New Techniques for Modeling Pavement Deterioration

MOHAMED Y. SHAHIN, MARGARITA M. NUNEZ, MARGARET R. BROTEN, SAMUEL H. CARPENTER, AND AHMED SAMEH

The ability to model and predict pavement condition accurately is critical to the success of pavement management systems. This paper evaluates the applicability of three mathematical curve-fitting techniques for modeling pavement condition deterioration behavior. The mathematical models investigated are: stepwise regression, *B*-spline approximation, and constrained least-squares estimation. The best features of each are integrated into an interactive format capable of operating within the PAVER pavement management system. Pavement sections from a given location consisting of the same pavement type, use, and rank are grouped into families. Models that filter obvious errors and statistical outliers from the data are applied to the family data. Both the *B*-spline approximation and constrained least-squares techniques are used in the screening process. A constrained least-squares curve-fitting technique is used to fit a family prediction model curve to the filtered data. Pavement condition prediction at the section level is accomplished using the position of a section relative to the family prediction model curve. Extrapolation of pavement family condition values beyond the maximum age available in a family data base is accomplished using a backtracking method. These methods were determined to give the most consistent and believable results from among a number of possible methods considered. These procedures constitute a complete method to model and predict pavement family behavior and pavement section behavior accurately. The method is being integrated into the PAVER pavement management system to improve the pavement evaluation process.

The ability to predict pavement deterioration accurately is critical to the success of any pavement management system. A successful pavement condition prediction technique provides a fundamental tool to aid in the planning and cost allocation of maintenance and rehabilitation activities. Detailed in this paper are the results of a study undertaken to develop a reliable pavement condition modeling technique.

The prediction methodology presented herein was developed for incorporation into the PAVER pavement management system (PMS) *(1)*. However, other management systems that use historical condition data can also effectively use this prediction technique. PAVER was designed to optimize fund allocation for pavement maintenance and repair. Currently, PAVER employs a straight-line extrapolation applied to each individual pavement section based on the results of previous condition surveys to model pavement condition. However, when pavement condition index (PCI) versus age is plotted, a curvilinear relationship with monotonically decreasing PCI is expected. Thus, a need clearly exists for a more realistic pavement condition modeling technique. The methodology presented in this paper is based on several years of research to improve the current pavement condition modeling and prediction procedures in PAVER.

M. Y. Shahin, M. M. Nunez, and M. R. Broten, U.S. Army Construction Engineering Research Laboratory, P. O. Box 4005, Champaign, Ill. 61820-1305. S. H. Carpenter, University of Illinois, 1203 NCEL., 208 North Romine St., Urbana, Ill. 61801. A. Sameh, University of Illinois, 306c Talbot Laboratory, Urbana, Ill. 61801.

SUMMARY OF PREVIOUS RESEARCH

Most PAVER data bases currently contain information pertaining to a one-time survey for each section during the lifetime of the pavement. Therefore, a critical need exists for a model to predict pavement behavior when good historical condition data for a section are unavailable. Several studies have been undertaken in the past to satisfy the need for an accurate method of pavement condition prediction.

Initially, stepwise regression models were used to predict pavement condition index (PCI). Several variables were incorporated into these models, including pavement type, condition rating, nondestructive testing (NDT) information, pavement construction, traffic information, pavement age, and pavement layer thicknesses. These models were evaluated and determined unsatisfactory for predicting PCI at the project level *(2)*. The resultant R^2 was very low and the residual standard error was high when stepwise regression was used to model these data. This was partially attributed to the large number of estimating errors associated with the models. Another drawback to these complex models is that they are difficult to update to reflect additional condition data as they are collected.

In an attempt to address these problems, Nunez and Shahin *(3)* took another approach to modeling pavement behavior. To combat the problem of insufficient section-specific condition history data, a family approach was developed. Pavement sections from a given PAVER data base that had the same pavement type, pavement use, and pavement rank were grouped into families. In this approach, sections with different ages and condition ratings are assumed to represent the deterioration in condition of a typical family section over time. Thus, by collecting each section age and condition, placing them chronologically, and fitting a curve through the points, a good idea of the total performance over time that is expected for all of the family sections can be obtained.

A family is defined by any combination of the following: *(a)* the pavement type: asphalt concrete (AC), portland cement concrete (PCC), asphalt concrete overlay on portland cement concrete (APC), and asphalt concrete overlay on asphalt concrete (AAC); *(b)* the pavement use: identified by the service rendered, such as roadways, streets, parking lots, runways, taxiways, or aprons; and *(c)* the pavement rank or functional classification, such as arterials, collectors, and local roads and streets. The criteria for pavement family selection are as follows:

Surface type

- AC = asphalt concrete
- PCC = portland cement concrete
- AAC = asphalt concrete overlay on asphalt concrete
- APC = asphalt concrete overlay on portland cement concrete

Branch use

- MT = motorpool
- ST = storage
- RO = roadway
- PA = parking
- RV = runway
- AP = apron
- HE = helipad
- TA = taxiway

Pavement rank

- A = principal
- B = arterial
- C = collector
- D = industrial
- E = residential

The family models were designed to allow continual updating as more data are gathered for a given location and entered in the data base.

This grouping of pavement sections with similar characteristics into families is the equivalent of including three additional variables in the model: pavement type, traffic, and pavement use. Furthermore, by grouping together pavements from individual geographic locations, climate is implicitly included as a variable in the family models. This family grouping makes it possible to develop reasonably accurate relationships for pavement condition prediction. Whereas other variables besides pavement type, traffic, pavement use, age, and climate certainly play a role in determining pavement condition, the availability of any such additional data in the PAVER data bases is currently limited or nonexistent. The present definition of family, which was used in the study by Nunez and Shahin *(3)* and also in this study, is provisional and will be further refined as more data become available. Efforts are not being made to incorporate a cause-of-distress factor, structural versus nonstructural, into the family definition *(4)*.

A screening procedure was designed *(3)* to examine the retrieved family data for obvious errors, and a statistical outliers analysis was implemented to detect any unusual observations. After the family data had been screened for these errors and outliers, polynomial curves were then fitted to the PCI versus age data points. It was found that a third-order polynomial curve best modeled the relationship of PCI versus age. This polynomial technique proved inadequate because a curvilinear relationship with monotonically decreasing PCI is expected when PCI versus age is plotted. The resultant third-order polynomial curve, when fitted to a family's data points, did not universally follow this known trend as the curve was not always smooth or monotonically decreasing.

Nunez and Shahin attempted to solve this problem by grouping data points within a family into 3-year averages and a third-order polynomial curve was fitted through these representative points. This was done under the assumption that the average pavement condition for the 3-year age ranges will influence the regression curve equally, regardless of the number of points in every range. However, the resultant curve was still not always smooth or monotonically decreasing. Therefore, the search for an applicable mathematical modeling scheme that would ensure decreasing PCI values as the age variable increases was expanded to include new methodologies.

APPROACH FOR NEW METHODOLOGY

Several requirements for the suitability of a particular method for pavement condition modeling and prediction must be met, including:

1. The PCI cannot be greater than 100 or less than zero,
2. The function must have a PCI equal to 100 at age equal to zero,
3. The function representing PCI versus age must be strictly decreasing as the age value increases,
4. The procedure must be suitable for automatic processing,
5. The procedure must be fast and capable of handling several hundred observations,
6. The procedure must be capable of being updated to reflect the addition of new condition survey data to the data base,
7. The procedure must be capable of being integrated into the current PAVER system without being obtrusive to the user, and
8. The procedure must accept user input information in developing family curves.

To develop an acceptable model for PCI prediction, several mathematical curve-fitting techniques were applied to historical pavement condition data. First the data were grouped into families and then subjected to a filtering procedure and an outliers analysis, as in the previous study by Nunez and Shahin *(3)*. Two polynomial curve-fitting techniques based on *B*-spline approximation and constrained least-squares estimation were applied to the filtered family data to determine their ability to satisfy the previously defined requirements. Extrapolation techniques that allow prediction of PCI values beyond the last PCI versus age data points were also studied. The following sections present the results of these investigations, including a summary of the drawbacks and advantages of using these curve-fitting and extrapolation techniques for the purpose of predicting future pavement condition.

PAVEMENT CONDITION DATA

PAVER Data Bases

To establish a modeling procedure that is representative and applicable throughout the United States, a comprehensive selection of pavements must be evaluated using that procedure.

The data bases used in this study came from cities, counties, and military installations that have adopted PAVER. Data in the PAVER system are stored on a section-by-section basis. Data normally available for each section include: pavement section identification, functional classification, age since construction or major rehabilitation, PCI, layer-material properties and, in some cases, traffic records. Table 1 summarizes the location of the data bases used in this study.

TABLE 1 SUMMARY OF THE NUMBER OF CASES BY DATA BASE

Data Base	No. of Cases
Abilene, Tex.	214
Ada, Idaho	742
Bellingham, Wash.	263
Billings, Mont.	93
Bloomingdale, Ill.	163
Bryan, Tex.	336
Calgary, Alta., Canada	891
Glenn Ellyn, Ill.	271
Hayward, Calif.	1,236
Niagara, Ont., Canada	617
Overland Park, Kans.	305
Tacoma, Wash.	314
Winnipeg, Man., Canada	113
Fort Eustis, Va.	221
Great Lakes, Ill.	559
Total	6,338

The PCI of a pavement is determined from the type, severity, and quantity of distress and is a composite index of the pavement's structural integrity and operating condition. The PCI is expressed as a numerical value ranging from 0 to 100 that has been divided into seven descriptive categories, ranging from failed to excellent. Based on the value of the PCI, an appropriate level of maintenance and rehabilitation can be recommended.

Family Definition: Data Retrieval and Organization

Data retrieval is accomplished with an automatic extraction program that selects information about the pavement section identification, PCI, and age. This information is retrieved based on the user-specified definition of a pavement family. A pavement family is defined as a group of pavement sections with the same pavement type, pavement use, and pavement rank. The ability of the users to set family definitions that may be unique for their particular location provides freedom to develop models specifically for that particular location.

Data Screening

Data Filtering Procedure

After data are retrieved it is necessary to filter out the inaccurate data. This is accomplished through a specially developed computer program. In the data-filtering procedure, the family data are first sorted by pavement section identification number, age, and PCI. When the same section is listed more than once, sequential cases of the same section are compared. If the PCI increases with age and the increase is greater than 20 points, the case with the higher PCI is removed to the errors file. This condition indicates that either an error is present in one of the records or that major rehabilitation has been performed between condition surveys, which would place this section in a different family of pavements. If a pavement section of the same age is listed more than once and the PCIs are the same, only one pavement section is retained. If the PCIs are different for the same section and age, all cases are removed to the errors file.

A further check on spurious data is done using a set of PCI boundaries *(3)*. These boundaries are defined by the maximum and minimum PCI values expected over the life of the pavements. A user can set these boundaries or use the default values, which are based on observations of available data bases combined with engineering judgment. If a record falls outside either the defined upper or lower boundary, the record is removed to the errors file. The user can then examine this file and check these data for possible problems. An adjustment to the PCI envelope can then be made to incorporate the data points that were removed to the errors file if, after examination, the user believes these data to be accurate. An example output from the data-filtering procedure is shown in Figure 1.

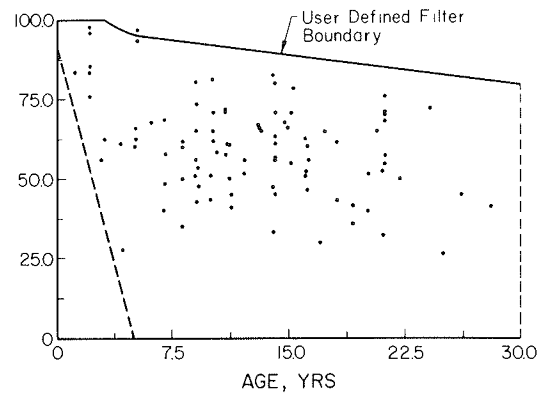

FIGURE 1 Example output of data-filtering procedure.

Outlier Analysis Procedure

The data-filtering procedure removes obvious errors from family data. To ensure appropriate model building, further examination of the data for removal of extreme observations is performed in the outliers analysis procedure. This step is important because cases with unusual performance can have a substantial impact on the statistical modeling of family behavior. The outlier procedure is performed by examining the residuals from PCI versus age curve fitting. In the previous study *(3)*, the residuals were calculated as the difference between the observed value and the value predicted by a linear regression model of PCI against age. In the current procedure, a constrained least-squares technique replaces the linear regression technique because the shape of the constrained least-squares model more closely resembles accepted pavement deterioration behavior.

A program was written to generate the residuals that are calculated as the difference between the observed PCI values and the predicted PCI values. The frequency distribution of residuals was found to be normal *(3)*, which allowed setting confidence intervals for scaling the spread of the data. The outlier program was developed to allow for various confidence limits (i.e., 95 percent), to be established by the user for removal of extreme cases. Figure 2 is an example output from the outlier analysis procedure.

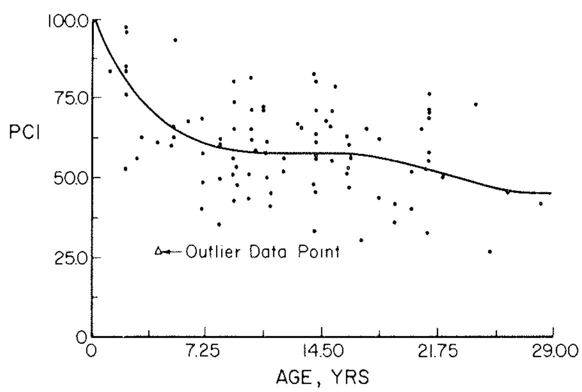

FIGURE 2 Example output of outlier analysis procedure.

DEVELOPMENT OF MATHEMATICAL MODELS

B-spline Approximation

To address the problems resulting from using a polynomial curve fitting procedure, a normalized *B*-spline function *(5)* was chosen to approximate the PCI/age data. Spline functions can best be explained as mechanical splines, which are flexible strips of elastic material. The mechanical spline is secured by means of weights at selected points, referred to as knots, through which the draftsman wishes the spline to pass. The spline assumes a shape that minimizes its potential energy. Elementary beam theory suggests that the function describing this mechanical spline is a cubic polynomial between each pair of knots. Adjacent polynomials join continuously with continuous first and second derivatives.

B stands for basis, or for the bell shape that characterizes such functions. A *B*-spline of degree k is a continuous function with its first $(k - 1)$ derivatives being continuous. In approximating several PCI/age families, it was found that *B*-splines of degree as low as 3 are sufficiently smooth to be useful.

An important consideration in using this mathematical model is the choice of knots for these *B*-splines. Sensible choices of the number and positions of the interior knots may often be estimated by examining the shape of the desired curve, but generally this is a process that requires advanced engineering judgment. Failures in this selection process can result in functions that are not strictly decreasing.

Initially, age data were averaged on an annual basis. This still produced an occasional positive trend in the function, as shown in Figure 3, a condition defined previously as unacceptable. In an attempt to rectify this situation, the age data were averaged on a 3- and 5-year basis. This produced smoother curves and a reduced occurrence of positive slopes, although functions that are not strictly decreasing can still result.

There are several causes for the *B*-spline function to take on a positive slope. One is the presence of more than one family in

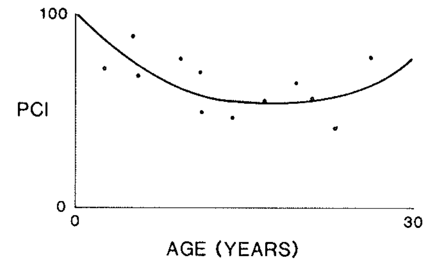

FIGURE 3 *B*-spline approximation: annual data averaging.

the retrieved data, as illustrated by Figure 4. Another possibility for an increasing trend in the *B*-spline function is a case in which a section is shown to have an unexpectedly high PCI at some age. The *B*-spline function will follow the data trend regardless of error, as shown in Figure 5. Note that when the extreme data point is removed, the *B*-spline function no longer exhibits a positive slope. Finally, a poor choice of the number and positions of the knots can result in functions that are not strictly decreasing.

Because of the complex nature of selecting the interior knots and the possibility of the occurrence of a positive trend in the function, the *B*-spline technique was not deemed suitable for inclusion in the pavement prediction modeling technique for PAVER. However, it can be used effectively in analyzing data and warning the user of potential problems and assisting him in identifying them. These uses of the *B*-spline approximation model will be discussed further in a later section. Another

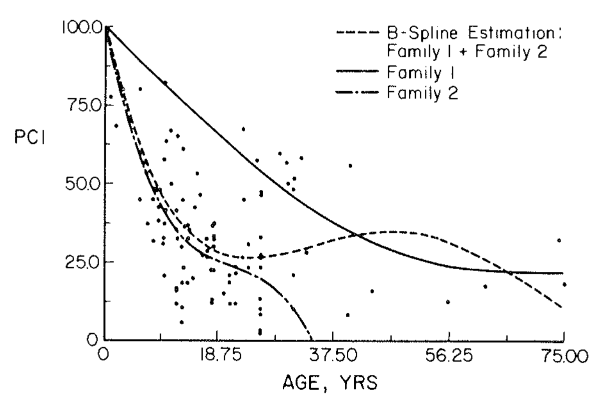

FIGURE 4 *B*-spline approximation: data file containing two families.

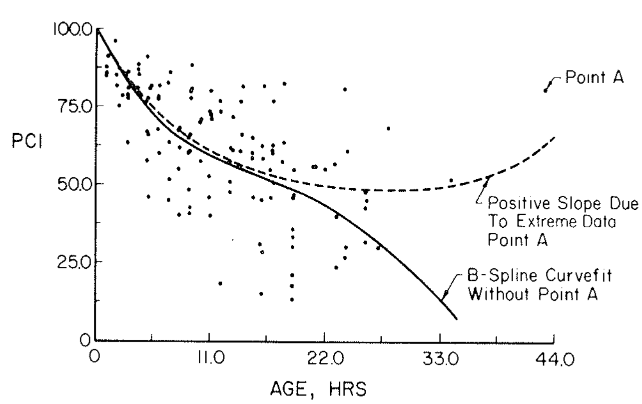

FIGURE 5 *B*-spline approximation: data file containing extreme data point.

approximation scheme, the constrained least squares, was adopted for integration into the PAVER prediction technique.

Constrained Least Squares

Constrained least-squares estimation *(6)* approximates the given observations using polynomial curve fitting. Data analysis showed no reason to use a polynomial of degree exceeding 4 or 5. The curves are mathematically constrained by the requirement that the first derivative of the curve at any age is kept negative or zero. This ensures that PCI values do not increase with age. The following constrained linear least-squares function was adopted:

$$\text{Minimize} \quad \sum_{k=1}^{N} [y_k - P(x_k)]^2$$

Subject to $P(0) = 100$ (initial PCI)
$P'(x_j) < 0$ for any x_j between 0 and maximum age.

Here, $P(x)$ denotes a polynomial value of degree υ at age x, i.e., $P(x) = \alpha_0 + \alpha_1 x + \alpha_2 x^2 + \ldots + \alpha_\upsilon x^\upsilon$.

Note that $\alpha_0 = P(0) = 100$.

Using constrained least-squares estimation to model the relationship between PCI and age of a pavement has several advantages. Whereas *B*-spline approximation may result in the best fit of data points when all data points are accurate and belong to a single family, constrained least squares avoids the potential problems that a *B*-spline function may present. Constrained least-squares curves, unlike *B*-spline curves, will never exhibit a positive slope; that is, the PCI values are not allowed to increase with age. Also, constrained least-squares estimation is simpler to use and, if the data file contains errors or data from more than one family, it is more accurate than the *B*-spline approximation.

In summary, the constrained least-squares estimation model can obtain reasonable accuracy without incurring the problems associated with *B*-spline approximations. Shown in Figure 6 are the results of fitting an accurate, single-family data file with *B*-spline and constrained least squares. It can be seen that the curves are reasonably similar. It is advantageous to use the constrained least-squares estimation on a file that contains data errors or mixed family data. The *B*-spline curve may reveal a positive trend in these cases, whereas the constrained least-squares curve will not. Thus, constrained least squares will more accurately predict normal pavement deterioration behavior.

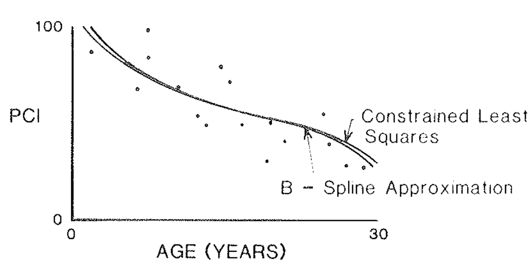

FIGURE 6 ***B*-spline versus constrained least-squares curve-fitting techniques: accurate data file.**

APPLICATIONS

Data Screening

Figure 7 contains the flow diagram depicting the iterative procedure a PAVER user will follow when defining pavement families and modeling their condition deterioration. First, the user defines the pavement family desired, under "Family Definition". Once all the data for that family have been retrieved, the data are filtered to eliminate inaccurate data in "Data Filtering for Errors". Next, the pavement family is subjected to an analysis of residuals based on constrained least squares "Outlier Analysis".

After all data have been screened in the data filtering procedure and outliers-analysis procedure, *B*-spline approximation is applied to the data to check PCI versus age slope. It was shown previously in this paper that a *B*-spline curve may take on a positive slope if errors are present in the data or if there is a mix of more than one family's data in the file. Although this ability to take on increasing PCI values as age increases limits the usefulness of *B*-spline approximation for prediction, it makes it very useful for identifying inconsistencies in the family's data file. If a positive slope does occur, the user has the option of returning to "Data Filtering", returning to "Family Definition", or continuing on to the portion of the program that applies a constrained least-squares curve-fitting technique to model the data.

Pavement Condition Extrapolation

The ability to predict pavement condition beyond the maximum age available in the data base is needed for PCI prediction into the future for the purpose of budget optimization.

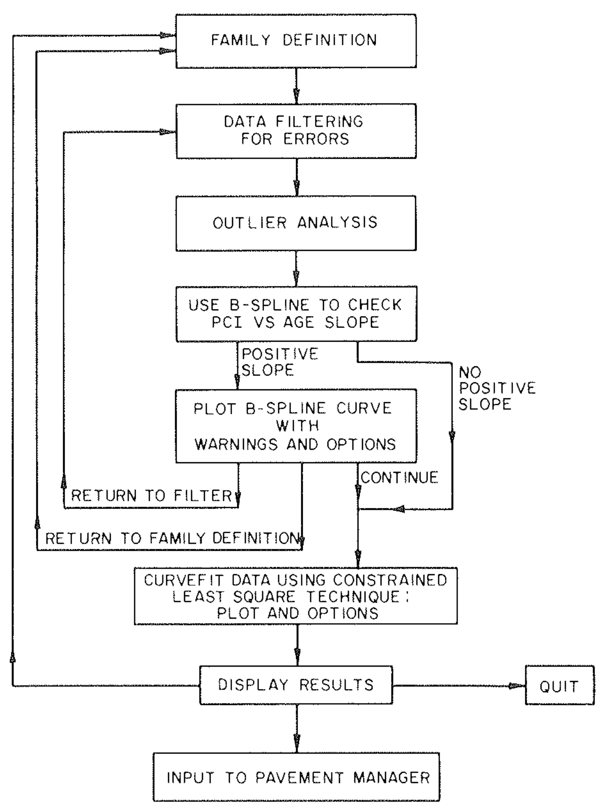

FIGURE 7 Flow diagram of family definition and curve-fitting procedure.

Both family condition and section condition must be extended beyond the available data. First, the family curve is extended beyond the age contained in the family data base. Next, pavement section condition must be predicted in a two-part procedure: *(a)* a curve must be established for a particular pavement section, and *(b)* this section curve must be extrapolated beyond the data points for that particular pavement section. This two-part section condition extrapolation is necessary to provide future predictions that provide crucial inputs into the maintenance and rehabilitation planning and economic analysis routines of PAVER. Both family and pavement section condition projection are discussed in the following sections.

Pavement Family Condition Extrapolation

The first approach taken to extend family prediction curves beyond the available data was based on using a polynomial developed from the constrained least-squares estimation. In order to use this constrained least-squares scheme for prediction beyond the maximum age in the observation table, the constraints were extended so that they would hold for any age between zero and the maximum age plus any desired number of years (e.g., 20). However, this approach produced curves with very steep slopes at the longer age values. An unacceptably rapid failure rate for the pavements resulted.

Applications of different straight-line extrapolations beyond the last point in the data base were also studied. Extension with a line of the same slope as that of the tangent to the curve at the last data point had the disadvantage of being unpredictable. The approach adopted was to use a backtracking method. The slope of the line between the last data point and the data point corresponding to an age 3 years before the last data point is determined. This slope is used to extend the curve beyond the last data point. This method, as shown in Figure 8, was determined to give the most consistent and believable results from among a number of possible methods considered.

Pavement Section Condition Extrapolation

The method used to extrapolate section condition is essentially the same as that proposed previously *(3)*. At the project level,

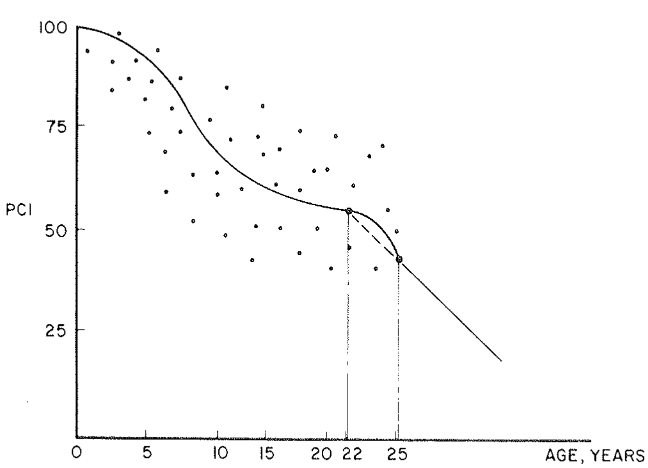

FIGURE 8 Pavement family condition extrapolation.

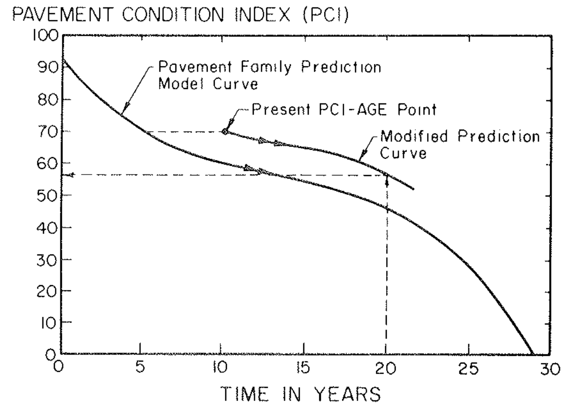

FIGURE 9 Pavement section condition extrapolation.

condition predictions are used to develop maintenance and repair alternatives. PCI prediction at the section level uses the pavement family prediction model curve. The prediction function for a pavement family represents the average behavior of all the sections of that family. Because climate, traffic, pavement type, and pavement use are the same for all sections in a pavement family, using the family curve should provide a reasonable estimate of section behavior.

The prediction for each section is done by using its relative position to the family prediction curve. It is assumed that the deterioration of all pavements in a family is similar and is a function of their present condition only regardless of age. A curve is drawn through the PCI-age points for the pavement section being investigated, parallel to the family prediction curve as shown in Figure 9. Mathematically, this is done by finding the roots of the polynomial function corresponding to the PCI value for the section. The PCI can then be determined at the desired future age.

SUMMARY

Pavement management relies heavily on the ability to predict pavement condition in the future based on historical data. By grouping pavement sections with the same pavement type, use, and rank together it was possible to develop reasonably accurate relationships for predicting pavement condition. The technique presented for modeling pavement deterioration behavior was developed specifically for use in PAVER, but it may be effectively used by other pavement management systems that use historical condition data.

After pavement sections have been drawn from a data base and grouped into families, the data pertaining to these sections are screened. A data-filtering procedure removes obvious errors and an outlier analysis procedure removes any exceptional pavements based on examination of the residuals of a constrained least-squares estimation model.

Once the initial screening has been accomplished, a *B*-spline approximation is applied to the data. If a positive slope results in the *B*-spline curve, it is taken as a warning. The data file is examined to determine whether more than one family of pavements is represented in the data file or if the presence of data errors is causing the increasing trend in the curve. The user then has the option of continuing the procedure or returning to the family definition or data-filtering phases of the program.

The technique selected for incorporation into PAVER to model the relationship between PCI and age of a pavement was constrained least-squares estimation. The procedure includes a least-squares estimation scheme that approximates the given observations with the constraint that the slope of the curve at any age must be kept negative or zero.

It is necessary to extrapolate pavement deterioration behavior beyond the maximum age available in a data base. Extrapolation of pavement family condition values is accomplished by using a backtracking method. PCI prediction at the section level is accomplished using the position of a section relative to the family prediction curve.

These procedures present a complete method to model and predict pavement family behavior and pavement section behavior. The present definition of a family is provisional and will be developed further as the data become more refined. This modeling technique was designed so that when more data are incorporated into the data base the model does not become obsolete but rather is improved by the increment in the number of observations. The modeling is simple, fast, and reliable. The model is being integrated into the PAVER program to improve the pavement evaluation process.

ACKNOWLEDGMENT

This project was funded by the U.S. Army Construction Engineering Research Laboratory.

REFERENCES

1. M. Y. Shahin and S. D. Kohn. *Pavement Maintenance Management For Roads and Parking Lots.* U.S. Army Corps of Engineers, Construction Engineering Research Laboratory, Champaign, Ill., Technical Report M-294, Oct. 1981.
2. D. E. O'Brien III, S. D. Kohn, and M. Y. Shahin. Prediction of Pavement Performance by Using Nondestructive Test Results. In *Transportation Research Record 943,* TRB, National Research Council, Washington, D.C., 1983, pp. 13–17.
3. M. M. Nunez and M. Y. Shahin. Pavement Condition Data Analysis and Modeling. In *Transportation Research Record 1070,* TRB, National Research Council, Washington, D.C., 1986, pp. 125–132.
4. A. A. Butt, K. J. Feighan, M. Y. Shahin, and S. Carpenter. Pavement Performance Prediction Model Using the Markov Process. In *Transportation Research Record 1123,* TRB, National Research Council, Washington, D.C., 1987.
5. C. E. Boor. *A Practical Guide to Splines.* Springer-Verlag, Inc., New York, N.Y., 1978.
6. C. L. Lawson and R. J. Hanson. *Solving Least Squares Problems.* Prentice-Hall, Englewood Cliffs, N.J., 1974.

The views of the authors do not purport to reflect the position of the Department of the Army or the Department of Defense.

Publication of this paper sponsored by Committee on Pavement Management Systems.