TRANSPORTATION RESEARCH

# RECORD

No. 1505

*Pavement Design, Management, and Performance*

# Pavement Monitoring and Evaluation Issues

*A peer-reviewed publication of the Transportation Research Board*

TRANSPORTATION RESEARCH BOARD
NATIONAL RESEARCH COUNCIL

NATIONAL ACADEMY PRESS
WASHINGTON, D.C. 1995

# Development of Performance Prediction Models for Illinois Continuously Reinforced Concrete Pavements

YING-HAUR LEE AND MICHAEL I. DARTER

A new predictive modeling approach is presented and the approach for localized failures in Illinois continuously reinforced concrete pavements (CRCP) is demonstrated. Some data retrieval guidelines from the Illinois Pavement Feedback System data base is first presented. A preliminary data analysis was conducted to assist in data cleaning and assessing the variability of the data before the analysis was performed. Several modern regression techniques ("robust" and "nonparametric" regressions) were introduced in a proposed new predictive modeling approach. The proposed modeling approach was used to develop an improved model for localized failures in CRCP. The resulting model includes several variables such as cumulative ESALs, slab thickness, content and methods of the steel reinforcement, and base type for the prediction of CRCP failures. A sensitivity analysis was also performed to illustrate the effect of these variables on failures. Slab thickness and steel content are by far the most significant variables affecting failures. Crack spacing had no effect.

Continuously reinforced concrete pavements (CRC pavements or CRCP) have been extensively constructed throughout the 1960s and 1970s in Illinois. Approximately 60 percent of the Illinois Interstate highways (the third largest mileage in the nation) was originally constructed as CRC pavements. The main incentive for constructing CRC pavements was the elimination of contraction joints to minimize joint-related distresses. The structural integrity of the concrete slab is maintained by allowing the pavement to crack randomly while providing reinforcement to hold the cracks tightly. The major distress types that occurred in CRCP are localized failures (including punchouts and steel ruptures) and major spalling of transverse cracks.

The causes and factors relating to localized failures in CRC pavements have been a topic among many investigators in past years (*1–3*). Various algorithms and numerical models have been developed in an attempt to describe the behavior of a CRC pavement under contraction restraints. The main focus points of these algorithms and models are the prediction of crack spacing, crack width, concrete stress, and steel stress due to environmental changes and external wheel load. The cracking behavior due to the percentage of longitudinal steel reinforcement, concrete strength, aggregate type, and other environmental factors has also been analyzed in a work by Zollinger (*4*).

Most of the maintenance activities on CRC pavements are directly related to localized failures (i.e., punchouts and steel ruptures.) It is often necessary to estimate these distress quantities for preventive design and pavement rehabilitation planning. In an attempt to relate the total number of failures to traffic loading, slab thickness, percent steel reinforcement, and subbase types, the first known predictive model based on actual CRCP performance data was developed using the Illinois CRCP data base compiled during the research project IHR-901 (*2*).

As part of the project Implementation of Pavement Feedback System, a study of the effect of different placement methods of reinforcing steel (i.e., tubes versus chairs) on the performance of CRCP in Illinois was conducted (*5*).

Y.-H. Lee, Department of Civil Engineering, Tamkang University, E725, #151, Ying-Chuan Rd., Tamsui, Taipei, Taiwan 25137, P.O.C. M. I. Darter, Department of Civil Engineering, University of Illinois, NCE L #12P, 205 N. Matthews Ave., Urbana, Ill. 61801.

## DATA PREPARATION

The Illinois Pavement Feedback System (IPFS) data base (1977–1991), containing the most complete source of pavement-related information for Illinois Interstate highways, is the main source of data used for this study. It contains detailed information about original and rehabilitation construction contracts, pavement inventory data, materials, historical traffic data, distress survey, condition rating surveys, and maintenance and rehabilitation records. The IPFS data base is currently implemented in the Illinois Department of Transportation (IDOT) mainframe system (VM "I" system) using the NOMAD2 data base management program. Automatic summary reports of the pavement information may easily be generated. For the purpose of this study, it was decided, however, to download all the summary section information, traffic history, distress records, and rehabilitation history to a PC and store in several PC-SAS datasets for further analysis. The PC-SAS SUMMARY, TRANSPOSE, and TABULATE procedures (*6*) were used heavily to summarize the information of interest and to provide more reliable data for this study.

### Design and Climatic Variables

The IPFS section summary data base includes codes for storing CRCP reinforcement data; however, few data on the type of steel reinforcement, diameter, spacing, and content are currently recorded in the data base. Fortunately, there exists some steel information, which was obtained from the IDOT district offices (*2*). This information was manually entered into the data base. In addition, IDOT's standard CRCP reinforcement designs (Standard 2225 and Standards 2225-1 to 2225-6) over the years were obtained for the rest of the CRC pavements for which no steel information was reported elsewhere. Generally speaking, the standard design of a given year was used to provide reinforcement data. The reinforcement type for these pavements was assumed to be deformed #5 and #6 bars for pavements constructed before and after 1981, respectively.

Different steel placement methods may affect the performance of CRCP pavements. Even though the IPFS data base includes a code for storing this information, very few data are recorded in the IPFS data base. Nevertheless, it is known that Illinois CRC pavements constructed before September 17, 1970, were mostly constructed using "chairs" with a few exceptions of "two-layer construction" for the steel placement. After this date pavements were constructed using "tubes" for placing reinforcement steels unless they were constructed in District 1, where chair placement was required. Therefore, this date was used as the cutoff point to determine whether the "chairs" or "tubes" placement method was used for a given CRC pavement.

Some additional construction data about drainage system, base type, and environmental data were also obtained. In addition, data on various subbase types, including granular, crushed stone, bituminous-aggregate mixture (BAM), and cement-aggregate mixture (CAM), were directly retrieved from the IPFS data base. Climatic data such as freezing index, average annual temperature, and average annual precipitation were also obtained.

### Traffic Calculation and Estimation

Traffic maps for average daily traffic, heavy commercial traffic, and multiple unit traffic volumes are published approximately every four years by the IDOT Office of Planning and Programming. The yearly traffic history recorded in the IPFS data base was determined by interpolating between those four-year periods.

Because the IPFS data base contains traffic information only up to 1987, it is necessary to estimate the traffic growth rate for each pavement section. A NOMAD2 program was written to perform automatically a huge array of regression analyses assuming that the traffic was increasing yearly with constant compound growth rates. With the estimated average daily traffic and ESAL traffic growth rates, the latest 1987 traffic data were then used as a starting point to predict the traffic into the future.

### Distress Quantities

The CRCP failures were recorded in various visible distress types, severities, amounts, and repairs in the distress data base. For example, a certain amount of medium-severity transverse cracking became high severity after a certain period of time; at the same time, some of the high-severity transverse cracking was corrected by full-depth repairs. To obtain a good single indicator of CRCP failures in each survey year, special efforts were conducted in deciding what distress types and severities should be included. After a considerable amount of effort and reexamination, the total number of CRCP failures (FAIL) per mile was defined as follows:

$$FAIL = 8.8\ PATCH + PUNCH + MHPOT + HTCRK \quad (1)$$

where

PATCH = all severities of permanent patches, percent area;
PUNCH = all severities of punchouts, #/mi;
HTCRK = high-severity transverse cracking, #/mi; and
MHPOT = medium- and high-severity of potholes and localized distresses, #/mi.

Because permanent patch deterioration was recorded in percent area of pavement surface, it was necessary to convert into number of patches per mile by assuming a $6 * 12\ ft^2$ per patch. In addition, high-severity transverse cracking was included and treated as an indication of steel rupture. Medium-severity transverse cracking was not included here because there were good chances that different surveyors rated low severity and medium severity inconsistently.

In this study, if the pavement was surveyed but without the aforementioned distress types recorded in the data base, the total number of failures was assigned to zero for later analysis. On the other hand, those unsurveyed pavement sections were excluded from the consideration, for example, pavements located in Chicago area (District 1) where detailed surveys were not possible because of the heavy traffic conditions.

### Additional Data from Old Vandalia Experimental Study

The longitudinal reinforcing steel content is known to have a very strong effect on the performance of CRCP pavements. Inadequate steel content often results in longer crack spacings, wider crack widths, and thus more punchout failures. To extend the range of analysis, several sections from the Vandalia CRCP experiment study (7) were also included in this study.

Eight sections of 7-in. and 8-in. experimental CRCP with 0.3, 0.5, 0.7, and 1.0 percent of longitudinal reinforcement constructed in 1947 to 1948 on US 40 west of Vandalia, Illinois, were studied over a 20-year service period (7). These pavements were placed directly on natural fine-grained soil and carried 4.27 million ESALs in 20 years. Many failures occurred in the 0.3 and 0.5 percent reinforcement sections in the 7-in. pavements.

Two major distress quantities, namely cracking (lin. $ft/1000\ ft^2$) and patching ($ft^2/1000\ ft^2$) were of particular interest (7) to quantify the equivalent total number of CRCP failures in these pavements. The reported cracking was defined as "cracks that are open or spalled at the surface to a width of 1/4in. or more for at least half the crack length, and sealed cracks" (7). Because most of them were still in a workable condition even after 20 years of service, only patching quantities were converted into total number of failures per mile for this study.

## PRELIMINARY DATA ANALYSIS

### Distress History and Additional Codes

A data cleaning process must be conducted before any regression analysis can be performed. With the help of graphical representation, failures were plotted against surveyed years for each section in the data base with additional information about route, direction, mileposts, D-cracking, slab thickness, and constructed year displayed. For example, a plot as shown in Figure 1 was used to examine the distress trends to eliminate possible data errors. The upper left-hand corner plot labeled '55-N-33.67-39.13-N-9"CRCP, conyr=75' indicated that a pavement located on I-55, northbound, mileposts 33.67 to 39.13, non-D-cracked, 9-in. CRCP, and constructed in 1975 was surveyed in 1985, 1987, and 1989 with approximate failures of 7, 6, and 15, respectively.

Each section was carefully examined. Two additional codes were assigned to each section to indicate the findings of the examination. The first code was used to indicate whether the total number of failures is reasonable according to the distress history. The second code supplements the first code to indicate which year of data to be

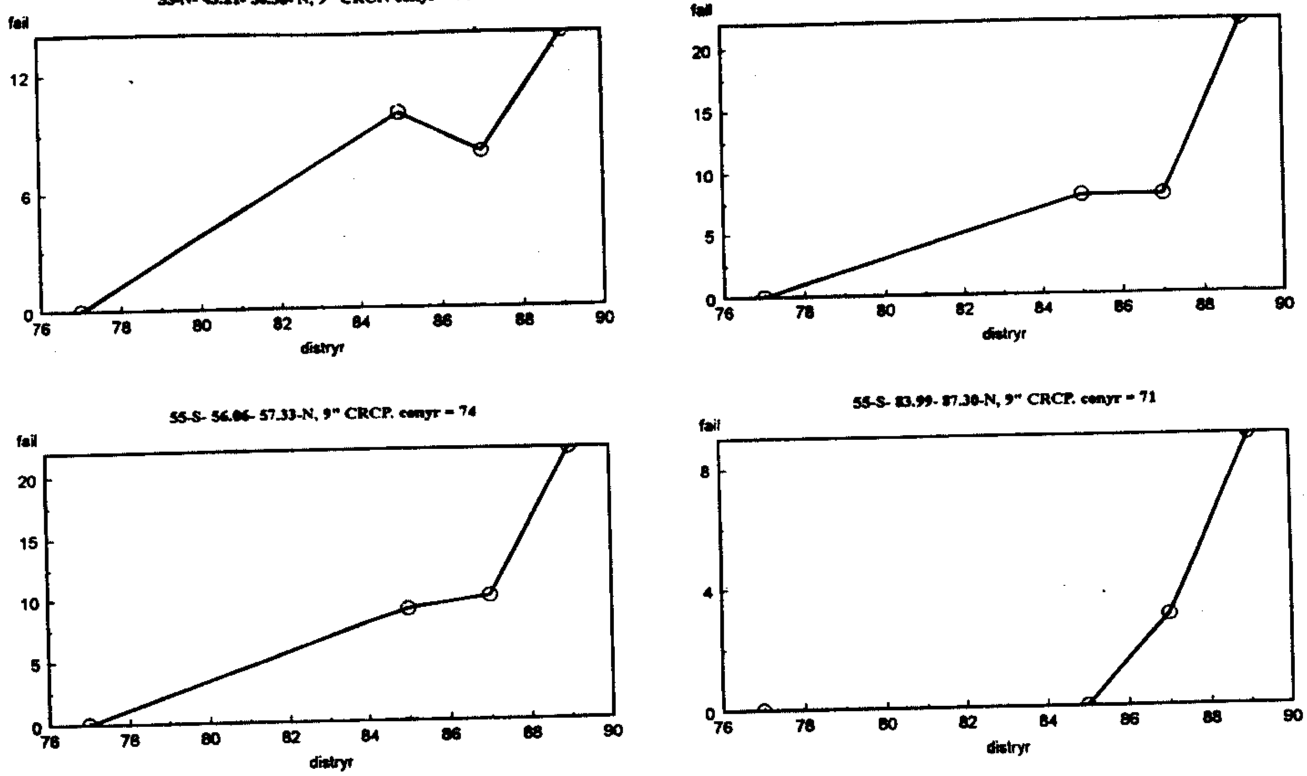


**FIGURE 1 Sample plots of distress history along I-55.**

deleted if necessary. Data correction was made in a way that could easily be traced back. A third code was also introduced to indicate the reliableness of the steel information. By doing so, different subsets of the final data base providing more reliable data might be analyzed for different purposes.

### Final Data Base with Non–D-Cracked CRC Pavements

D-cracking is a serious problem in Illinois Interstate concrete pavements. The presence of this type of distress causes the pavement to deteriorate prematurely. The number of sections with and without D-cracking based on different slab thicknesses was summarized after deleting those sections having unrealistic distress history, a particular suspicious data point, or unreliable calculated crack spacings. The results showed that approximately 60 percent of 7-in. IPFS sections were D-cracked whereas 40 percent for 8-in. and 20 percent for 9-in. pavements were D-cracked. No 10-in. CRCP pavements were recorded as D-cracked pavements. After carefully cross-examining the data, excluding all D-cracked pavements, and deleting seven IPFS data points with extremely large numbers of failures (greater than 100 failures per mile) a data base with 586 data points was finally created for later analysis.

### Correlation of Variables

A matrix plot containing the most important variables considered in this study are given in Figure 2. The variable correlations can be visually inspected through these plots. In addition, trimmed correlation matrixes showing the variable correlations after a certain portion of influential data points or possible outliers are eliminated were also obtained.

CRCP failures (indicated as "distr") were strongly correlated with age and cumulative ESAL (cesal) as expected even after 20 percent of data was trimmed. However, the correlations of slab thickness (pavthk), steel content (percent reinforcement, area of longitudinal reinforcement, and bonded area) to CRCP failures change dramatically. These are good indications of having influential data points in these variables, which should be used with extreme caution in later analysis.

The interrelationship between age and cumulative ESAL is also evident. The strong correlation between slab thickness and area of reinforcement almost guarantees the presence of one or the other in the later CRCP failure model development, but never both together to avoid strong collinearity in the model. This is also true for bonded area and percentage of reinforcement. Among these most important variables to CRCP failures, slab thickness, percent of reinforcement, and cumulative ESAL were chosen in the final CRCP failure model.

### Factor Space and Limitations

The final data base was mainly constructed from the in-service pavements, which satisfies certain design guidelines. Thicker pavements were designed to carry heavy traffic loadings. A relatively constant percentage of longitudinal reinforcement was often used.

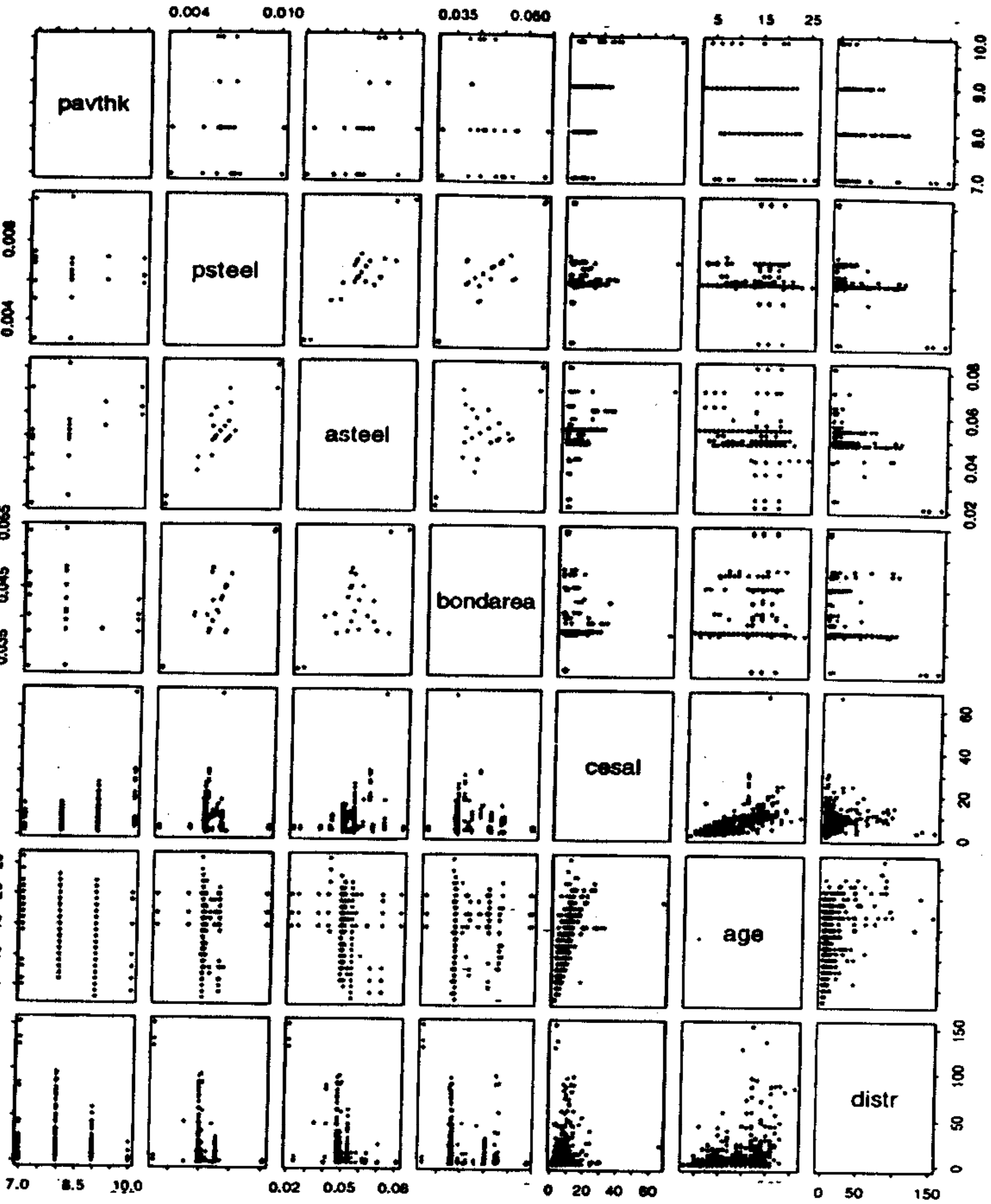


**FIGURE 2** Matrix plot with most important variables to CRCP failures (distr = failures/mi).

For example, there was limited range of reinforcement content in the data except for the Vandalia experimental sections. Vandalia experimental sections provided some very low and very high steel contents for 7- and 8-in. pavements. However, there existed no such steel contents for 9- and 10-in. in-service pavements. Most of the final data base was 8- and 9-in. pavements, which had a fixed reinforcement content of approximately 0.60 to 0.62 percent.

In addition, the Vandalia experimental data were obtained under very low traffic (i.e., up to 4.3 million ESALs in 20 years). There existed no very low or very high steel contents under very high traffic loading conditions. Knowing that the deficiencies of the unbalanced factor space of the final data base exist, extreme care should be used during the modeling process as well as deriving conclusions beyond the range of data.

### Variability of Data

Because most of the final data base was 8-and 9-in. pavements with constant reinforcement contents of approximately 0.60 to 0.62 percent, a preliminary examination of the relationships among CRCP failures, age, and cumulative ESALs was performed. Under similar design conditions (BAM subbase, drainage system, and "tubes" reinforcement placement method), the performance of these two groups of 8- and 9-in. pavements showed high variation.

This high variation might be the result of some other hidden variables not considered in this study. However, it is also believed that the current practice of sampling only 10 percent of the entire project may be the major source of these variations. For example, three identical 1-mi pavement sections when surveyed in their first 0.1 mi and recorded to have 0, 1, and 2 failures might end up with 0, 5.3, and 10.5 failures per mile.

It is recommended that a full-length survey might be the best solution to minimize this high variability in the total number of failures counted. If this is not possible, a much higher portion of the project such as 20 to 30 percent should be surveyed instead so that the accuracy of the IPFS data base can be improved.

## PROPOSED NEW PREDICTIVE MODELING APPROACH

The proper selection of regression techniques is one of the most important factors to the success of prediction modeling. Traditional "parametric" regression techniques such as linear and nonlinear regressions require imposing a parametric form on the functions and then obtaining the parameter estimates. With the multi-dimensional pavement engineering problems in mind, several unresolved deficiencies were frequently identified in the use of traditional stepwise regression and nonlinear regression. These include problems in the selection of correct functional form, large influence of potential outliers, violations of the embedded statistical assumptions, and failure to satisfy some engineering boundary conditions.

Because of the innovation of computers and the almost unlimited computing power, several ingenious iterative regression techniques in the area of "robust" and "nonparametric" regressions have been developed in the past 10 years and have gradually gained popularity. They are useful especially in situations in which large data contamination and little knowledge about the shape and the form of a function exist. For this study, particular attention is focused on the following advanced modern regression techniques (*8*):

1. Least median squared (LMS or "Robust") regression (*9,10*): a robust regression technique, extremely powerful in detecting outliers in either response variable or predictor variables;
2. Alternating conditional expectations (ACE or "Expectation") (*11*): a nonparametric regression technique, providing optional variable transformations to maximize the squared multiple correlation ($R^2$); and
3. Additivity and variance stabilization (AVAS or "Stabilization") (*12*): a nonparametric regression technique, transforming both sides of the additive model to achieve constant error variance assumption.

Without imposing an unjustified parametric assumption, nonparametric regression techniques strive to estimate the actual functional form that best fits the data through the use of scatter plot smoothers. They can be excellent supplements to traditional parametric regression techniques, especially in suggesting proper transformations of the response variable and the predictor variables to help uncover the underlying relationships and to satisfy some applicable boundary conditions.

A new statistical package named S-PLUS, which has been widely used by statisticians for data analysis, was selected because of the availability of these new regression techniques.

### Multiple Linear Regression

Multiple linear regression is one of the most time-honored and widely used regression techniques for the study of linear relationships among a group of measurable variables. Suppose there exists a true model to describe the relationship between response variables ($y$s) and explanatory variables (or predictors, $x$s) (*13*):

$$y_i = x_i^T\beta + \epsilon_i \qquad i = 1, \ldots, n \tag{2}$$

where $x_i^T$ is the $i$th row of the ($n \times p$) matrix $X$ of the column of 1s if including an intercept and the explanatory variables. The superscript $T$ denoting the transpose of the column vector $x_i$ is required because of the usual convention that all vectors are represented by column vectors. $\beta$ is a ($p \times 1$) vector of unknown regression coefficients and $p$ and $n$ are the number of parameter estimates in the model and the total number of observations, respectively.

The basic assumptions are usually that the random errors ($\epsilon$s) are mutually uncorrelated and normally distributed with zero mean and constant variance, and additive and independent of the expectation function. For any arbitrary $\beta$ value of $\hat{\beta}$, the residuals $r_i(\hat{\beta})$ can be determined by the following expression:

$$r_i(\hat{\beta}) = y_i - x_i^T\hat{\beta} \qquad i = 1, \ldots, n \tag{3}$$

Based on those assumptions, multiple regression tries to find a set of parameters $\beta$ such that the sum of the squared residuals given in Equation 4 is minimized, which is also best known as the least squares (LS) method.

$$RSS(\hat{\beta}) = \sum_{i=1}^{n} [r_i^2(\beta)] = r_1^2(\hat{\beta}) + r_2^2(\hat{\beta}) + \ldots + r_n^2(\hat{\beta}) \tag{4}$$

### Nonlinear Regression

Practical real-world problems are often found to be nonlinear in nature. Because of its favorable feature of handling a complicated nonlinear model, nonlinear regression has been widely used as a modeling technique. However, nonlinear models are more difficult to specify and develop than linear regression models. "Some models are difficult to fit, and there is no guarantee that the procedure will be able to fit the model successfully" (*14*).

Suppose there exists a true model that best describes the relationship between response variables ($y$s) and explanatory variables ($x$s), (*14,15*):

$$y_i = F(\beta, x_i) + \epsilon_i \qquad i = 1, \ldots, n \tag{5}$$

where

$F(\beta, x_i)$ = nonlinear function based on the predictors,
$\beta$ = ($p \times 1$) vector of unknown regression coefficients to be estimated,
and $n$ = total number of observations.

Similar to linear regressions, the disturbance (or error) term is usually assumed to be additive, mutually uncorrelated, and normally distributed with zero mean and constant variance. For any arbitrary $\beta$ value of $\beta$, the residuals $r_i(\beta)$ are

$$r_i(\hat{\beta}) = y_i - F(\hat{\beta}, x_i) \qquad i = 1, \ldots, n \tag{6}$$

Unlike linear regressions whose parameters can be explicitly estimated by a closed-form expression, nonlinear regressions must use an iterative routine to find the best parameter estimates ($\hat{\beta}$) such that the sum of the squared residuals as given in Equation 4 is minimized.

### LMS or "Robust" Regression

Recently, new robust statistical techniques have been developed to avoid the large influence of outliers. The results of these methods are still trustworthy even if a large amount of data is contaminated. They are extremely useful in identifying a group of potential outliers in a single attempt. Of many robust regression techniques, the least median of squares estimator introduced by Rousseeuw (*9*) and Rousseeuw and Leroy (*10*) is the most robust with respect to outliers in the dependent variable as well as outliers in the independent variables or predictors.

It is assumed that the true model and the residuals are the same as those given in Equations 1 and 2, respectively. The LMS estimator ($\hat{\beta}$) attempts to minimize the median instead of the sum of the squared residuals defined as follows:

$$RMS(\hat{\beta}) = \underset{i=1}{\overset{n}{\text{med}}} \{r_i^2(\hat{\beta})\} = \text{med}\,[r_1^2(\hat{\beta}), r_2^2(\hat{\beta}), \ldots, r_n^2(\hat{\beta})] \tag{7}$$

As Rousseeuw stated, LMS regression first tries to fit most of the data and then discover the potential outliers. The LMS method has a breakdown point of 50 percent, which is the largest possible value, whereas the LS method has a breakdown point of 0 percent. The breakdown point of a regression estimate is defined as the largest fraction of data that may be replaced by any arbitrary values without causing arbitrary parameter estimates. This means that the LMS estimates still continue to follow the trend of most of the data even when almost half is arbitrarily corrupted. Geometrically, it corresponds to finding the narrowest band that covers at least half of the observations such that potential outliers are discovered (*10*).

The advantages of LMS regression are obvious especially when analyzing field-collected pavement data that may contain as much as 10 to 20 percent contaminated data. "Robust" regression provides a more objective way to help identify some potential data and model problems. These problems include actual data errors, data from a different population, and the inadequacy of the regression model due to missing some important variables.

Thus, once these trouble data points are identified, more detailed investigations should be conducted to find out why and how they are different from the other tentative good ones. Some trouble data points may possibly be identified as errors and subsequently be corrected or excluded from the analysis. The possibility to include other explanatory variables or other model forms in the model should also be fully investigated to improve the fit including all the data points. It should be emphasized, however, that no data should be deleted without having justifiable reasons to do so. By conducting these analyses in an iterative manner, it is strongly believed that more reliable predictive models may be developed.

### ACE or "Expectation" Algorithm

An algorithm to find the optimal transformations of the response variable and the predictor variables such that the $R$-square of a multiple regression is maximized was introduceful. Suppose that there exists a true additive model given as follows:

$$\theta(y) = \phi_1(x_1) + \phi_2(x_2) + \ldots + \phi_p(x_p) + \varepsilon \tag{8}$$

where

$y$ = response variable,
$x_1, x_2, \ldots, x_p$ = predictor variables,
$\theta(y), \phi_1(x_1), \phi_2(x_2), \ldots, \phi_p(x_p)$ = unknown nonparametric transformation functions of each variable to be estimated, and
$\varepsilon$ = random error.

The proposed algorithm uses a sophisticated supersmoother while it alternatively changes the conditional expectation functions to minimize the fraction of variance ($e^2$) not explained by regressing $\theta(y)$ on $\phi_1(x_1), \ldots, \phi_p(x_p)$. This algorithm is often called alternating conditional expectations algorithm. The $e^2$ given in the following expression is also called the goodness-of-fit measure (*8*).

$$e^2(\theta, \phi_1, \ldots, \phi_p) = \frac{E\left[\theta(y) - \sum_{i=1}^{p} \phi_i(x_i)\right]^2}{E\theta^2(y)} \tag{9}$$

This "nonparametric" algorithm will only give back data-dependent estimates of variable transformations that are not restricted to any particular functional form. However, the data analyst might be able to estimate particular parametric transformation for each variable by plotting the suggested transformed variables versus the original ones. The traditional Box-Cox transformation technique can be used for this purpose. If the suggested transformations are so desirable that a single family of power transformation is not adequate, polynomial regression, nonlinear regression, or any other curve fitting techniques can also be applied separately for the transformation of each variable. Thus, the $R$-square of the final additive model is optimally maximized.

The "Expectation" algorithm provides a fully automated routine to assist in selecting the optimal form of transformations for response and predictor variables. However, it should be used with caution, especially in the presence of outliers. Furthermore, because a smoothing technique often requires a certain number of degrees of freedom, the transformed vectors might be highly unstable if the number of observations is not large enough. The ACE algorithm may still produce strong looking transformations when there is little or no relationship between the predictors and the response variable. This problem may be detected by the resulting relatively small $R^2$. More detailed discussions of its applications and limitations are given elsewhere, (*8,11,16*).

### AVAS or "Stabilization" Algorithm

In 1987, Tibshirani (*12*) successfully introduced the additivity and variance stabilization algorithm by applying the same alternatively backfitting techniques used in the Expectation algorithm. In contrast to the ACE algorithm that tries to maximize the squared multiple correlation ($R^2$), the Stabilization algorithm strives to achieve the constant error variance assumption of regression and also improve the model fit.

Assuming that there exists an additive model of the same form given in Equation 8, the AVAS algorithm tries to achieve the following two goals simultaneously (*8*):

$$E[\theta(y)\,|\,x_1, x_2, \ldots, x_p] = \sum_{i=1}^{p} \phi_i(x_i), \tag{10}$$

$$\mathrm{VAR}[\theta(y)\,|\,\sum_{i=1}^{p} \phi_i(x_i)] = \text{Constant} \tag{11}$$

where $E[z \mid w]$ and $\mathrm{VAR}[z \mid w]$ stand for the conditional expectation function and the conditional variance function of $z$ given $w$, respectively.

The AVAS transformation is more flexible than the traditional Box-Cox method and is often better suited to regression problems than the ACE algorithm (*8,12*). Because the objective functions that they attempt to optimize are different, the ACE and the AVAS transformations behave differently for different situations. It is too early to say which one is better. In general, some evidence shows that the AVAS algorithm is better behaved than the ACE algorithm when there is little or no relationship between the predictors and the response variable.

## Proposed New Predictive Modeling Approach

For practical engineering problems, often little knowledge about the true functional form is available and the data collected are also confounded with substantial errors. So far there exists no regression algorithm that can perform outlier detections and variable transformations simultaneously to minimize these problems. To develop a more reliable predictive model for such complicated problems, it is proposed to incorporate the Robust regression, Expectation, and Stabilization algorithms into the modeling process.

The Robust regression is proposed because of its favorable feature of analyzing highly contaminated data by detecting outliers from dependent and independent variables. Through the iterative use of the combination of these outlier detection and nonparametric transformation techniques, it is believed that some potential outliers and proper functional forms may be identified. Subsequently, traditional regression techniques can be used more easily to develop the final predictive model.

The basic procedures and concepts behind the proposed modeling approach are briefly discussed. First, assume a plausible linear model relating the response variable to the explanatory variables. Then, apply Robust regression to delete some potential outliers based on the assumed model form. Subsequently, apply the Exception and Stabilization algorithms to find possible variable transformations that best fit the remaining data. The transformed vectors are then plotted against each original variable. In addition, a plot of the predicted versus the actual values and a plot of the residual versus the predicted values are also provided.

Through visual inspections, the reasonableness of the suggested transformations and the goodness of the fit can be easily accessed. Because the suggested transformations for each variable are displayed in two-dimensional plots, they can be properly formulated using traditional Box-Cox transformation, linear (or polynomial) regression, and nonlinear regression. These tasks are relatively easy, because they involve only one variable at a time.

Then, revise the assumed linear model using the suggested transformations and repeat the entire process until the detected outliers and the suggested transformations are acceptable. Finally, traditional linear regression is used to get the final regression statistics and diagnostics of the additive model using the transformed variables. Step-by-step procedures for the proposed new predictive modeling approach are summarized in Figure 3.

The potential outliers detected by Robust regression are temporarily excluded from the subsequent Expectation and Stabilization trials to minimize the influence of possible data errors. However, these data points should be added back to the original data base when analyzing the next trial linear model form. This is because the potential outliers detected by Robust regression may be affected by the assumed trial linear model form, but they may not be actually bad data points. The resulting transformations suggested by the Expectation and Stabilization algorithms may also be affected by excluding those data points. During each iteration, however, some erroneous data may be identified and subsequently eliminated form the analysis. These procedures can be routinely performed until an acceptable model is obtained.

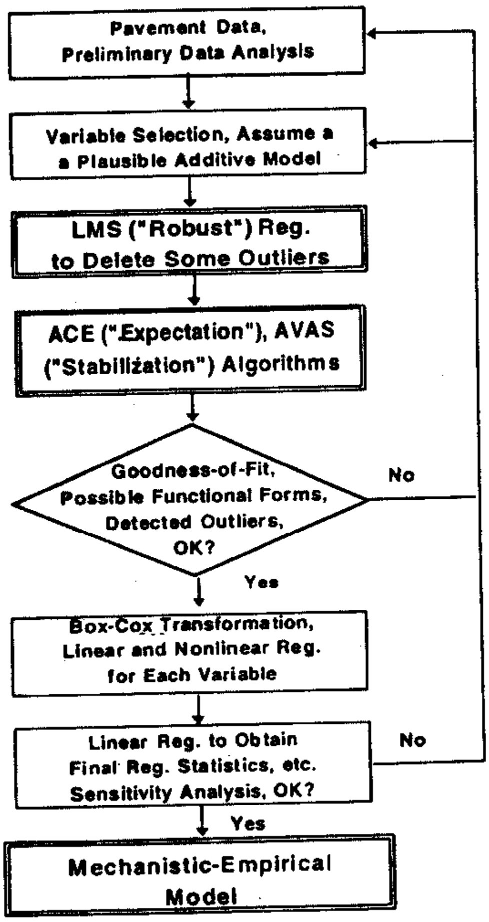


**FIGURE 3 Proposed modeling procedures for outlier detection and selection of proper functional forms.**

## PROPOSED CRCP PERFORMANCE PREDICTION MODELS

As expected, developing a predictive model to adequately fit this type of data is not an easy task. Some preliminary trials using the traditional linear and nonlinear regression techniques have had difficulty in achieving a satisfactory model for the data.

The proposed new predictive modeling approach as described in the previous section was followed closely and used routinely to develop an improved model. This approach proved to have substantial improvement over the use of traditional regression techniques alone in an attempt to uncover the underlying relationships together with the consideration of the possible influence of outliers.

The results of many LMS regression trials indicate that a large portion of the zero failures in the data is questionable in spite of different model forms analyzed. Most of these zero failures were forced into the final data base because of some evidence showing that the given pavement section was surveyed but did not have any of these failures recorded in the original IPFS data base. By excluding these forced-in data points, a better model with more reasonable predictions could be developed, although very high variations were still present.

Several dozen predictive models using different model forms were developed with similar prediction accuracy. The final proposed model for predicting the number of CRCP failures on a per mile basis is given as follows:

$$\begin{aligned}\log_e(\text{FAIL}) = {} & 6.8004 - 0.0334 * \text{PAVTHK}^2 - 6.5858 * \text{PSTEEL} \\ & + 1.2875 * \log_e(\text{CESAL}) - 1.1408 * \text{BAM} \\ & - 0.9367 * \text{CAM} - 0.8908 * \text{GRAN} \\ & - 0.1258 * \text{CHAIRS} \qquad \text{statistics: } R_2 = 0.44, \\ & \text{SEE} = 1.06, N = 408 \end{aligned} \tag{12}$$

where

FAIL = total number of failures in outer lane, #/mil;
THICK = CRCP slab thickness, in.;
PSTEEL = longitudinal reinforcement, percent;
CESAL = cumulative ESALs, millions;
BAM = 1 if subbase material is bituminous-aggregate mixture, 0 otherwise;
CAM = 1 if subbase material is cement-aggregate mixture, 0 otherwise;
GRAN = 1 if subbase material is granular, 0 otherwise; and
CHAIRS = 1 if chairs used for reinforcement placement, 0 if tubes used.

The regression summary outputs and the goodness-of-fit of the proposed model were presented elsewhere (*5*). Notice that a small number of 0.1 is added to the actual total failures ("distr") to avoid numerical difficulties. This model also satisfies the boundary condition of resulting zero failures when no traffic exists.

Some plots showing the sensitivity of the various factors in the proposed model are presented in Figures 4 and 5. Figure 4 shows the relationships among cumulative ESALs, slab thickness, reinforcement content, and total number of failures per mile (fit). Figure 5 shows the effects of reinforcement placements and different base types. The general trends of the effects appear to be reasonable. Note that the plots are extended a bit beyond the range of the actual data to show how the model performs. The reinforcement content of a given CRC pavement has a large effect on the occurrence of failures.

The proposed model also includes the type of reinforcement placement (CHAIRS). The use of chairs results in fewer total failures; however, the difference is not significant. Even though the analysis does not provide a lot of support for the placement method

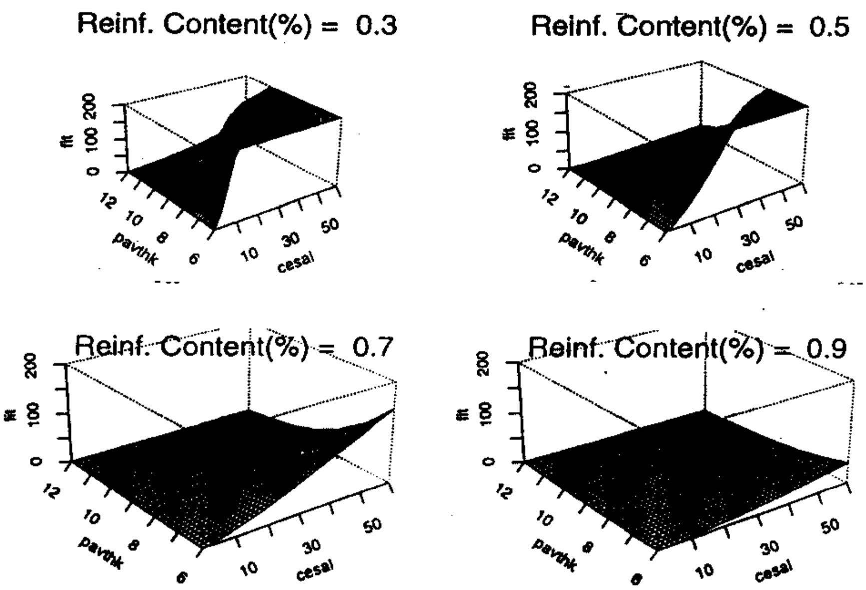


**FIGURE 4 Three-dimensional sensitivity analysis for reinforcement contents (plot truncated at 200 failures/mi).**

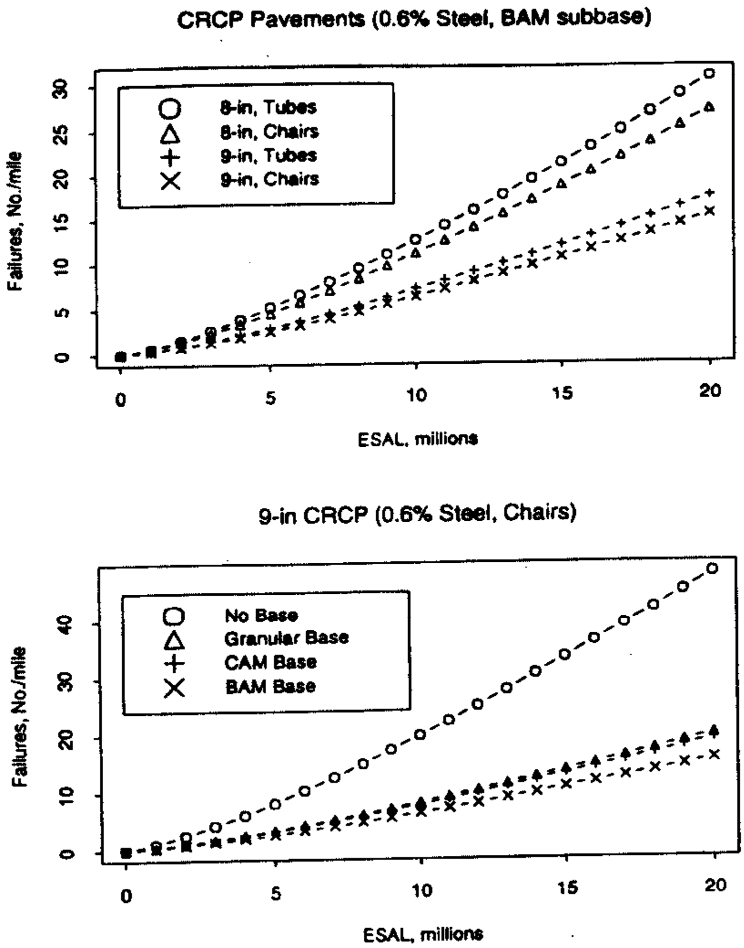


**FIGURE 5 Two-dimensional sensitivity analysis; reinforcement placements and base types.**

having a significant effect on the development of CRCP failures, this does not necessarily prove that this is the case. As previously discussed, the data base does not provide a clean separation of two identical groups of CRC pavements (i.e., one with tube placement and the other with chair placement). Many factors may be interacting with thickness, traffic loading, or other factors to cause the true effect of placement type to be hidden.

Some research performed in Illinois on I-70 has clearly shown that the depth of reinforcing steel greatly affects the crack width and thus the breakdown of cracks and development of CRCP failures. Thus, if the tube placement method results in a greater variation in depth of steel, there would likely be a greater chance for more failures. This may be the case even though the data did not clearly show this to be true.

The effects of different base materials were also investigated. A CRC pavement with BAM base has the best overall performance, which also agrees with previous findings (*17*). On the other hand, a conclusion different from the previous literature (*17*) is derived from this analysis that CAM base has about the same effect on the development of CRCP failures as granular base.

## SUMMARY AND CONCLUSIONS

A study of the factors affecting the performance of CRCP was conducted using the in-service IPFS data base. Detailed guidelines for data preparation are provided. The entire performance records of bare CRC pavements in Illinois were retrieved. In addition, some of the old Vandalia experimental sections were included in this study because they provided additional ranges in steel content (0.3 to 1.0 percent). The data were cleaned carefully to remove sections that had D-cracking, questionable data (high failures), very short sections, and so forth because these would only increase the potential errors.

A preliminary data analysis was conducted to assist in data cleaning, assess the variability of the data, and understand the interrelationships between variables before actually performing the regression analysis. Very high variations of the data are evident, suggesting that the current practice of surveying 10 percent of the entire pavement network may be inadequate.

Several modern regression techniques (robust and nonparametric regressions) were introduced in a proposed new predictive modeling approach with detailed step-by-step guidelines. The proposed modeling approach was routinely used to derive a more reliable predictive model. The resulting model includes several variables such as cumulative ESALs, slab thickness, content and methods of the steel reinforcement, and base type for the prediction of CRCP failures. A sensitivity analysis was also performed to illustrate the effect of various factors in the model, which also appeared to be reasonable.

## ACKNOWLEDGMENTS

This research was originally sponsored by IDOT. The authors are grateful to David Wheat, David Lippert, and James Hall for their assistance. The support obtained from Tamkang University and the University of Illinois to finish this research work is also greatly acknowledged.

## REFERENCES

1. Ma, J., and B. F. McCullough. *NCHRP Report 177-9: CRCP-2, An Improved Computer Program for the Analysis of Continuously Reinforced Concrete Pavements.* TRB, National Research Council, Washington, D.C., Aug. 1977.
2. LaCoursiere, S. A., M. I. Darter, and S. A. Smiley. *Performance of Continuously Reinforced Concrete Pavement in Illinois.* Illinois Cooperative Research Report 901-1, University of Illinois, Urbana, Dec. 1978.
3. Zollinger, D. G., and E. J. Barenberg. *Thickness Design Procedure for Continuously Reinforced Concrete Pavements.* Illinois Cooperative Highway Research Report IHR-518-3, University of Illinois, Urbana, May 1988.
4. Zollinger, D. G. *Investigation of Punchout Distress of Continuously Reinforced Concrete Pavement.* Ph.D. Thesis, University of Illinois, Urbana, 1989.
5. Lee, Y. H. *Development of Pavement Prediction Models.* Ph.D. Thesis, University of Illinois, Urbana, 1993.
6. *SAS User's Guide: Basics.* Ver. 5 Edition, SAS Institute, Inc, Cary, N.C; 1985.
7. Burke, J. E., and J. S. Dhamrait. A Twenty-Year Report on the Illinois Continuously Reinforced Pavement. On *Highway Research Record 239,* HRB, National Research Council, Washington D.C., Highway 1968, pp. 197–211.
8. *S-PLUS for Windows: User's Manual.* Vol. 1 and Vol. 2, Statistical Sciences, Inc, Seattle, Washington, March 1993.
9. Rousseeuw, P. J. Least Median of Squares Regression. *Journal of the American Statistical Association,* Vol. 79, 1984, pp. 871–880.
10. Rousseeuw, P. J., and A. M. Leroy. *Robust Regression and Outlier Detection.* Wiley Series in Probability and Mathematical Statistics, John Wiley & Sons, Inc., New York, 1987.
11. Breiman, L., and J. H. Friedman. Estimating Optimal Transformations For Multiple Regression And Correlation (With Discussion). *Journal of the American Statistical Association,* Vol. 80, 1985, pp. 580–619.
12. Tibshirani, R. Estimating Transformations for Regression via Additivity and Variance Stabilization. *Journal of the American Statistical Association,* Vol. 83, 1987, pp. 394–405.
13. Weisberg, S. *Applied Linear Regression.* Second Edition, Wiley Series in Probability and Mathematical Statistics, John Wiley & Sons, Inc., 1985.
14. *SAS User's Guide: Statistics.* Ver. 5 Edition, SAS Institute, Inc., Cary, N.C., 1985.
15. Bates, D. M., and D. G. Watts. *Nonlinear Regression Analysis and Its Applications.* Wiley Series in Probability and Mathematical Statistics, John Wiley & Sons, Inc., New York, 1988.
16. Buja, A. Remarks on Functional Canonical Variates, Alternating Least Squares Methods and ACE. *The Annals of Statistics,* Vol. 18, No. 3, 1990, pp. 1032–1069.
17. Heinrichs, K. W., M. J. Liu, M. I. Darter, S. H. Carpenter, and A. M. Ioannides. *Rigid Pavement Analysis and Design.* Federal Highway Administration Contract DTFH-61-85-C-00103, University of Illinois, Dec. 1988.

*Publication of this paper sponsored by Committee on Pavement Monitoring Evaluation, and Data Storage.*