

Serviceability Prediction From User-Based Evaluations of Pavement Ride Quality

SUKUMAR K. NAIR and W. R. HUDSON

ABSTRACT

Presented in this paper are the results of research conducted to develop predictive serviceability equations to upgrade those currently in use by the Texas State Department of Highways and Public Transportation (SDHPT). The method has been based on the serviceability-performance (S-P) concept. Experiments were designed to study two types of variables, one associated with the rating process and the other related to pavement characteristics. The rated sections were profiled using the new Model 690D surface dynamics profilometer (SDP). From the profile data, a family of profile summary statistics called root-mean-square vertical accelerations (RMSVAs) was computed. A calibrated Mays meter and Walker accelerometer device (SIometer) were also operated on these sections. A multiple linear regression procedure was used to develop reliable serviceability equations (with good predictive capabilities) by regressing the mean panel ratings on the set of RMSVA indices. Correlation analysis of the Mays meter and SIometer measurements with the panel ratings showed that the calibrated Mays meter predicts panel ratings better than the SIometer. The best prediction of the panel ratings, however, is achieved by the 690D profilometer.

Road riding quality or roughness has special significance as it has been shown to directly affect vehicle operating costs and road safety. In previous studies, relationships have been developed between pavement serviceability and user costs. Sponsored by the World Bank, recent studies in developing countries have provided valuable quantification of road deterioration, vehicle operating costs, and road maintenance policy on road roughness. In this light, the importance of accurate and reliable measurement of road roughness cannot be overemphasized.

The serviceability of a pavement is largely a function of its roughness. Results from the AASHO Road Test (1) have shown that nearly 95 percent of the information about the serviceability of a pavement is contributed by the roughness of its surface profile. Roughness has been defined as the distortion of the pavement surface that contributes to an undesirable or uncomfortable ride (2). The American Society for Testing Materials has defined roughness as "the deviations of a pavement surface from a true planar surface with characteristic dimensions that affect vehicle dynamics, ride quality, dynamic pavement loads, and pavement drainage (e.g., profile, transverse profile, cross slope, and rutting)" (3).

In 1968 the Texas State Department of Highways and Public Transportation (SDHPT) conducted a rating session in order to obtain serviceability equations using the 1965 version of the surface dynamics profilometer (4). Since then, these equations have been the basis for the evaluation of Texas highways. With the recent purchase of the highly sophisticated new Model 690D surface dynamics profilometer, it is necessary to upgrade the roughness evaluation system

by incorporating its new capabilities in updated serviceability equations.

Another significant consideration here is the change in the average passenger vehicle. Over the years there has been a noticeable shift in vehicle population from big, heavy automobiles to smaller, lighter ones. Hence it is essential that the changes in ride-quality judgments be reflected in serviceability predictions.

RATING: A SYSTEMS APPROACH

The rating process that results in an evaluation of pavement ride quality is a complex phenomenon. Examining it from a systems standpoint, the process involves three subsystems: the vehicle, the road surface profile, and the rater (highway user). The dynamic interactions between these subsystems are responsible for the output responses and characteristics of the system. In order to understand the rating process, it would appear appropriate to study the interactions between the stimuli and these receptor systems. Consider a rater in a rating situation (Figure 1) being subjected to the physical stimulus (S), the vibrations that are being imparted to him by the vehicle. Each vibration triggers certain events in his mechanical energy receptor systems. Thus, the physical continuum evokes a corresponding sensory continuum. When the same stimulus (vibration) is presented to the same rater on different occasions, it will not always produce the same magnitude of the variable on the sensory continuum. This is where the subjectivity of the rating process is realized. Three continua are the stimulus or physical continuum (S), physiological or subjective continuum (P), and the judgmental continuum (J).

However, as has been previously reported, there exist certain basic problems typical of serviceability ratings. These possible errors (listed next with

S. Nair, Center for Transportation Research, The University of Texas, Austin, Tex. 78712. W.R. Hudson, Dewitt C. Greer Centennial, The University of Texas, Austin, Tex. 78712.

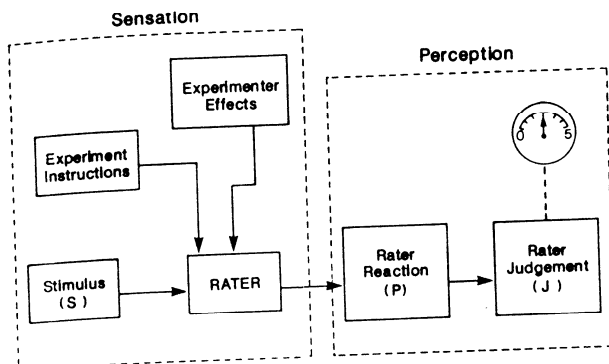


FIGURE 1 Systematized concept of rating.

suggested ways to avoid them) were kept in mind during the planning stages of the study so as to minimize or do away with them completely:

1. The error of leniency, which refers to the constant tendency of a rater to rate too high or to low for whatever reasons; remedied by statistical transformation of rater variance.
2. The halo effect, which refers to the tendency of raters to force the rating of a particular attribute in the direction of the overall impression of the object rated; avoided by accuracy and exactness in definitions.
3. The error of central tendency, which refers to the fact that raters hesitate to give extreme judgments of stimuli and tend to displace individual ratings toward the mean of the group; taken care of by introducing the judgment continuum as distinct from the sensory continuum.
4. The error of anchoring, which refers to the endpoints of the scale being rated; overcome by using accurate definitions.

EXPERIMENTAL DESIGN

In laying out the experiment design, the first step was to look through the "window" of applicability, the inference space. This is defined as that space within which the results of the study may be applied. When selecting the panel, if a wide distribution of members from various parts of Texas are chosen randomly, then the panel could be considered as representative of the people of the state of Texas.

Considerations such as the type and number of variables to include in the statistical analysis relating the profile data to the panel rating data are significant for future applications of the serviceability formulas. It is important to choose pavement test sections that include a wide range of wavelengths, as well as to consider this range of wavelengths as they influence Texas raters' judgments of ride quality. With these ideas in mind, 17 factors, along with their corresponding levels, were enumerated by a research team and a group of pavement engineers from the Texas SDHPT who were familiar with pavement roughness. To reduce the complexity of the analysis, some of the factors were studied in screening experiments and the remainder were studied in the main rating sessions.

SCREENING EXPERIMENTS

The following factors and corresponding levels were considered to be studied in screening experiments:

Factor	Level 0	Level 1
Vehicle type	Car	Van
Position in car	Front	Rear
Rater's age	< 35 years	> 35 years
Rater's sex	Male	Female
Time	Night	Day

Factors associated with the type of pavement sections are listed as pavement variables as follows:

Factor	Level 0	Level 1
Pavement type	Black	White
Surface texture	Coarse	Fine
Location of road	Rural	Urban
Maintenance	Unpatched	Patched
Functional class	Low	High
Surroundings	Poor	Good
Road width	Narrow	Wide
Lane position	Inside	Outside

In order to simplify the analytical procedure and also to make the computational procedure amenable to mainframe computer capacities, it was decided to drop the factors (a) functional class and (b) lane position. A panel of four engineers surveyed existing sections to obtain information about these factors and clarifications as to the levels of each factor were provided. The need to locate more sections was realized. The objective was to fill the full factorial (2⁶ or 64 sections) as completely as possible, recognizing that certain sections are impossible to exist or simply do not exist and then to run analyses of variance (ANOVAs), knowing that the factorial is not balanced. The sections used in the factorial for screening the experiments are shown in Figure 2.

			Patched				Unpatched				
			Wide		Narrow		Wide		Narrow		
			Good	Poor	Good	Poor	Good	Poor	Good	Poor	
White	Fine	Urban					B4, B17,				
		Rural					B6, B7, B16				
	Coarse	Urban		B24, B15			B2, B8	G1, B20			
		Rural	B12, B13				B3, B11	B14			
Black	Fine	Urban	A7, A20	A6, A25	A3	A9	5, B19	A11	A5		
		Rural		G2	A8	40, A10	7, 33	A4, A15	A14	33, 44	
	Coarse	Urban	A24, C1	A1	A23	A13	19, 23, B18	37, A2	A21	15	
		Rural	A22	A18, B1	39, A16	6, B22	36, 9	A12, A19	8, 38	2, 3	

FIGURE 2 Sections used in the factorial for screening experiments.

MAIN RATING SESSIONS

The factors considered for analysis were

1. Rater profession,
2. Function in car,
3. Vehicle wheelbase,
4. Time of day,
5. Rater fatigue, and
6. Vehicle speed.

Selection of the Rating Panel

The selection of the panel was dictated by the strictest considerations as follows:

1. Panelists should represent the typical Texas traveling public,
2. Panelists should have a wide range of highway travel experience, and
3. Panelists should have no undesirable (biased) attitudes toward road travel in general.

Members of the rating panel consisted of personnel from the State Department of Highways and Public Transportation, the Center for Transportation Research (CTR), and volunteers from the general public.

The maximum size of the rating panel was 20, with 15 raters (5 vehicles, 3 to a car) and 5 drivers. The rating panel included both men and women of different ages with a wide range of driving and riding experience.

Selection of Vehicles

Two types of vehicles were selected to study the effect of vehicle wheelbase length: a subcompact model (Plymouth Horizon) and a mid-sized model (Mercury Zephyr and Ford Fairmont). These vehicles were taken as representative of typical vehicles owned and operated by an average middle class Texan. Two subcompacts and three mid-sized vehicles were used in the main rating sessions. Equal wear and tear of the vehicles (within each size category) was considered in their selection.

Selection of Sections

For the main rating sessions, the specific combination of characteristics in a section was not so important as the range of roughness of the section itself. The idea here is that to be able to predict serviceability indices from different kinds of roughness characteristics, it is essential to incorporate sections with these (and all other possible) characteristics. One important need, therefore, was to obtain as wide a range of roughness (serviceability indices) as possible. Because the existing pavement sections did not span the roughness spectrum, it became necessary to launch a search. With this objective in mind, roads in eight counties were surveyed. This search resulted in the location of 100 sections in all 8 counties, 77 of which were flexible and 23 of which were rigid pavement sections.

Rating Sessions

The rating method employed was similar to that used in the screening experiments, except for some improvements. The instructions to the raters were revised, based on experience from the screening sessions as to words or cues that were obfuscating and

that raised a number of questions. The same instructions given to drivers in the screening sessions were given to drivers in the main rating sessions. Of course, because the drivers were required to rate the sections according to the experimental plan, they were also required to attend the training session.

A major enhancement in the training session was the use of videotaped instructions. By using videotaped instructions, the researcher hoped to achieve higher reliability through consistency and standardization. Again, the script for the videotape had to be carefully drafted to ensure the minimization of miscues and vague definitions. This technique was also used to alleviate some of the possible errors arising from scale construction such as anchoring effect, and so forth. After the classroom briefing, the panelists were driven over some of the sections. This orientation session was more extensive than the one in the screening experiments. The same rating form (see Figure 3) was used, except for the "age" column.

To analyze the factors vehicle speed and rater fatigue, seven sections with two levels of roughness were chosen, and runs were made corresponding to the levels of the variables. (The "tired" level of the variable rater fatigue corresponded to runs in which raters were at a continuous rating stretch of more than 2 hr, whereas the "fresh" level corresponded to less than 1.5 hr.) Rater profession and vehicle wheelbase could also be included in the factorial. To examine the effect of vehicle wheelbase, raters in the subcompact cars were switched to the mid-sized cars (and vice versa) and thus rated nine sections in both cars. Eight sections were chosen to be rated both in the morning and afternoon so that the effect of time (morning versus afternoon) could be investigated.

Of the 171 sections that were rated, 129 were flexible and 42 were rigid pavement sections. A maximum panel of 20 raters rated these sections in 5 vehicles over a period of 13 days.

Roughness Measurement

The profilometer was set for normal operating conditions, specifically the following:

1. Accelerometer filter wavelength: 200 ft,
2. Sampling frequency: 6.00 in.,
3. Profiling distance: 0.2-mi sections, and
4. Profiling speed: 20 mph.

Roughness measurements were also made using the Mays meter and the SIometer. Runs were made after proper calibration procedures were followed and under normal specified operating conditions.

ANALYTICAL PROCEDURES

Individual rater performances were examined by plotting the mean individual ratings against the mean panel ratings for each rater (Figure 4 shows a typical plot). The mean individual rating represents the average rating (mean over runs) for each section for that rater. The mean panel rating (PSR) is obtained by taking the mean of all the mean individual ratings of all the raters for a particular section. Thus each point on a rater performance graph corresponds to a test section wherein the vertical axis value represents the individual's mean rating for that section and the horizontal axis value represents the mean rating of the panel as a group (for that same section).

SERVICEABILITY RATING CARD

Test Section _____ Date _____ Time _____

Rating Scale

Very Good	Acceptable on the (a) Interstate system <table border="1"> <tr><td>Yes</td></tr> <tr><td>Undecided</td></tr> <tr><td>No</td></tr> </table> (b) Secondary system <table border="1"> <tr><td>Yes</td></tr> <tr><td>Undecided</td></tr> <tr><td>No</td></tr> </table>	Yes	Undecided	No	Yes	Undecided	No	Car/Make Model _____
Yes								
Undecided								
No								
Yes								
Undecided								
No								
Good								
Fair		Position in the car: DR RF LR RR						
Poor		Name _____						
Very Poor		Age _____						

Comments: _____

FIGURE 3 Rating form used in the screening sessions.

Each graph gives an idea as to how each rater performed in comparison to the group. It is not necessary that all points lie on the equality line, but at the same time a point with a large deviation does indicate that for that section, that rater was at variance with the rest of the group for some reason, mostly subjective differences in perception and judgment.

Careful examination of each of the rater performance plots was made to check for discrepancies or abnormalities. No extreme outliers were singled out although it was noted that some of the raters differed with the panel as a group. For instance, one rater may generally tend to rate most pavements better than the others, but then, that is quite reasonable within the limits of acceptable subjective variation. If, however, this variation was found to be of a consistently high order, then the inclusion of this rater in the panel would be reviewed.

In order to analyze the data, mixed model, nested analyses of variance procedures were used. This technique allowed for the testing of hypotheses about the significant differences of means of various variables. The analysis of rating and pavement-related variables was performed using the generalized linear model (GLM) procedure available in the Statistical Analysis System (SAS). Two levels of roughness were chosen corresponding to sections that had PSRs greater than 2.5 and less than 2.5. The main effects of these variables on rating are given in Table 1. These factors are tested against sections S(G) (sections nested within roughness).

Measurements using the Model 690D surface dynamics profilometer are recorded as a road profile, which provides a complete signature of the road surface. In a previous study (5), a profile summary statistic termed root mean square-acceleration (RMSVA) that simulated the response of a typical Mays meter was developed. As a set of indices, RMSVAs can reveal many of the characteristics associated with road roughness.

Thus, from the left and right wheelpath profiles obtained from the operation of the new profilometer, RMSVA values were computed and the left and right wheelpath RMSVAs for each baselength were averaged. For each section, the indices were computed for baselengths of 0.5, 1, 2, 4, 8, 16, 32, 64, and 128 ft.

In addition to the homogeneity and normality, sometimes the question of additivity arises, and in keeping with a sound statistical approach, a nonadditivity test was performed. Tukey's test for nonadditivity (6) was used for this purpose on the main rating data. The interaction effect turned out to be insignificant, indicating that the effect of raters and sections is not multiplicative and that there is no indication that the data need to be transformed.

Multiple regression models were used to relate the profile summary statistics to the panel ratings, and a rigorous statistical procedure (7) was used to select the best candidate in each case.

In selecting prediction equations obtained through standard least-squares regressions, the following criteria were employed:

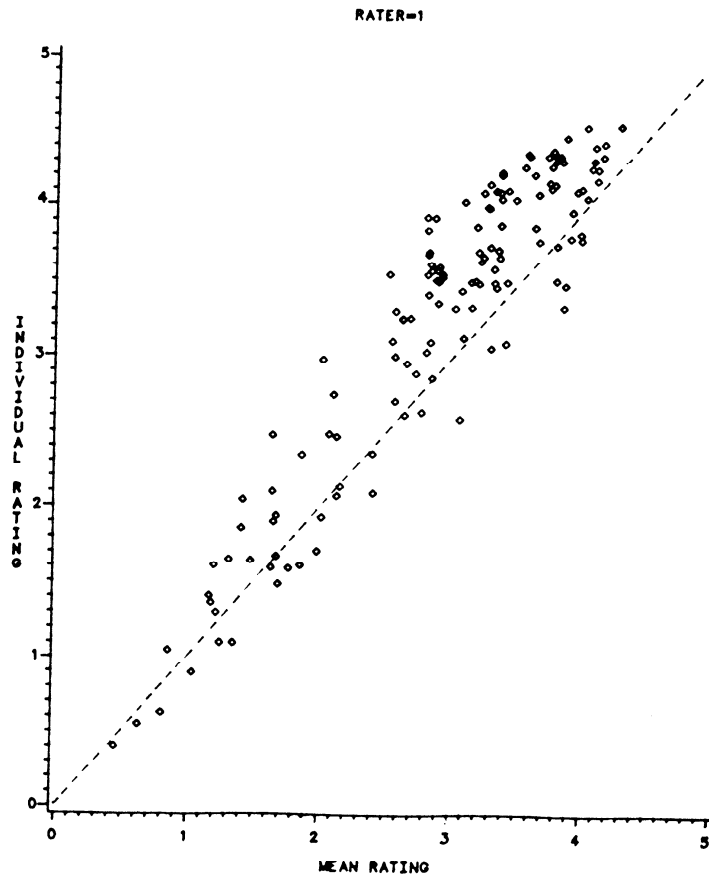


FIGURE 4 Individual rater performance, Rater 1.

TABLE 1 Results of Analysis of Rating Variables

Variable	Effect on Rating
Position in car	Not significant
Rater's sex	Not significant
Rater's age	Not significant
Time (night-day)	Not significant
Rater's profession	Not significant
Function in car	Not significant
Vehicle speed	Not significant
Time (a.m. - p.m.)	Not significant
Vehicle wheelbase length	Significant
Vehicle size	Significant
Rater fatigue	Significant
Pavement type	Significant
Maintenance	Significant
Surface Texture	Not significant
Location of road	Not significant
Road width	Not significant
Surroundings	Not significant

1. The value of R^2 achieved by the least-squares fit;
2. The Mallows' C_p statistic; and
3. The values of s^2 , the residual mean square.

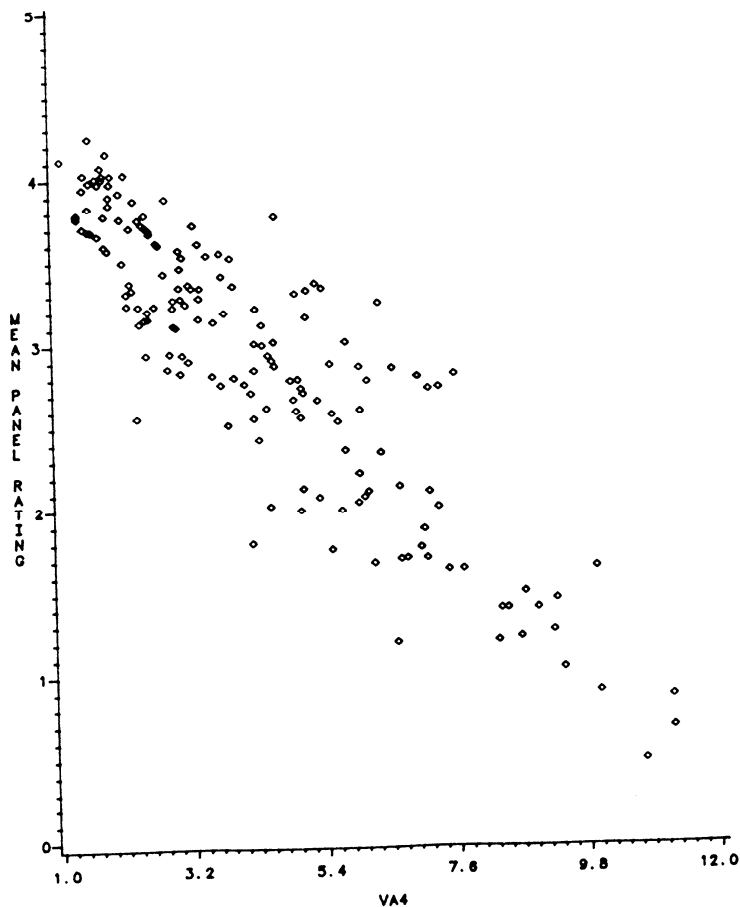
The residual sum of squares can be broken up into lack of fit and pure error sum of squares. The C_p statistic is a reflection of the adequacy of the model, that is, whether or not the model is biased. Equations with considerable lack of fit will show up significantly above or below the $C_p = p$ line on a C_p versus p (number of terms in the model) plot.

The residual mean square, s^2 , provides an estimate of the variance about the regression, which is presumed to be a reliable unbiased estimate of the error variance. For this study, this presumption is valid, considering the large number of degrees of freedom. This procedure alleviates the problem of unreliable inferences that result from using stepwise regression. Before running the regressions, the individual regressor variables were plotted against the PSRs. Figure 5 shows the relationship between the mean panel ratings and one of the independent variables, VA_4 . Overall, it may be observed that as the RMSVA increases, the rating decreases; here again, the physical meaning of the RMSVA concept is manifested in that, with higher amplitudes, the degree of discomfort (or roughness) as perceived by the user is greater.

Linear multiple regression analyses were performed on the data with the mean panel ratings as the independent variable and the RMSVAs (average of the mean left and right wheelpath values) in the base-lengths of 0.5, 1, 2, 4, 8, 16, 32, 64, and 128 ft as dependent (regression) variables. Separate analyses were carried out as follows:

1. Overall data,
2. Overall data with a forced dummy variable for pavement type,
3. Flexible sections, and
4. Rigid sections.

For each analysis, all possible numbers and combinations of terms were included in the regression

FIGURE 5 PSR versus VA₄.

models and R^2 and C_p values were generated for each model. From this list, the best model was selected based on the R^2 , C_p , and s^2 criteria. Diagnostic checks were made on each of the selected equations to verify that the assumptions of regression were fulfilled satisfactorily. Plots of the predicted values and the residuals were examined and the normality of errors was checked using the W-statistic or the D-statistic, as appropriate.

FINDINGS AND CONCLUSIONS

Based on the controlled experiment designs discussed earlier in this paper, the main effects of the variables associated with the rating process were found to be significant or not significant at the 0.01 α -level as shown in Table 1.

The conclusion is that the position of the rater in the car (whether in the front or rear) does not influence the rating. Similarly, it can be concluded that whether the rater is male or female, young or old, riding in the vehicle or driving the vehicle, or is a technically experienced person or not, has no effect on his or her rating; also, whether the rating is done during any particular time of day has no effect on the rating. These findings support the relationship between roughness (as manifested through the road surface and vehicle characteristics) and the rating of ride quality. The finding that vehicle speed has no effect on rating appears contrary to

expectation; however, it should be expected that the interaction between vehicle speed and road roughness would be significant. The conclusion here is that the rater's receptor system adjusted for the range of levels considered (30 mph versus 50 mph) in such a way that there was no significant difference in his or her ratings.

The variables that indicated a significant effect on rating at the 0.01 α -level were vehicle wheelbase, vehicle size, and rater fatigue. The effect of different vehicle characteristics on the perceptions of ride quality is exemplified here. It was found that raters expressed lower ratings (as much as 1.5 serviceability units) while riding in short wheelbase vehicles compared with longer wheelbase vehicles. The role played by vehicle characteristics in the rating process has been shown by the significance of the effect of vehicle size on rating. The significance of rater fatigue as a variable demonstrated the sensibility of the rater vis-à-vis the condition of his or her receptor system.

Two pavement-related variables, pavement type and maintenance, were found to have significant effects at the 0.01-level, whereas surface texture, location of road, road width, and surroundings had no significant effect on ratings. Using regression analyses, a set of serviceability prediction equations was developed. The best formulas obtained are listed next (PSI refers to present serviceability index and VA_b is the measure of root-mean-square vertical acceleration associated with baselength b , ft).

Overall (167 sections):

$$\text{PSI} = 4.42 + 1.55 \cdot 10^{-3} \text{VA}_{0.5} - 0.311 \text{VA}_4 - 3.35 \text{VA}_{64}$$

with $R^2 = 0.86$, $s^2 = 0.10$.

Overall (with dummy variable PTYPE):

$$\text{PSI} = 4.31 + 0.039 \text{VA}_2 - 0.504 \text{VA}_8 - 822 \text{VA}_{128} + 0.366 \text{PTYPE}$$

with $R^2 = 0.88$, $s^2 = 0.09$.

Flexible (125 sections):

$$\text{PSI} = 4.43 - 0.016 \text{VA}_2 - 0.237 \text{VA}_4 - 0.4 \text{VA}_{32} - 10.4 \text{VA}_{128}$$

with $R^2 = 0.89$, $s^2 = 0.10$.

$$\text{PSI} = 5.00 - 0.0029 \text{VA}_{0.5} - 0.2609 \text{VA}_4 - 5.006 \text{VA}_{64}$$

with $R^2 = 0.82$, $s^2 = 0.15$.

Rigid (42 sections):

$$\text{PSI} = 4.34 - 0.092 \text{VA}_4 - 0.47 \text{VA}_8$$

with $R^2 = 0.73$, $s^2 = 0.03$.

The Mays meter data and the mean panel ratings showed good correlation for flexible sections (correlation coefficient $r = -0.91$). The correlation coefficient for rigid sections was found to be much lower ($r = -0.513$). Regression analysis on the overall sections indicated an R^2 value of 0.79, compared with 0.859 for the profilometer [refer to Nair (7) for prediction equations]. From this it was concluded that

1. The Mays meter can predict PSR better on flexible sections (with an R^2 of 0.83) than on rigid sections (R^2 of 0.26),
2. The SIometer can predict PSR better on flexible sections (with an R^2 of 0.56) than on rigid sections (R^2 of 0.11), and
3. The 690D SDP is by far the best overall predictor of PSR.

For all of the foregoing discussions, it should be remembered that Mays meter and SIometer data correlations have been obtained only after properly calibrating these devices.

The major development of this research study is a set of equations relating the ride quality of pavement sections to pavement roughness. This was achieved by relating a set of roughness summary statistics (RMSVAs associated with different wavelengths) obtained from the pavement profiles to the mean panel ratings. The study showed that up to 88 percent of the variation in PSR can be explained by the roughness variables; this a very high degree of linear association (a correlation coefficient of -0.94) between PSR and roughness as characterized by the set of RMSVAs. Thus, this study further attests to the serviceability-performance (S-P) concept in general, and to the validity of using road profile measurements to predict PSRs and to obtain indices of serviceability in particular.

ACKNOWLEDGMENTS

The authors would like to express appreciation for the cooperative efforts of Gary Graham, the Texas SDHPT representative. Also appreciated is the staff of SDHPT's D-10 Research Technical Services for helping furnish Mays meter and SIometer data for the study. The authors are also pleased to acknowledge

the combined efforts and support of the Center for Transportation Research at the University of Texas, Austin, and the Texas SDHPT in cooperation with the Federal Highway Administration, U.S. Department of Transportation.

REFERENCES

1. The AASHTO Road Test: Report 5, Pavement Research. HRB Special Report 61-E, HRB, National Research Council, Washington, D.C., 1962.
2. W.R. Hudson. High-Speed Road Profile Equipment Evaluation. Research Report 73-1. Center for Highway Research, University of Texas, Austin, Jan. 1966.
3. Standard Definition of Terms Relating to Traveled Surface Characteristics. ASTM E867. American Society for Testing and Materials, Warrendale, Pa., 1982.
4. F.L. Roberts and W.R. Hudson. Pavement Serviceability Equations Using the Surface Dynamics Profilometer. Research Report 73-3. Center for Highway Research, University of Texas, Austin, April 1970.
5. D.W. McKenzie, W.R. Hudson, and C.E. Lee. The Use of Road Profile Statistics for Maysmeter Calibration. Research Report 251-1. Center for Transportation Research, University of Texas, Austin, Aug. 1982.
6. V.L. Anderson and R.A. McLean. Design of Experiments. Marcel Dekker, Inc., New York, 1974.
7. S.K. Nair, W.R. Hudson, and C.E. Lee. Development of Realistic and Up-to-Date Pavement Serviceability Equations Using the New 690D Surface Dynamics Profilometer. Research Report 354-1F. Center for Transportation Research, University of Texas, Austin, Aug. 1985.

The contents of this paper reflect the views of the authors, who are responsible for the facts and the accuracy of the data presented herein. The contents do not necessarily reflect the official views or policies of the Federal Highway Administration. This paper does not constitute a standard, specification, or regulation.

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

Discussion

R. M. Weed*

This discussion pertains to both the paper and the project report (1) summarized in the paper. The authors have done many things extremely well and, in several ways, have illustrated how a thorough statistical analysis should be performed. However, there is one particular area in which a further refinement may be desirable. This involves the use of multiple linear regression and a perplexing result that obviously was of concern to the authors. After commenting on certain aspects of mathematical modeling,

*New Jersey Department of Transportation, 1035 Parkway Avenue, Trenton, N.J. 08625

I would like to suggest what might be a more appropriate theoretical model.

The primary goal is to find the mathematical model that most accurately describes the process being investigated. To this end, it is appropriate to use every resource available, including any prior knowledge of that process. In the case of pavement serviceability rating (PSR, the Y data) versus vertical acceleration (VA, the X data), it can be reasoned that if a pavement were so smooth that no vertical acceleration could be detected, it would be appropriate to rate it at a value of PSR = 5. At the other extreme, there is essentially no difference between a very high VA value and a slightly higher VA value; for all practical purposes both would correspond to PSR = 0. This suggests that the appropriate mathematical model will originate at (VA = 0, PSR = 5) and decline in some way to eventually become asymptotic to the X axis. (This assumes, of course, that the roughness-measuring device is sufficiently responsive to justify such a relationship. This must be confirmed by an examination of the data. If the device is not sufficiently responsive, it is unlikely that any mathematical model would be very useful, and the use of a different device would be indicated.)

A further consideration is a philosophical one. It is believed by many analysts that known prior knowledge (the engineering constraints on the intercept and the asymptote in this case) should take precedence over empirical statistical measures. In other words, if it were believed that a model of a particular form was fundamentally correct, it would be chosen in favor of a competing model that happened to have a higher correlation coefficient. The rationale is that, while the competing model might appear to have greater predictive power for this particular data set or range of data, the more theoretically appropriate model is likely to perform better in the long run (especially if, as so often happens with published research findings, it should be used outside the range of data from which it was generated).

A basic exponential decay function that is capable of satisfying the constraints on the intercept and the asymptote is given by Equation 1 in which A and B are constants and e is the base of natural logarithms. Other similar forms may also be used. If several different wavelengths of VA must be accounted for, then Bx in this expression would be replaced by a function f(x). Figure 6 illustrates three general shapes that these functions can take.

$$Y = Ae^{-Bx}$$

(1)

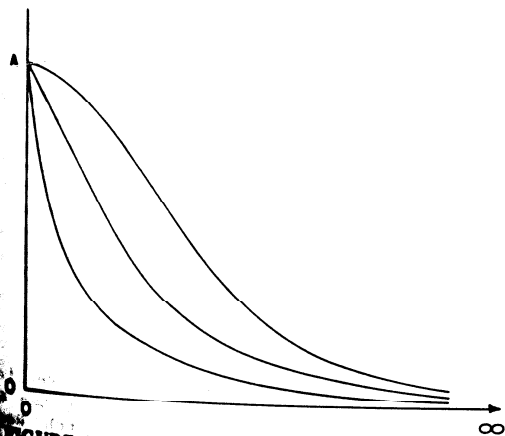


FIGURE 6 Examples of exponential decay curves.

If possible, because of the collinearity problem that the authors have duly noted (1,p.115), it might be both practical and desirable to choose a single wavelength on which to base the equation. All the data plots (1,pp.116-124) exhibit the same general trend and PSR versus VA_8 , for example, would appear to have the most uniform variability about a fitted line.

The decision to use a linear model was apparently made by visual inspection (1,p.115) of the data plots. However, in most of the plots presented in the report, it is very easy to visualize how the exponential forms in Figure 6 might more appropriately be fitted. (It should be noted that failure to locate the Y axis at VA = 0 in these figures tends to mask this effect.) One of the figures, PSR versus VA_{16} , is reproduced here as Figure 7 to illustrate an approximate exponential fit. (Actual exponential fits obtained by least squares with similar PSR data may be seen in the discussion of the paper by Janoff elsewhere in this Record.)

It is always disconcerting whenever an analysis produces a result that is inconsistent with a (presumed) known fact. The authors are appropriately concerned about the positive coefficient for $VA_{0.5}$ in the overall equation for PSI, reproduced here as Equation 2. This implies that, if VA_4 and VA_{64} were held constant, the PSI improves as $VA_{0.5}$ increases, an obviously incorrect conclusion as they have noted. Although the authors state that there is no evidence of incorrect specification of the model, this result in itself may be an indication that the chosen form of the model (linear) is inappropriate. Their attempt to obtain a better model was unsuccessful, most likely because only linear models were considered.

$$PSI = 4.42 + 0.00155 VA_{0.5} - 0.311 VA_4 - 3.35 VA_{64} \quad (2)$$

The authors' investigation of a model with the intercept forced to be at PSI = 5.0 was certainly sound and logical but, unfortunately, had little chance for success unless it was also recognized that an exponential model was needed. In the form given by Equation 1, for example, this would involve setting A = 5 and determining the coefficient B by regression.

Still another troublesome factor may come into play. Figure 8 shows an example illustrating how data conforming to the same fundamental exponential relationship can produce two distinctly different equations when linear regression is used. If flexible pavements tend to be smoother than rigid pavements (as is the case in the state of New Jersey), they lie farther up on the exponential curve. If analyzed by linear regression, they produce a steeper slope than that obtained from the data representing rigid pavements. In this case, depending on where these lines cross, either could be falsely perceived to serve better than the other. An effect similar to this may also be present in the Texas data.

A final point is very speculative. Roughness measurements produced by wavelengths that are similar in length are stated to be correlated (1,p.115), resulting in a collinearity problem among the several independent variables of the multiple regression. Because the measurements in question deal with a vibrational phenomenon, there may also be some sort of harmonic relationship between wavelengths that are integral multiples of each other. It is possible that the positive coefficient in Equation 2 is the result of correlation among the measurements obtained from different wavelengths. If it is absolutely necessary to include more than one wavelength, perhaps

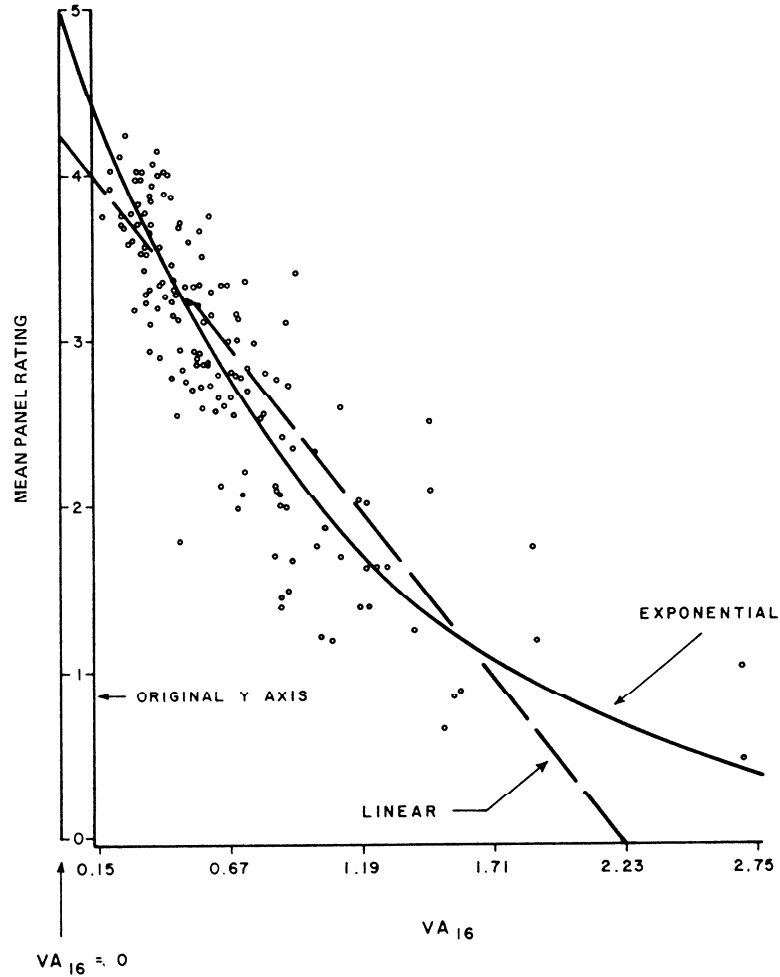


FIGURE 7 Approximate linear and exponential fits.

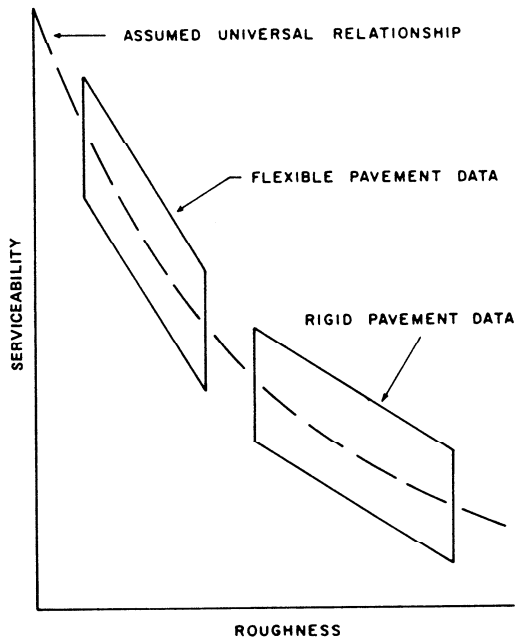


FIGURE 8 Conceptual illustration of a condition that can produce different linear regression lines.

it would be better to choose wavelengths that cannot produce harmonic frequencies.

In view of the apparent problem with a linear model, as well as the possible collinearity problem, it may be inappropriate to conclude that rigid pavements are perceived to serve better than flexible pavements (1,p.133). The use of a dummy variable representing pavement type in the PSI equation (1,p.130) may be unnecessary if an exponential model is used. It is also conceivable that some of the other conclusions regarding performance of the various types of equipment might be altered to some extent with an exponential model.

In summary, the authors have conducted a very thorough experiment and have made an important contribution to the understanding of the perception of pavement serviceability. Consideration of exponential models that satisfy the fundamental engineering constraints, and a concerted effort to avoid the collinearity problem, may serve to further enhance their efforts.

REFERENCE

1. S.K. Nair, W.R. Hudson, and C.E. Lee. Realistic Pavement Serviceability Equations Using the 690D Surface Dynamics Profilometer. Research Report 354-1F. Center for Transportation Research, University of Texas, Austin, Aug. 1985.

Authors' Closure

We are most appreciative of Weed's discussion of our paper. In general, we agree with his comments and thank him for his very careful review of our work. It is generally true that every available modeling resource should be used in conceptualizing a problem and we did attempt to do this.

It is not necessarily true that boundary conditions such as proposed of $VA = 0$, $PSR = 5$ will always govern such concepts and it presumes too much knowledge of the problem. In reality, a $VA = 0$ may not, for example, be attainable on a pavement nor may it be considered perfect by the average rater. Furthermore, any time individual raters are forced to use a scale that is bounded, such as 0 to 5, it is almost precluded that the average of a set of 10 or more raters can ever equal 5.0. For these reasons, it is difficult, in our opinion, to consider forcing the model through an origin of $VA = 0$, $PSR = 5.0$.

We concur wholeheartedly that the exponential form should be considered. We chose to consider and use the basic linear form because we were replacing

an existing equation with the linear form, and the sponsors desired minimum acceptable change. We do agree, however, that the exponential model is a worthwhile model to examine, and we will examine that possibility as time permits. There is not sufficient time, however, to do so for this closure.

Review of the relationship between vertical acceleration and serviceability rating leads to the conclusion that several wavelengths are necessary to fulfill the correlation. Therefore, Equation 1 as proposed by Weed will involve a more complicated function of X .

The comments about the relative smoothness of rigid and flexible pavements in the state of New Jersey are not applicable to Texas pavements. In general, Texas pavements in the study exhibited similar roughness ranges for rigid and flexible. We will certainly attempt to further investigate the concept of nonharmonic wavelengths as further work is permitted.

In summary, we greatly appreciate Weed's contribution and his thoughtful review and ideas in extending this work. We will certainly take them into account as additional work progresses.