An Asphalt Pavement Rating System Based on Highway Maintenance Engineers' Experience

DANIEL S. TURNER, JAMES V. WALTERS, TERRY C. GLOVER, and EDWARD R. MANSFIELD

ABSTRACT

Research was conducted at The University of Alabama for the State of Alabama Highway Department (AHD) to define the relationship between asphaltic pavement distress and the opinions of experienced highway maintenance engineers. Seasoned maintenance engineers evaluated a series of sites across the state to determine the required level of maintenance. Thirty-one inspectors rated 1,086 sites in terms of minor maintenance and overlays of structural rehabilitation. The ratings were based on a linear scale calibrated to maintenance needs through a Delphi study. Following the inspections, the engineers' ratings were correlated to distress measurements at each site. An extensive regression study produced a model that was strong, with an R-square value of 0.74 and a small standard error. The research staff and highway department engineers were quite pleased with the strength of this model. The AHD now measures pavement distress for a test section of each lane-mile, and applies the predictor model to yield a numerical score equivalent to the rating of maintenance engineers. This is the key to the required level of maintenance. It allows prompt tabulation of the required statewide maintenance, and development of a priority listing for completing the work.

The State of Alabama Highway Department (AHD) is currently developing a Pavement Management Program (PMP) to systematically measure and evaluate the condition of roadways within the state. This effort is intended to develop a methodology to identify those roadways most in need of immediate maintenance.

The AHD hopes to change pavement management decisions from a subjective judgment by widely scattered individuals to a more sophisticated process that incorporates managerial decision policy and engineering analysis into an optimization system.

OBJECTIVES

The ultimate goal of this project was to examine the pavement distress at specific sites and to relate these observations to a rating scale for maintenance or replacement priorities. Five specific objectives were involved in reaching this goal. They are as follows:

1. Devise a rating scale to convert subjective ratings of pavement condition (by experienced maintenance engineers) to a numerical scale using the Delphi method.
2. Conduct subjective evaluations of existing pavements across the state by having a significant number of district, division, and maintenance engineers rate multiple sites using the preceding scale.
3. Obtain pavement distress data for the sites visited by the district, division, and maintenance engineers.
4. Perform a multiple regression analysis as well as an analysis of variance to relate the ratings of experienced engineers to pavement distress measurements.
5. Develop a simple methodology (regression equation or serviceability index) to provide a priority system for selecting maintenance actions for state roads.

BACKGROUND

For any roadway, the condition of the pavement deteriorates during its service life. In general, the life cycle is an S-shaped curve, but the specific steps and degrees of curvature vary from road to road. An example curve is shown in Figure 1. The curve illustrates that early in a pavement's life, it is stable and of high quality, but with time it begins to deteriorate, reaching a point where this deterioration increases significantly over a short time span. As it approaches the end of its useful life, the deterioration rate begins to stabilize, but the pavement is considered to be in poor condition.

As highways age, the maintenance requirements change. Referring to points A, B, and C (shown in Figure 1), highway engineers would like to define the locations on the life-cycle curve where a highway just begins to need (a) routine maintenance (pothole patching, crack sealing), (b) resurfacing (overlay), and (c) major structural work. The numerical values of these three points establish a yardstick for measuring pavement distress. Experienced engineers could use these values for any given roadway to determine the type and amount of maintenance required at the present time, and to forecast future maintenance needs. Unfortunately, we have not yet learned how to accurately define points A, B, and C, even though there are many ongoing research efforts concerning this issue.

The University of Alabama, P.O. Box 1468, University, Alabama 35486.

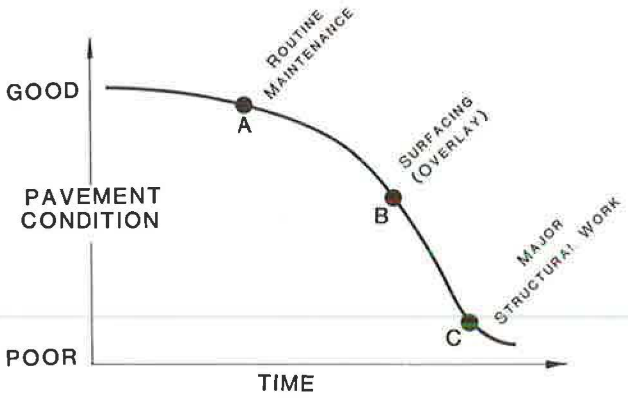

FIGURE 1 Typical asphaltic pavement life cycle.

THE ALABAMA PROGRAM

Substantial annual expenditures are required to preserve and maintain Alabama's highway system investment. Resurfacing of approximately 700 mi of roadway cost $24 million in 1983 and rose to $28 million for the same number of miles in 1984. Available funds for pavement maintenance were limited. Yet, as of the end of 1984, 35 percent of the current highway system needed repaving at an estimated cost of $198 million.

Overview

The AHD utilized a steering committee to tailor its PMP to fit its specific needs. The PMP concentrated on two areas: (a) safety, and (b) structural performance. The safety aspects of the program were handled by an intensive skid measurement program and an analysis of pavement characteristics that contributed to roadway friction. The skid value study was not a portion of the current research.

The second emphasis area in Alabama's PMP was pavement structural performance, as measured by ride quality and observations of physical distress. This article outlines the research effort into relating ride quality-physical distress to required levels of maintenance, using the experience of seasoned maintenance engineers.

Data for the PMP

Before Alabama's PMP could be implemented, a suitable data base had to be assembled to describe existing pavement conditions. Certain questions had to be answered about the data including

1. How many variables should be measured?
2. What will each distress variable be called?
3. How should each distress be measured?
4. How many levels of each distress should be established?
5. How will each level be defined? and
6. How will uniformity of data be ensured?

Because no national standards have emerged for PMP data, these were difficult questions to answer.

The development of the distress collection system began with a review of roads throughout the state to visually examine and assess their condition. Following that, a series of distress variables was selected and defined, as shown in Table 1. All the necessary measurement techniques were developed and tested from September through November 1983. To guarantee the consistency and suitability of these physical measurements at each site, the AHD published a guidance booklet for field inspections. The Research Division of the AHD Bureau of Materials and Tests engaged four crews to perform the site surveys. Those crews were trained intensively for one month (March 1984) in safety and in the identification and measurement of structural defects. A series of practice surveys was conducted to ensure consistency of ratings. At each AHD division, a local engineer was added to the crew to help with site identification and safety.

TABLE 1 Physical Distress Measurements at Each 200-ft Site

Distress Type	Measurement
Alligator cracking[a]	
Level	
1	Hairline cracks
2	Pattern of pieces lightly spalled
3	Pattern of pieces severely spalled
Patching (example photographs used to guide this rating)	
Level	
1	Good
2	Fair
3	Poor
Raveling	
Level	
1	Slight (localized, wheelpaths, or entire lane)
2	Moderate (localized, wheelpaths, or entire lane)
3	Severe (localized, wheelpaths, or entire lane)
Bleeding	
Level	
1	Localized
2	Wheelpaths
3	Entire lane
Block cracking[b]	
Level	
1	Hairline to 1/8-in.
2	1/8-in. to 1/4-in.
3	1/4-in. +
4	Spalled
Transverse cracking[b]	
Level	
1	Hairline to 1/8-in.
2	1/8-in. to 1/4-in.
3	1/4-in. +
4	Spalled
Longitudinal cracking[b]	
Level	
1	Hairline to 1/8-in.
2	1/8-in. to 1/4-in.
3	1/4-in. +
4	Spalled
Outer rut depth (in ft from start of site)[c]	
Location	
1	0
2	50
3	100
4	150
Inner rut depth (in ft from start of site)[c]	
Location	
1	0
2	50
3	100
4	150
Shoulder	
Type	
Condition	
Separation	
Drop-off	
Heave	

[a] Measure the area, in square feet, of each distress level.
[b] Measure the length in feet, of each distress level.
[c] Measure in hundredths of an inch.

A significant product of the PMP was the development of a comprehensive data management system. Pavement distress data, construction history, geometric data, traffic volume data, and ride quality information [present serviceability index (PSI)] were gathered and stored in computer files. The most important of these is the pavement distress data gathered in the statewide survey of highways. This time-consuming task of data collection requires a biennial inspection of a 200-ft long test site for each of the 30,000 lane-miles of AHD-maintained roadways.

Data collected by AHD crews were proofed, stored in computer files, edited, and reformatted during the course of the state road survey. In the final data set, each site was represented by a 145-character-wide field. On the completion of site visits, this represented some 4 to 5 million characters of computer-filed data. This comprehensive data set was delivered to the University to form the basis for the statistical analysis.

DEVELOPMENT OF A RATING SCALE

The success of this project depended on the establishment of a rating scale to describe AHD engineers' evaluations of the physical condition of any pavement site. This scale described the required level of maintenance by a simple numerical rating. The project staff used the Delphi technique to gather expert opinions and form the rating scale.

THE DELPHI TECHNIQUE

The Delphi technique is used to gather a consensus opinion (on a technical topic) from a homogeneous group of experts. Where no historical data exist, these expert opinions represent a good way to make a difficult judgment. In this case, the difficult decision was how to define points A, B, and C on the pavement life curve (see Figure 1). The experts were engineers from the AHD whose daily work activity called for such judgments. The expert opinions were solicited through a questionnaire, and were then refined through an iterative process using additional questionnaires.

The Questionnaire

The initial questionnaire was carefully worded to enhance the probability of prompt and accurate responses by the participating experts. For example, to avoid imposing monitor views and preconceptions on the respondents, all evaluation sheets were carefully worded to reduce implied responses; verbal discussion and instructions strongly denoted that the interest was in AHD engineers' perceptions of the rating scale. Other safeguards were taken, including engaging a scaling expert to help devise the rating form. The resulting questionnaire included a short set of instructions, a 0-to-100 scale, definitions of the three types of maintenance (points A, B, and C), and blanks into which the inspector was to insert values for the three points. The form was also stratified into three traffic volume levels to create a three-by-three matrix requiring nine entries by the expert.

Conducting the Delphi Technique

The AHD submitted 63 names of employees at several levels of management to participate in the Delphi survey. The initial round was initiated by mail in September 1984. The AHD engineers received an invitation to a pavement management conference and a copy of the Delphi questionnaire. They were asked to complete the questionnaire and return it to the research staff.

As the engineers arrived at the conference, they were advised of the results of the initial round and asked to complete a second questionnaire. The second round gave them an opportunity to revise their answers in light of the group's response. On the next day of the conference, attendees were informed of the results of the second round, and questionnaires were distributed for the third round of the Delphi procedure.

All three Delphi rounds are summarized in Table 2. The table indicates that the procedure functioned as intended. For each of the three maintenance levels, the variance among respondents was reduced on each round as the opinions of experts converged. Point A (routine maintenance) fell at 75.91, point B (major surface work) fell at 56.66, and point C (rehabilitation) fell at 37.89. The research staff and the AHD concluded that these values constituted an adequate rating scale on which to base the field evaluations.

TABLE 2 Summary Descriptive Statistics for Delphi Rounds

Question	Round 1	Round 2	Round 3
Beginning to need routine maintenance?			
Mean	77.07	76.58	75.91
Standard deviation	6.96	4.29	4.29
Variance	48.435	18.426	18.418
Beginning to need major pavement surface work?			
Mean	55.19	57.18	56.66
Standard deviation	10.74	8.21	6.97
Variance	115.297	67.361	48.588
Beginning to need rehabilitation?			
Mean	36.50	37.45	37.89
Standard deviation	13.54	8.85	6.26
Variance	183.238	78.209	39.159

THE TRAINING CONFERENCE

In fall 1984, 57 participants attended the first Pavement Management Conference at the AHD training facility in Montgomery, Alabama. There were three overall aims: (a) to explain the PMP and conduct a group discussion on pavement distress, materials, and maintenance treatments; (b) to conduct the remaining Delphi rounds; and (c) to conduct a pilot field test of the proposed pavement inspection procedures. The Delphi study has already been reported and will not be repeated here. The other portions of the conference are recapped in the following paragraphs.

CONFERENCE DISCUSSIONS

Top management officials of the AHD opened the conference by explaining the goal of the PMP and emphasizing their commitment to it. Next, engineers from the AHD Research Division carefully reviewed the development of the PMP, including the ongoing program to gather distress data. For the remainder of the afternoon, the University research team conducted a group discussion, showing slides of the major distress types and asking for feedback on the significance of the various levels of distress. There was a great deal of interaction among the participants. They were especially interested in treatments that other maintenance engineers had found to be successful.

A segment of the conference was devoted to explaining the life curve of a typical asphalt pavement, including the location of points A, B, and C from Figure 1. The engineers were led to see that the Delphi procedure would create a "yardstick" that could be used to measure pavement conditions anywhere in the state. They were then exposed to the elementary statistical concepts that would relate physical distress measurements to inspector ratings to create the final road rating scale. Finally, the pilot field test was introduced, and the participants were dismissed for the afternoon.

PILOT FIELD STUDY

The maintenance engineers had been told that they would visit sites across the state to evaluate pavements for required maintenance actions. The Delphi-derived scale was to be used for this evaluation. Because this was a new procedure, it was tested at the conference through a pilot study. Eight sites of varying distress were selected for the study and a rating form was developed.

The research staff explained the use of the form, the site locations, and inspection procedures, then dispatched the groups to perform the trial inspections. Research team members accompanied the inspectors to observe their techniques and record their discussions. The trial ratings were returned to the conference center, where they were compiled and returned to the participants in the form of histograms. A strong central tendency was noted among the ratings. The research staff explained that this was anticipated and that the pilot study results were highly satisfactory. The remaining steps in the research project were outlined, and participants were dismissed with the understanding that they would soon be called on to begin the field study.

PARTICIPANT SELECTION FOR STATEWIDE PAVEMENT INSPECTION

Back at the University, more formal analyses of the pilot field survey and the Delphi experiment were conducted to determine inspector variability, so that the final participants could be selected to perform the field evaluations. Histograms, scatterplots, and simple linear regression techniques were used to identify the AHD engineers whose pilot study ratings were the most consistent. Of the 65 potential inspectors (maintenance engineers, district engineers, division engineers, and materials engineers), only 31 were needed for the final program. Engineers were removed from the pool if their rating scores varied excessively from the group average, if they had excessive personal variability in ratings from site to site, or if they had health restrictions or prior commitments. Of the remaining possible inspectors, a random sampling without replacement was used to select the final 31 participants required for the balanced incomplete block design.

At this stage of the project, the rating scale had been prepared and tested, AHD engineers had been trained in inspection techniques, and the overall research procedure had been verified through the pilot study. It was time to proceed with the full field study.

STATE ROAD SURVEY

Introduction

Sites for the field survey were selected from all candidate sites for which distress data were available. The criteria for selecting sites included the number of candidates and the types of distress prevalent in each division. Locations for the field examinations were thus representative of all sites available.

Because it was physically impossible for each AHD investigator to visit each study site, a statistical sampling procedure was used for assigning engineers to sites. This procedure, the Balanced Incomplete Block (BIB) Design ensured that each site was visited often enough to provide statistical validity. It also ensured enough overlap between inspectors to isolate and identify sources of variability. The specific research procedures utilized during this part of the study are described in detail in the remainder of this section.

FIELD SITE SELECTION

Raw distress data supplied by the AHD were sorted by division, route, beginning milepost, and county. Because data collection was not completed for all divisions, the number of sites was determined for each division based on the percentage of data present. An appropriate adjustment factor was assigned to give each division a weighted value that it would support for the random selection of field sites.

From the evaluations of 16,994 lane-miles (51 percent of the total state) of roadway available at the time of selection, frequency distributions were prepared to determine the number of unique categories of distress data. Using the frequency distributions, the distributional characteristics, and the knowledge gained from the conference in Montgomery (concerning priority of importance in distress factors), the principle distress in each division was identified. Sets of relationships among variables were examined using an elaboration analysis to find the primary indicators and the patterns.

A contingency table was prepared to examine the data resulting from a proportional selection of sites. If the primary distress variables were used to select sites, certain combinations of distress would be underrepresented. To overcome some of this problem so that a more diversified selection might be randomly generated from the divisionally stratified data base, weighting factors were used, and the ranges with small proportions were combined in hopes of favoring some random selection in those categories where only minimal data were available.

Several levels were defined for the primary variables. This would ensure that inspectors saw a range of conditions in the field, resulting in high and low rankings to increase the chances for a strong regression fit. Using a computer program prepared by the research staff, the proportioned divisional data, distress levels, and weighted distress factors from the cross-tabulation analysis were entered for each division in an attempt to simulate distress data combinations of all types in the final sample. The stratified random selection chose 326 sites of the 16,994 available, which allowed for replacement sites if they were needed later in the study. These 200-ft long sites were plotted on an Alabama road map using pins of different colors to represent the proportional groupings of categorized distress available for selection of sites. Pins were placed on the map to represent available AHD engineers so that the research staff could gain an appreciation for the geographical distribution of sites and inspectors before designing the field inspection program.

As a result of cost constraints, it was not feasible to use one design over the entire state. To minimize expense, an attempt was made to hold the travel of the participants to a reasonable limit. An alternative design was implemented. The state was divided into three geographical regions: North, Mid, and South. All the inspectors would visit some sites in the Mid section and some in their base region. Hence, three separate BIB designs were constructed. In design 1 (North), 80 sites and 16 experts were used with each site inspected by 6 experts and each inspector visiting 30 sites. In design 2 (South), 15 experts each visited 28 of the 60 sites, and each site was visited by 7 experts. Design 3 (Mid) consisted of 31 sites and all 31 inspectors (16 from North and 15 from South), with each expert visiting 6 sites. Hence, experts from the North would visit 36 of the 111 sites available to them and experts from the South would visit 34 of the 91 sites in their region.

PARTICIPANT AND SITE MATCHING

Constraints were used in assigning field workers to specific sites. No experts were assigned to any sites within their home districts. Division personnel were not assigned to any sites within their entire divisions. A computer program was written to choose sites and inspectors from the available pool and produce a compatible program of site visits that would minimize travel while honoring the constraints involving visits within their home districts. The final design was an acceptable program whereby the 16 North experts visited 30 sites in their own region and 6 sites in the Mid region. The 15 South experts visited 28 sites in their own region and 6 sites in the Mid region. This overlap in the Mid region allowed an assessment of variability among the North and South inspectors.

Survey Formulation and Reporting Form

All the preparation for the field study had to be completed in a minimal time because the approaching cool weather would alter the distress conditions of the asphaltic pavement. It was desirable to have a minimum of 1 week's delay after the conference to allow inspectors to reflect on the rationale behind the scale's creation, and to remove the effects of any possible peer pressure created at the conference.

Forms detailing the site information were prepared for each location. In addition to supplying a site rating at a specific milepost, additional data required of the inspector included the date, time of day, weather condition (sunny, cloudy, or overcast), and pavement condition (dry, damp, or wet). These forms were then mailed to inspectors who began to review the sites in the middle of October. Approximately 5 working days spread over a 2- to 3-week period was the anticipated time that each inspector would need to devote to the field evaluation. After the evaluations were performed, the rating sheets were returned to the University for analysis.

ANALYSIS OF FIELD STUDY DATA

The purpose of the statistical analysis was to assess the consistency of the inspectors in using the rating scale, and to develop a regression model that related the observed rating scores with the physical measurements of distress. Each of the inspectors had a record sheet for each site visited; hence, there were 1,086 records [(36 x 16) + (34 x 15)]. After the data were delivered to the University, each sheet was reviewed for compliance with the correct inspection procedure, and then its information was added to a computer data file.

Variance of Inspection Data

To determine the amount of variability among inspector ratings at each field site, a computer program was developed and applied to each region. The Analysis of Variance tables and the estimates of variance components are given in Table 3. In all three regions, the null hypothesis that the variance component was zero was rejected because the F-statistic (ratio of the explained variance to the unexplained variance) for inspectors adjusted for sites was larger than the calculated p-value (the probability that the hypothesis was correct).

TABLE 3 Analysis of Variance Tables and Estimates of Variance Components

Source of Variation	df	F-ratio	Estimates of Variance Components	Standard Deviation
Sites				
North region	79	13.29	134.7	
Mid region	30	29.61	271.4	
South region	59	23.38	192.6	
Inspectors adjusted for sites				
North region	15	7.67	17.3	
Mid region	30	3.46	29.9	
South region	14	4.70	9.2	
Error				
North region	385	69.3	9.30	
Mid region	125		62.9	9.63
South region	346			8.54

Note: Data are uncorrected.

The estimated variance for inspectors and the standard deviation of ratings for a given site are given in the table. The variances ranged from 9.2 to 29.9, although standard deviations were grouped from 8.54 to 9.63. Had the variance component for inspectors been negligible, each of the six or seven scores given by the inspectors at each site could have been used as separate responses in the model-building phase of the study. This would have resulted in more degrees of freedom for estimating unexplainable variance. As it was, the response score used for each of the 171 sites was determined by the mean of scores across inspectors. Hence, the variability of individual inspectors was averaged to get a representative score for the site.

Further Examination of the Data

The standard deviations for each site were calculated and compared to the base value found for each region. Sites with large or small standard deviations (greater than 13.0 or less than 3.0) were reviewed to determine the causes of the large and small variations, respectively. Items that were noted at these sites included

- Weather conditions,
- Road conditions,
- Time of day,
- Comments by inspectors, and
- Maintenance remarks.

In general, it was found that ratings were distinctly higher where weather and wet pavement were factors. Low visibility in the dawn and dusk hours influenced ratings taken during these times of day, causing them to be higher. Raters' comments confirmed that the actual distress had not been clearly visible to the observer. Comments also indicated that at a few locations, resurfacing projects had been scheduled for sites during the course of the observation period. Types of distress listed by observers helped to confirm the inspector's presence at the correct sites, while remarks were useful in interpreting an inspector's logic for rating assignments.

At eleven of the sites, the value of the standard deviation of ratings was three or less. These were the type of responses the research team had hoped to get for all locations. The inspection sheets for these sites were examined in great detail to determine the reasons for uniformity of ratings. Five locations were found to be freshly resurfaced, leading to high ratings. Three roads were found to have large amounts of distress, leading to low ratings. One site had just received crack sealing treatment and inspectors zeroed in on point A on the rating form. In general, ratings for these sites were given in good weather and good daylight, and close together in time.

On the other hand, there were 13 sites that had standard deviation values greater than 13.0. The individual rating sheets were reviewed to find reasons for inconsistencies. One location was immediately noted to have been resurfaced after four inspectors had seen it, but before two visitors. The ratings of the last two inspectors were removed from the study.

For three additional locations, a detailed analysis of comments revealed that one inspector had gone to the wrong test section. The erroneous ratings were deleted from the data set. At two other sites, the remarks indicated that inspectors held widely varying views on the importance of rutting. Some believed that severe rutting deserved immediate treatment; some inspectors were unconcerned. Scrutiny of the remaining data sheets showed some evidence of bias toward high traffic volumes, instances where one inspector viewed the site in distinctly different weather or light conditions, and one case of an inspector who was always higher than others. Where the research staff consistently noted erroneous or biased data, the values were removed from the study or adjusted to the value of the remaining site observations.

Development of a Prediction Model

Four guidelines were adapted for the regression study. These were designed to produce a final model (a) that had statistically significant variables (2-tail, p-values of 0.05); (b) whose coefficients estimated the true contribution of that variable as suggested by its deduction points from the rating equation; (c) whose estimated regression coefficients were logically ordered; and (d) that minimized the number of variables necessary for accurate prediction.

To conduct the analysis, the mean of the ratings given by the inspectors was regressed on the physical distress variables. The intent was to determine the functional form of the relationship between the score and a set of predictor variables. Within this process, several diagnostic measures were used to determine such things as validity of assumptions, influential observations, need for transformation of variables, and existence of collinearity among predictor variables. A variable selection procedure was then used to determine a subset of variables that might do as well as the full set.

A final item was added to the data just before the initiation of the regression analysis. Pavement roughness data (PSI measurements) were obtained from the AHD and blended into the data set. This provided another element to extend the possibilities of obtaining a strong regression analysis.

Initial Analysis

The first model estimated the multiple linear regression equation using all the independent variables while identifying the best subsets of predictor variables. (In this case, "best" was defined in terms of sample R-squared, adjusted R-squared, or similar measures.) Multiple regression techniques were used to determine which of the variables (distress and roughness) were the best indicators of pavement condition. Analysis of the raw data using the original 31 variables produced a relatively strong coefficient of multiple determination, R-squared = 0.747. This coefficient was an indicator of a good fit between the inspector's ratings and the distress data. No single variable appeared to dominate the regression analysis, although alligator cracking, transverse cracking, and severe raveling were strong contributors to the value of R-squared for most models. Analyses of residual plots and partial residual plots did not reveal any violation of regression assumptions, or the existence of curvilinearity within any of the predictor variables.

The initial analysis suggested the need for further refinements of the regression relationship. There was a linear pattern among some variables within this group; these included a correlation between outer and inner rut depths, and between rut depths at the various locations within the test site. Another difficulty involved variables for which there were few observations in the data set. In these cases, t-statistics remained insignificant and regression coefficients were positive, signifying that the distress variable added to the site rating rather than deducted from it. Any form of distress present should reduce the site's rating, not increase it. Another confounding issue was that the least severe level of some variables (level 1) was more significant than higher levels (levels 2 and 3) for the initial models. For these reasons, further statistical analyses were undertaken.

Model Testing

Multivariable regression analyses were performed in an attempt to improve the correlation by using combinations of parameters considered to have a significant effect on pavement distress. Various transformations and subequations were utilized to improve the regression. Although a complete discussion of these analyses is beyond the scope of this paper, several examples will be discussed for illustrative purposes.

Because correlation was present among the many rutting measurements taken at each test section, average values were used for the outer and inner wheel paths. Later, an examination of t-statistics showed that inner rut measurements had an insignificant contribution to the predictor equation, and this variable was deleted from the study. There were very few observations of level 3 longitudinal cracking, so levels 3 and 4 were combined into a new variable that functioned adequately in the prediction model. Similar transformations were used to produce significant variables for levels 3 and 4 of transverse cracking, and levels 2 and 3 of patching.

Dummy variables were used to assess the effects of certain factors. For example, inspectors were uncertain as to the importance of rutting. Comments on inspection forms indicated occasional fears of base failure under the pavement. A dummy variable was used to mark potential areas of base failure (i.e., roadways with outer rut depths of large value). However, because the t-statistics proved this to be a nonsignificant variable, and other effects of the transformation appeared minimal, it was removed from the study.

Several attempts were devoted to providing a logical ordering of the various levels of variables. For alligator cracking, residual plots showed that a distinct segregation existed at a level of 400 ft^2. Above this value, additional square footage made no further contribution to the model. The research staff developed two models to allow for this. One equation truncated all values above 400, while the second equation used a logarithmic relationship to account for the diminished contributions at upper levels. Both of these equations produced a good fit to the data with corresponding ordering of the distress levels.

Further studies were conducted on bleeding, raveling, block cracking, combinations of variables, and nested variables. In general, these studies contributed to the research staff's understanding of the data set and strengthened the predictor model.

Recommended Model

A large number of models were prepared and evaluated to determine the regression equation that would best fit the AHD's purposes. Some 26 combinations of the original variables and subsets thereof produced a range for R-squared from approximately 0.68 to 0.74. The standard error of estimate changed by only 0.5 throughout the process. The recommended model is quite complex. No single variable or pair of variables dominated the statistical analysis. Rather, many variables contributed to the strength of the regression models. This confirms what AHD engineers have long suspected: that many items must be considered in establishing the condition of a given pavement and in recommending appropriate maintenance treatment.

The following equation is presented as the model that best represents the data base for this study. This model is based on extensive investigation of the variables and their associated ratings relevant to the computed statistics. A solid R-squared of 0.74289 is accompanied by a reduced standard error of estimate of 8.10430. Variables that contributed most significantly to the equation's coefficient of multiple determination included (a) LNALL2, (b) BLK2RD, (c) ROUGH, and (d) TRAN12. All t-statistics were significant at less than the 2-tail, $p = 0.035$ level. The equation that represents the pavement distress as indicated by these statistics is

$$\begin{aligned}\text{Rating, } Y = 95.5727 &- 5.5085\,[5-\text{ROUGH}]\\ &- 1.5964\,\text{LNALL1} - 1.9629\,\text{LNALL2}\\ &- 2.9795\,\text{LNALL3} - .01630\,\text{PAT2RD}\\ &- .07262\,\text{BLK2RD} - .2220\,\text{AVGOUT}\\ &- 3.4948\,\text{RAVL31} - 7.5269\,\text{RAVL32}\\ &- 11.2297\,\text{RAVL33} - .03032\,\text{LONG12}\\ &- .05484\,\text{LONG34} - .53050\,\text{TRAN12}\\ &- .69736\,\text{TRAN34}\end{aligned}$$

where

ROUGH = roughness or PSI,
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), ≤ 400 ft^2,
BLK2RD = block cracking (all levels summed), ≤ 400 ft^2,
AVGOUT = outer wheelpath rutting (all levels averaged), in 10^{-2} in.,
RAVL31 = severe localized raveling (Code: 0 = none, 1 = present),
RAVL32 = severe wheelpath raveling (Code: 0 = none, 1 = present),
RAVL33 = severe entire lane raveling (Code: 0 = none, 1 = present),
LONG12 = longitudinal cracking (level 1 + level 2), in ft,
LONG34 = longitudinal cracking (level 3 + level 4), in ft,
TRAN12 = transverse cracking (level 1 + level 2) number of cracks, and
TRAN34 = transverse cracking (level 3 + level 4) number of cracks.

SUMMARY

The University of Alabama conducted a research project for the AHD to develop a rating scale for asphalt pavement maintenance actions. The scale was developed by correlating the experience of AHD maintenance engineers with measurements of pavement distress. A statistical analysis produced a regression model that was quite strong. The following key points document the success of the research project.

Research Procedures and Findings

Following are findings from the AHD study as well as notes on the research procedures:

1. The AHD had carefully designed a program to obtain distress data items thought to be pertinent to pavement management. This research project confirmed that these variables were appropriate, that the AHD's measurement scales and classifications were well founded, and that the data was of good quality.

2. The AHD gathered distress data at 30,000 sites across the state. This represented a substantial and significant data base on which to conduct the research.

3. AHD division engineers, district engineers, maintenance engineers, materials engineers, and other central office personnel participated in the study. The range of jobs and experiences provided a good background for the study.

4. A Delphi procedure was used to design a rating scale for field evaluation of test sites. The opinions of experts (AHD engineers) were used to establish the numerical values on the scale.

The Delphi procedure was found to be very suitable. During the three rounds of the procedure, the experts' opinions converged, and all common modes of Delphi technique failure were overcome.

5. A training conference was conducted in Montgomery, Alabama. The AHD participants were able to discuss common problems, distress factors, and the success of previous maintenance treatments. This conference provided an excellent forum for interaction among participants.

A pilot study was conducted to train the participants in field procedures for rating roadways. Eight sites were used in the study, which was beneficial in validating the Delphi scale, ensuring that the rating procedure was functional, measuring the variability between evaluators, and providing insight into the reasons for this variability.

6. From the distress data gathered by the AHD, a stratified random selection procedure was used to identify more than 300 initial sites for field evaluation. This ensured a wide variety of distress combinations illustrative of all types found across the state. These sites proved to be excellent for the purposes of the research project.

7. Of the AHD engineers participating in the training conference, 31 were selected for the field evaluation process. Statistical procedures were used to provide a suitable combination of geographic and experience characteristics among the engineers used in the study. The final selection of evaluators was based on a random process.

8. A comprehensive field study was designed. The initial sites were screened and reduced to 171, and the 31 AHD engineers were each assigned approximately 35 sites to evaluate. A balanced incomplete block design was used to ensure that there was sufficient overlap of inspectors and sites so that the variance caused by both of these factors could be measured and evaluated.

9. Variance among the evaluations for any given site was analyzed, along with the variance from site to site for any given inspector. Each inspector's comments and recommended maintenance treatments were reviewed to identify causes of variance, and to determine the consistency of data.

10. An intensive regression analysis was conducted to identify the relationship between the measured distress data and the opinions (ratings) expressed by AHD engineers during the field evaluations. The various parameters were examined individually and in combination with others. Estimates were obtained for the contributions of each of the variables.

11. A model was recommended for use by the AHD after extensive testing and verification.

12. The prediction model was quite strong, with a coefficient of multiple determination (R^2) of 0.75 and a Standard Error of Estimate of 8.104. The t-statistic for individual variables were all highly significant at the 2-tail, $p < 0.05$.

13. The recommended model was quite complex. No single variable or pair of variables dominated the statistical analysis. Rather, many variables contributed to the strength of the regression model. This confirms what AHD engineers have long suspected: many items must be considered in establishing the condition of a given pavement and in recommending appropriate maintenance treatment.

14. The research staff was able to account for the contribution of all variables to the predictor model. However, care must be used in interpreting distress values for two of the variables because (a) cracks in asphalt overlays over rigid pavement were not considered important by the engineers who participated in the field study although the same types of cracks were considered very important in full-depth asphalt, and (b) evaluators could not completely agree on the importance of rutting in determining the condition of any pavement.

15. The regression model may be applied to distress data for any site to compute a rating equivalent to the opinion of AHD engineers. This rating provides a ready tool for comparison of various sites.

Recommendations

1. The research staff recommended that the AHD apply the regression model to distress data gathered during the evaluation program. The model could be added to the current PMP edit program residing on the AHD's computer. This would allow automatic calculation of ratings during the data input process.

2. For each classification of roadway, the staff recommended that the AHD tabulate the number of miles at each ranking value. This tabulation provided a ready index to the condition of state roads, and yielded strong estimates of the types and amounts of required maintenance.

3. The research staff recommended that the rating values be used for preliminary assessment of sites competing for maintenance funds. Sites should be grouped in a comparative manner (by road type, geographical location, etc.) before using the rating.

4. The rating provided an estimate and not an absolute score. It should be used only as a preliminary method for comparing sites within roadway categories. The advantage of the model is that it provides a ready and quick tool to eliminate sites that do not need attention and to focus on those most deserving of further study.

Detailed field investigations should be performed as the final selection process. These detailed investigations provide more data over the complete length of a project than is possible in the 200-ft sample length used in the model.

5. The research staff recommended that the AHD continue its program of gathering distress data. The data should be gathered annually for the first three years, and then should be shifted to a biennial basis.

6. The staff recommended a comprehensive analysis of the year-to-year change in ratings for each site around the state. This study would yield the rate of change of pavement condition as a product of changes in variables such as traffic volume and pavement distress.

7. The project staff recommended that the AHD hold another conference on pavement maintenance in the coming year. Engineers should be allowed the opportunity to compare their actions with those of their peers to identify appropriate treatment techniques. This conference would be an appropriate means of introducing trend analysis and other tools that may be used to analyze the data available through the Pavement Management Program.

Following the closure of the research project, the AHD adopted the methodology recommended in the project report. The prediction equation was added to the computer edit program to automatically generate a rating value equivalent to the beliefs of maintenance engineers. This rating value is the primary data item used by AHD managers in selecting and determining the priority of sites while preparing each year's PMP.

ACKNOWLEDGMENTS

Funding for this research was provided by the AHD. Appreciation is expressed to Frank Holman, Alton Treadway, and Billy Allen of the AHD for their technical assistance and definitive guidance during the project period. Without the full cooperation and assistance of many AHD engineers, the field inspections would not have been completed and this project would not have been possible. The authors also gratefully acknowledge the contributions of Dean McLure, Willie Peoples, and Bette Jefcoat toward the successful completion of this research.

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