approach, regression modeling, does not suffer from these limitations and has the potential to provide improved PCR values from automated data. The approach was to use the same computed PCR values from the ALDOT ground-truth data using equations 4.1 and 4.2. A modified equation, having the same form of equation 4.1, was then calibrated through multivariable, non-linear, least squares regression with the automated Pathway data.

Prior to performing the recalibration, the existing equation (4.1) was evaluated to judge the current status of the ALDOT PCR equation using the automated data. Figure 4.7 compares the Pathway and ALDOT PCR values using equation 4.1. The line of equality and lines indicating ±10 PCR points show that a large number of data points (63%) fall within this practical range of tolerance. However, there does appear to be some bias in the data with 33.2% exceeding the +10 tolerance limit with only 3.8% below the -10 limit.

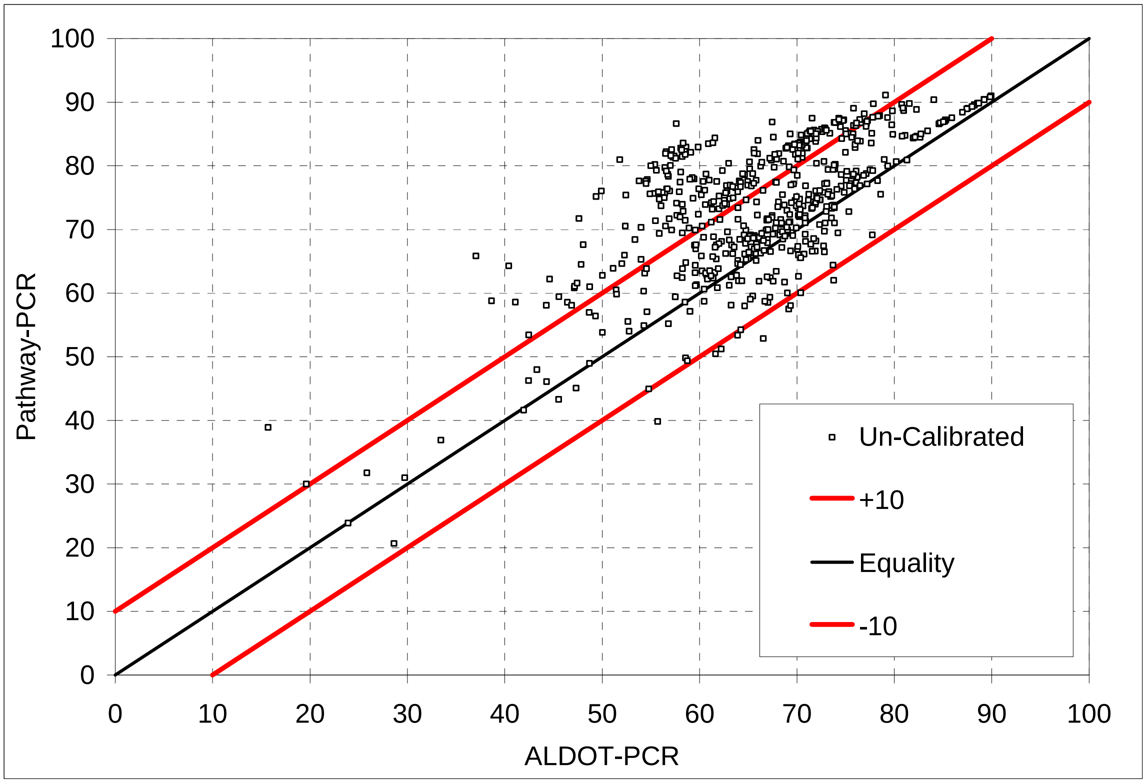

FIGURE 4.7 PCR Comparison Using Original PCR Equation.

Multivariable non-linear least squares regression was conducted in Excel to arrive at a set of recalibrated regression coefficients. The result was:

$$PCR = 95.5727 - 10.9994*(5.0\text{-}ROUGH) - 0.3948*LNALL1 - 0.8347*LNALL2 - 0.5027*LNALL3 - 0.2423*AVGOUT - 0.03032*LONG12 - .05484*LONG34 - 0.2916*TRAN12 - 0.69736*TRAN34 \quad (4.3)$$

Table 4.7 compares the regression coefficients from the original model (equation 4.1) to those of the recalibrated model (equation 4.3). Note that a number of coefficients remained unchanged between the two models. It was decided to fix the intercept coefficient to maintain the same starting score for both models before taking deductions for particular pavement distresses. This prevented achieving scores higher than reported with the original equation. The longitudinal cracking (LONG12 and LONG34) and high severity transverse cracking (TRAN34) were included as variable parameters in the regression process but were left unchanged by the regression analysis. Inspection of the coefficients indicates that the recalibrated equation puts a greater emphasis on pavement smoothness (ROUGH) with less (or the same) emphasis on the cracking terms. There was a slight change in the rutting term (AVGOUT). It is important to emphasize these changes are a function of the supplied ALDOT and Pathway data sets. Future data sets may result in different coefficients and trends. However, this procedure can be easily applied to newly developed data as it becomes available as demonstrated in the next section.

TABLE 4.7 Regression Coefficient Comparison

Coefficient	Original (equation 4.1)	Recalibrated (equation 4.2)
Intercept	95.5727	
ROUGH	-5.5085	-10.9994
ALL1	-1.5964	-0.3945
ALL2	-1.9629	-0.8347
ALL3	-2.9795	-0.5027
AVGOUT	-0.2220	-0.2423
LONG12	-0.03032	
LONG34	-0.05484	
TRAN12	-0.5305	-0.2916
TRAN34	-0.69736	

Ultimately, it is not the coefficients, but the quality of the PCR computations that is of vital importance. Figure 4.8 shows the result of the recalibration in the same format as the previous figure. After recalibration, 86% of the data were within ±10 PCR points and the data are more centered on the line of equality than the plot comparing ANN-generated data with the automated data. The data also do not suffer from the stacking seen in the ANN models (Figures 4.4 – 4.6).

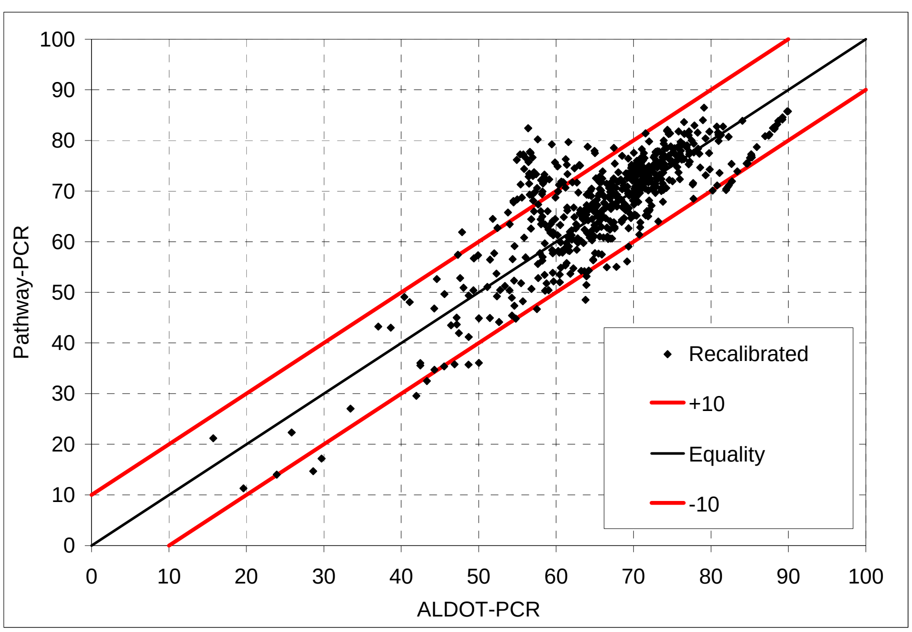

FIGURE 4.8 PCR Comparison Using Recalibrated PCR Equation.

5 FURTHER MODEL CALIBRATION

Pavement condition data collected in 2011 were made available to further calibrate the PCR equation. The data, provided by the ALDOT Pavement Management Division within the Bureau of Materials and Tests, were from ten quality control sites where ground-truth and Pathway measurements were conducted. Table 5.1 lists the 2011 test sites.

TABLE 5.1 2011 Test Sites

ROUTE	Starting Milepoint	Ending Milepoint	Paired Data Points
AL0006	181	181.3	30
AL0008	100	100.3	30
AL0008	165	165.3	30
AL0015	132	132.3	30
AL0015	156	156.3	30
AL0021	124	124.3	30
AL0053	104	104.3	30
AL0140	2	2.3	30
AL0185	19	19.3	30
AL0223	18	18.3	30
Total			300

Unlike the 2009/2010 data set, the 2011 data set included ALDOT-measured IRI and rut depths. This enabled a four phase evaluation to examine the following recalibration scenarios:

A. Replace ALDOT IRI & rutting data with Pathway data. This was the approach taken with the 2009/2010 data set.
B. No replacement of data.
C. Replace only ALDOT IRI data with Pathway data.
D. Replace only ALDOT rut data with Pathway data.

The data for each of the four scenarios was subjected to the same multivariable least-squares regression procedure to develop scenario-specific regression equations. Table 5.2 summarizes the regression terms while Figure 5.1 provides a direct comparison between PCRs and Figure 5.2 shows the residuals as a cumulative distribution. It should be noted that the recalibration resulted in a positive coefficient for the LONG12 variable for each scenario. This was likely an artifact of the regression process specific to this particular data set. Since this implies that the PCR increases with an increase in longitudinal cracking, it was decided to fix the coefficient for LONG12 to the original value (-0.0303). It is suggested that if positive coefficients are found in future recalibrations, they should be fixed to the original value and recalibration conducted again.

Regardless of the scenario (A, B, C or D), the results of further calibration were generally good. Using the ±10 PCR points as a practical limit, the worst case scenario was when no data were replaced resulting in 82.6% within the ±10 point range. As expected, the best case scenario (A) was where both the ALDOT and Pathway PCR were computed from the same rutting and IRI measurements which had 84.1% of the vendor-generated PCRs within ±10 of the ALDOT PCRs. This result was similar to the recalibration conducted with the 2009-2010 dataset.

TABLE 5.2 Regression Parameter and Accuracy Summary

	Scenario			
Coefficient	A-Replace IRI & Rut	B-No Replacement	C-Replace IRI Only	D-Replace Rutting Only
Intercept	95.5727			
ROUGH	-9.8339	-8.6376	-9.8358	-8.6357
ALL1	-0.5491	-0.7837	-0.6387	-0.6941
ALL2	-1.9321	-2.0792	-2.0171	-1.9943
ALL3	-1.9826	-1.9208	-2.0566	-1.8469
AVGOUT	-0.1678	-0.1569	-0.1218	-0.2029
LONG12	-0.0303 (*note – fixed to original value*)			
LONG34	-1.9042	-1.6201	-2.3476	-1.1767
TRAN12	-0.3668	-0.4259	-0.4007	-0.3920
TRAN34	-0.6974	-0.6974	-0.6974	-0.6974
%Within ±10	84.1	82.4	83.3	82.6
%Within ±5	63.0	59.3	62.2	59.6

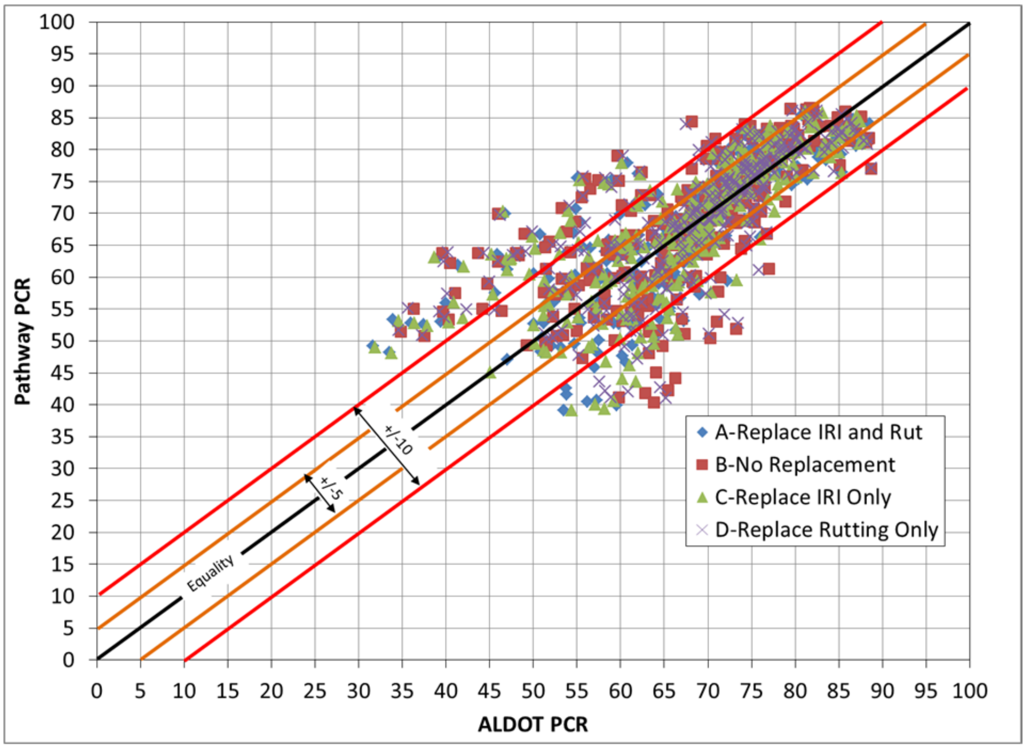

FIGURE 5.1 PCR Recalibration Using 2011 Data.

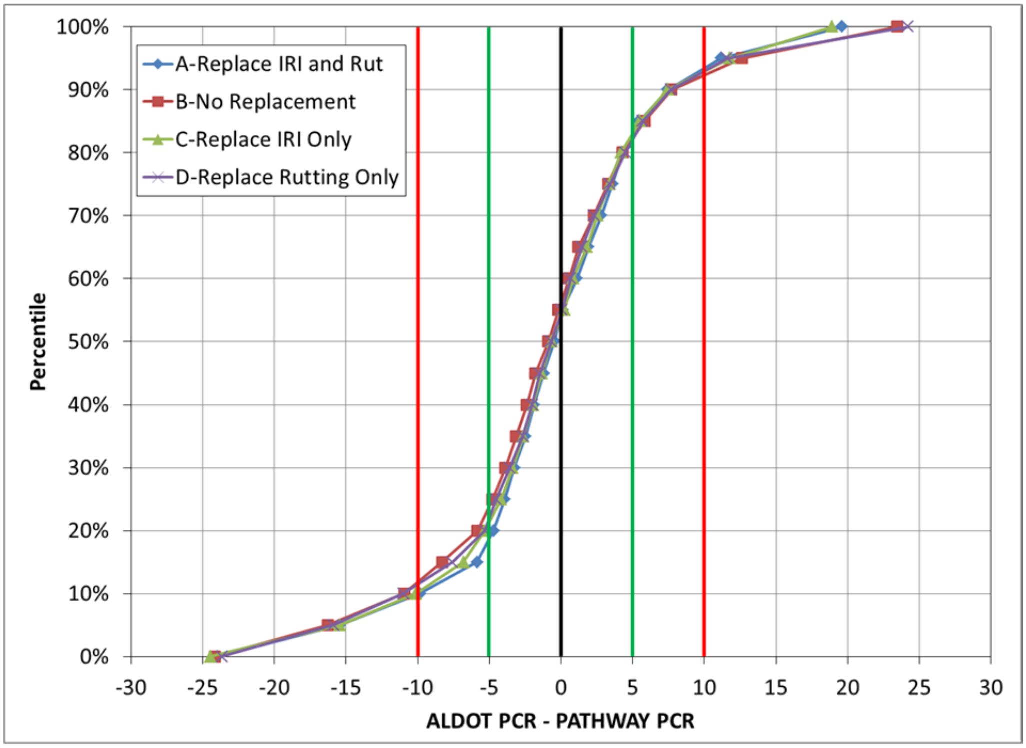

FIGURE 5.2 PCR Cumulative Residual Distribution.

Though Figures 5.1 and 5.2 do not show large practical differences between the data sets, regardless of the scenario, statistical testing was conducted to validate this conclusion. Two-tailed paired t-tests (α=0.05) were conducted for each scenario to evaluate the null hypothesis that the mean difference between ALDOT and Pathway PCR values was zero. Table 5.3 summarizes the results. The only scenario with statistically significant differences was B (No-replacement); though the null hypothesis was just barely rejected (as indicated by a p-value of 0.0355, as shown in Table 5.3). The others were all statistically-equivalent. Since past studies had shown generally good agreement between ALDOT and the vendor-generated IRI and rut depths, it is recommended that the vendor-provided IRI and rut depth data be used to compute the ground-truth PCR. This approach appears to provide the best correlation between ALDOT and vendor-generated PCRs by focusing primarily on differences between cracking measurements. However, it is critical that the replacement only be done after the agency has deemed the IRI and rut depth data sets from the vendor acceptable using ground-truth measurements.

TABLE 5.3 Paired t-Test Results

Parameter	Scenario			
	A.2-Replace IRI & Rut	B-No Replacement	C-Replace IRI Only	D-Replace Rutting Only
Observations	300	300	300	300
ALDOT Mean PCR	67.6	68.2	67.6	68.2
ALDOT PCR Variance	120.0	117.1	120.4	115.3
Pathway Mean PCR	68.3	69.2	68.4	69.0
Pathway PCR Variance	112.3	111.5	115.7	108.1
Pearson Correlation	0.78	0.71	0.77	0.73
p-value	0.1011	0.0355	0.0679	0.0500
Result	Accept Null	Reject Null	Accept Null	Accept Null

6 CONCLUSIONS AND RECOMMENDATIONS

The primary objective of this research was to develop a methodology for updating ALDOT pavement condition ratings to reflect past experience in pavement management using new means of distress data collection**.** Both artificial neural network modeling and recalibration of the original ALDOT PCR equation were conducted using data collected from 2009 through 2011. Based upon this investigation, the following conclusions and recommendations are made:

1. The ANN modeling proved unreliable using the 2009 data. Repeated training on the same data set yielded variable results that were not deemed acceptable. It is not recommended, at this point, to use ANN modeling for PCR prediction.
2. Recalibration of the original ALDOT PCR model, with a few parameters removed, yielded acceptable results using the 2009/2010 data. After recalibration, 86% of the vendor-computed PCR values were within ±10 points of the ALDOT-computed value.
3. Further calibration with more recently collected data, using presumably better data collection and analysis equipment and techniques, yielded similar results. 84.1% of the data were within ±10 points. For the purposes of prioritizing roadway segments for maintenance and rehabilitation, this was deemed sufficiently accurate. Furthermore, no statistically significant differences were detected between ALDOT and vendor-generated PCR scores when using common IRI or rut depth or both to compute the PCR value. It is recommended that the practice of using vendor-provided IRI and/or rut depth data continue once the vendor-provided data have been validated.
4. It is recommended that ALDOT continue with their ground-truth measurements. These measurements are critical to check assure the quality of the vendor data and provide the necessary data to conduct recalibration on an as-needed basis.

REFERENCES

1. BW Mining, http://bwmining.com/wp-content/uploads/2012/10/artificial_neural_network.png, accessed December 11, 2013.
2. Eldin, N. N. and Senouci, A. B. (1995). A pavement condition-rating model using backpropagation neural networks. *Microcomputers in Civil Engineering,* 10(6), pp. 433-441.
3. U.S. Federal Highway Administration (FHWA) (2003). *Distress Identification Manual for the Long-Term Pavement Performance Program*. Report No. FHWA-RD-03-031.
4. GDOT (1996). *Road Surface Management Pavement Condition Evaluation System*. Georgia Department of Transportation.
5. Holman, F.L. (1995). *Determining IRI and PSR for Use in Alabama*. Alabama Department of Transportation.
6. Leiva-Villacorta, F. (2012). *Advanced Computing Techniques in Structural Evaluation of Flexible Pavements Using the Falling Weight Deflectometer.* PhD Dissertation, Auburn University.
7. Loizos, A. and Karlaftis, A. (2006). Investigating pavement cracking: an evolutionary modular neural network approach. *Transportation Research Record*, *No. 1970*, pp. 207-214.
8. Roberts, C. A. and Attoh-Okine, N. O. (1998). A Comparative analysis of two artificial neural networks using pavement performance prediction. *Computer-Aided Civil and Infrastructure Engineering*, 13(5), pp. 339-348.
9. Saltan, M., Tigdemir, M. and Karasahin, M. (2002). Artificial neural network application for flexible pavement thickness modeling. *Turkish Journal of Engineering and Environmental Science*, 26, Tubitak, Turkey, 246-248.
10. Smith, R.E., Freeman, T.J., Pendelton, O.J. (1996). *Evaluation of Automated Pavement Distress Data Collection Procedures for Local Agency Pavement Management*. Report GC10470. Washington State Department of Transportation, Olympia, Washington.
11. Timm, D.H. and McQueen, J.M. (2004). *A Study of Manual vs. Automated Pavement Condition Surveys*. Auburn University Highway Research Center, Report No. IR-04-01.
12. Turner, D.S., T.C. Glover, J.V. Walters and E.R. Mansfield (1985). *A Pavement Rating Procedure*. TSM Report No. 112-85, BER Report No. 352-39.
13. WSDOT (1999). *Washington State Highway Pavement: Trends, Conditions and Strategic Plan*. Washington State Department of Transportation.