Description and Evaluation of Alaska's Pavement Rating Procedure

ROBERT L. McHATTIE AND BILLY G. CONNOR

Pavement condition rating methods used on Alaska's roadways since 1978 are described and examined. The methods are intended to provide the specific performance data necessary to optimize construction and maintenance planning and the allocation of available funds. Rating elements include simplified measurements of ride roughness, fatigue (alligator) cracking, patching, and rut depth. These features are reported individually and are also combined with traffic data to indicate more general levels of roadway serviceability. Field evidence shows that a high degree of variability exists in the measurement of cracking, patching, and rutting. Coefficients of variation above 20 percent were estimated for each type of rating element from experimentally repeated measurements. On a given road section estimates of fatigue cracking made by 15 crews differed by up to twice the calculated average. Rut depth measurements were typified by calculated standard deviations of about half the mean value. Report findings suggest that great care be exercised on future pavement performance inventories. Standardization techniques are suggested that should improve manual rating methods. Mechanized or electronic data acquisition techniques must be developed to eliminate human error.

The Alaska Department of Transportation and Public Facilities (ADOTPF) initiated use of newly developed pavement rating procedures during its 1978 highway inventory. The purpose of this study is to evaluate the statistical validity of individual measurements that comprise it. The current Alaskan rating attempts to quantify surface fatigue cracking, patching, and wheelpath rutting as an aid to planning design, construction, and maintenance. The amount of error associated with measurements of pavement distress is examined, and improvements are suggested that can be incorporated into future inventory work.

The research data base used consisted of data and experience accumulated from two complete inventories of the Alaskan paved highway system conducted during 1979-1981. The study also examines results of repetitive sampling conducted specifically for this project on five typical pavement sections located near Fairbanks, Alaska.

DEVELOPING ALASKAN PAVEMENT RATING PHILOSOPHY

During the winter of 1977-1978 the planning division of ADOTPF decided to revise its existing highway inventory procedure to fill the need for accurate, quantitative data for programming highway maintenance and construction funds. The department's research section was commissioned to produce a practical inventory that would stand the scrutiny of statistical evaluation.

As a first step, the literature was researched to see how other states and foreign transportation agencies had negotiated the same ground. A method for rating pavements was first developed for use in the AASHO Road Test of the late 1950s to early 1960s. Pavements are classified numerically based on the subjective observations of engineering specialists and normal highway users (1). The rating scale was arbitrarily set between 0 and 5, where 0 is extremely poor and 5 is perfect. The key distress manifestations selected are surface deterioration, ride roughness, rutting, cracking, and maintenance patching. This rating technique produces a number termed present serviceability rating (PSR) for classifying a given section of road. Figure 1 (2) indicates the number of individual raters necessary, statistically, to estimate the true value of PSR by using the completely subjective AASHO method. This figure indicates that for one or two raters the error associated with estimation of PSR is greater than 1. The error can range ±1 from the true value; therefore, the full range of possible estimation is two, which represents one-third of the total 0 to 5 scale.

The AASHO researchers then took the next logical step of converting the rating from a subjective to an objective method by deriving a regression equation that closely matches PSR panel scores. Independent variables for the regression equation consisted of standardized measurements of fatigue cracking area, maintenance patch area, wheelpath rut depth, and longitudinal surface variation (roughness). The road surface condition values calculated by the regression equation are termed present serviceability index (PSI).

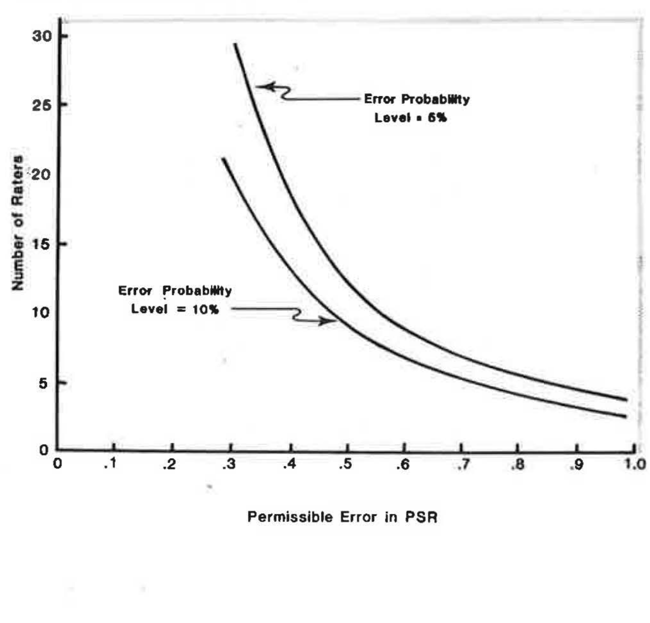
$$PSI = 5.03 - 1.91 \log(1 + SV) - 1.38RD^2 - 0.01(C + P) \quad (1)$$

where

SV = mean slope variance in the two wheelpaths as measured absolutely by a longitudinal profilometer (in./mile x 10^6),
RD = mean rut depth (in.),
C + P = cracking + patching (ft^2/1,000 ft^2 total surface).

Most pavement rating methods developed subsequent to the AASHO study, including Alaska's, are related in some degree to the original AASHO form and were intended to provide key performance feedback to the overall pavement management process. Generation of Alaska's rating scheme was expedited by a summary and critique of highway agency pavement management practices. A federally sponsored workshop was held in Tumwater, Washington, in November 1977 to examine the existing state of the art in the field of pavement management systems (PMS). United States and

Figure 1. Estimating PSR.



Canadian representatives were invited providing they were actively implementing and, therefore, experienced in a PMS program. At the time ADOTPF was attempting to devise a rating method for asphalt concrete pavements, the Tumwater conference report was by far the most comprehensive source of information concerning rating schemes available (3). The Tumwater report not only discussed various field methods but also compared them critically. Rating system elements were suggested that provided the best input to the overall PMS.

Assuming that PMS would be the ultimately intended use of pavement inventory data, the following consensus emerged from the Tumwater conference:

1. Ride roughness should be rated objectively.
2. Structural capacity should be rated, but whether to rate structural capacity on the basis of deflection tests or surface distress measurements was not clearly decided.
3. Pavement distress should be rated. This includes measurement of rut depth, cracking, and patching.
4. Rut depth measurements were considered along with skid testing to provide an indication of road safety. Rut measurements should, therefore, be included in any highway rating scheme.
5. The use of a single classification number such as PSI was said to provide a valid measure of pavement condition.
6. Little standardization of terminology and measurement technique exists among the available systems of pavement rating when these systems are examined in detail.

Each of the preceding points was considered seriously before development of the Alaskan rating system. Table 1 (3) indicates the salient features of the road rating methods used by the U.S. states and Canadian provinces represented at the Tumwater conference.

The objectives and basic rating elements listed below were chosen by ADOTPF from background research and a definition of departmental needs. They guided the development of Alaska's inventory rating by providing use targets. Only the most commonly recognized pavement condition indicators were selected for consideration as elements in Alaska's rating procedure.

ADOTPF decided that a pavement condition (rating) must

1. Provide information for planning and ordering the priority of rehabilitative design and maintenance of existing pavements,
2. Provide information on the relative condition of total highway mileage within various jurisdictions for budgetary apportionment purposes, and
3. Provide design feedback information.

Table 1. Pavement monitoring features and evaluation.

Agency	Surface Condition	Roughness of Ride	Skid Resistance	Structural Capacity	Rating System	Primary Decision Criteria
Arizona	Crack survey	Mays ridemeter on annual basis	Mu meter-500 ft at each mile post	Dynaflect-3 locations/mile	Pavement management information system	Compares major maintenance alternatives
California	Pavement condition survey based on alphanumeric rating	Ride score not part of pavement distress	Measured periodically		Alphanumeric rating combines severity and extent of defects	Defects compared to repair strategies and costs
Florida	Structural defects of cracking, rutting, and patching	Mays ridemeter correlated with CHLOE profilometer			Combined ride rating and defect rating	Adjusted pavement rating evaluated for priority programming
Kentucky	Used as feedback for design deficiencies	Roughness index correlated to PSI. Use ride-quality meter or GM profilometer		Road rater for specific design evaluation	Correlation of several factors for design input	Input used to develop overlay design
New York		Vehicle response profiler is heart of system			Pavement serviceability system, based on correlation with known serviceability levels	Aimed at identifying budget needs, failed pavements, effectiveness of expenditures
Pennsylvania		Mays ridemeter used to develop serviceability	ASTM skid trailer	Road rater		
Texas	Distress survey based on vehicle-mounted camera-visual distress rating	Mays ridemeter correlated with Surface Dynamics profilometer	Skid trailer	Dynaflect for critical locations	Relative design, ratio of allowable 18K axle loads to those predicted for next 20 years	All highways must carry their traffic safely and comfortably
Utah	Pavement distress based on 11 observed parameters	PCA roadmeter on 1-mile increments	Mu meter, 0.5-mile sections tested every 2 miles	Dynaflect for predicting remaining life	Present serviceability index	Overall priority ranking for preventative rehabilitation
Washington[a]	Pavement condition survey every 2 years covering entire network	PCA roadmeter on all sections	ASTM skid trailer for high accident locations-considered separately	Limited use of Benkelman beam	Combined structural rating and ride score	Tabulate rehabilitation strategies and costs based on pavement condition
Ontario	Pavement condition rating, 1 to 2 year cycle	Subjective riding comfort index		Dynaflect, random sample locations in need of rehabilitation	Subjective, pavement condition rating	Required overlay prediction based on expected performance
Saskatchewan	Annual surface condition rating	PCA roadmeter at intervals of 1 month to 1 year		Benkelman beam data used for overlay design	Condition rating system used to order priority of projects for overlay or scaling	Preventative maintenance is primary goal

[a]Photo logging of entire system.

Literature review plus common sense pointed to the need for a rating method that would characterize the road condition adequately and allow a high degree of reproducibility at a minimum cost. The data must provide true reproducible characterization of pavement condition that changes from year to year in a rational manner (i.e., pavements should not appear to heal anomalously with time unless maintenance has actually been done). The rating technique, therefore, had to be as simple as possible and include the largest practical sampling of each road section.

The following were chosen as rating parameters by Alaskan researchers:

1. Fatigue cracking (alligator cracking),
2. Major patching (at least full lane width),
3. Wheelpath rut depth, and
4. Ride roughness as measured by the Mays ridemeter.

Fatigue cracking was selected as a rating parameter because it is an excellent indicator of structural condition and load-life potential. Design-life vehicle load capacity is said to be reached when significant alligator cracking becomes apparent. Fatigue cracking is also often associated with unacceptable rutting, rough rides for the vehicle, and desintegration of the pavement surface.

Major patching needed to repair a host of problems, including fatigue cracking, embankment settlement, and rutting, gives a general picture of the maintenance effort required on a given road section. Patching is also a principal source of surface roughness and usually becomes cracked and potholed with time.

Wheelpath rutting is generally considered important in terms of driver safety and travel costs. A consensus of available literature indicated that rutting deeper than approximately 0.5 in. is a safety hazard that can cause hydroplaning on wet road surfaces at high vehicle speeds. Rutting also has an effect on vehicle steering and reduces the mechanical life of chassis components. Deep rutting usually accompanies advanced alligator cracking and signifies that pavement structural soil layers (base or subbase) have been loaded beyond capacity. This condition is aggravated through use of materials subject to extensive moisture-related softening (thaw weakening).

Ride roughness is measured because it is the characteristic of pavement that is of primary concern to the driving public. The combination of differential settlement and leveling patches is common to all parts of Alaska and is the major cause of roughness felt by the driving public. Ride roughness is measured objectively on a continuous basis by using available technology such as the Mays ridemeter.

Some recognized surface distress features were disregarded in order to simplify the rating process. These include raveling, longitudinal cracks, thermal cracking, shoving and bleeding, potholes, and deflection. Statewide skid measurements in 1975 indicated that the materials used in Alaskan roadbuilding provided consistently high skid numbers. Reasons for this include a high degree of aggregate hardness and limited potential for asphalt bleeding because of Alaska's relatively cool air temperatures.

Testing of deflection statewide will ultimately become part of the normal inventory process. This process began in 1982 and will require approximately five years per statewide cycle. The falling weight deflectometer is currently being used to collect inventory data.

RATING AND SCORING PROCEDURES USED SINCE 1978

A discussion of pavement rating methods used by ADOTPF since 1978 and an explanation of how field data are manipulated for purposes of scoring and reporting follow. Figure 2 illustrates the manner in which raw field data are transformed into a useful pavement inventory report.

Development of Field Methods

The rating process is done as two separate operations, each requiring the use of a two-person crew. Phase 1 consists of measurement of ride quality. The Mays ridemeter trailer is currently being used because of the relatively low cost of gathering data and its reasonably good repeatability. The trailer-mounted meter provides a standardized vehicle, suspension, and tire type. In phase 2 the surface distress features are measured. These include alligator cracking, full-lane patching, and rut depths.

The Mays ridemeter can provide a continuous sampling of highway roughness at 50 mph automatically. Studies of the repeatability of this test have been made by others and are beyond the scope of this report. The objective nature of ridemeter measurements suggests that they be considered a relatively reliable element of the current pavement inventory.

Methods for measuring alligator cracking and major patching were evaluated initially on seven sections of roadway near Fairbanks. Each section was divided into 0.1-mile subsections that were rated independently. Full-width patching was characterized on the basis of total length (density); fatigue cracking was typified by both density and severity. A type-1 or type-2 classification was adopted for cracking of lesser or greater severity. Alligator cracking was defined as cracking that is visible while driving 7 to 10 mph. It is measured as the total percentage of the road section length that exhibits cracking, regardless of wheelpath location. Histograms were constructed from field data (Figure 3) to show the frequency distribution of fatigue cracking for the subsections within each mile. The distribution of cracking is strongly polymodal (showing no single mean value) and bounded to both the 0 and 100 percent occurrence level. Distributions of fatigue cracking are obviously non-Gaussion in character. Based on these data Table 2 gives the probability of a random selection of a 0.1-mile sample that will predict the true mean condition of each section of roadway. The probability is obviously small in all cases. In view of these data, a 1-, 2-, 3-mile or more length of paved road could not be rated accurately for fatigue cracking based on measurements in a randomly selected subsection several hundred feet long. The normal assumption of a 10 to 20 percent sampling density is of no value in this case. Fatigue cracking, therefore, must be measured by continuous observation through each mile of roadway. All data collected subsequent to the initial trial have supported this decision.

Full-width patching was observed to have a distribution of occurrence similar to that of fatigue cracking and it was also decided that this feature could be properly characterized only by continuous observation.

The frequency of rut depth measurement was also examined briefly before development of the rating method through multiple readings taken on each of eight 1-mile-long pavement sections near Fairbanks. Rut depth averages ranged from 0.185 to 0.244 in. The standard deviations of the sample ranged between 16 and 35 percent of the sample means and the plotted frequency distributions of rut depth measurements appeared reasonably indicative of normal

(Gaussian) behavior. Based on these trials the assumption was made that rut depth measurement could be evaluated by normal statistical techniques. Sampling frequency was addressed through the statistical method used for estimating a true mean value from a small sampling. An estimation of true population average is given by

$$\mu_o = \bar{X} \pm ST \sqrt{N} \quad (2)$$

where

μ_o = true population average (i.e., true average rut depth),
X = average rut depth as determined from sample,
S = standard deviation of sample,
N = number of measurements constituting the sample, and
t = Student's t-value for a given confidence level and N.

Figure 2. Elements of Alaskan pavement inventory.

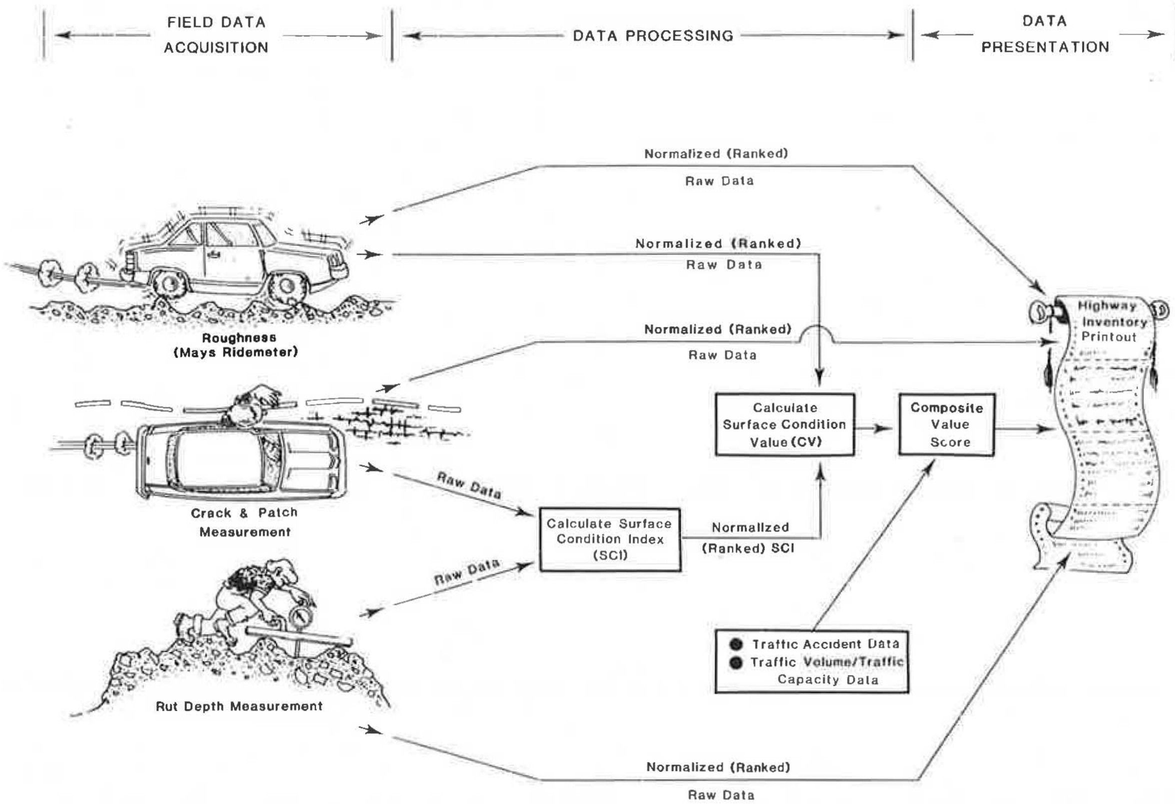
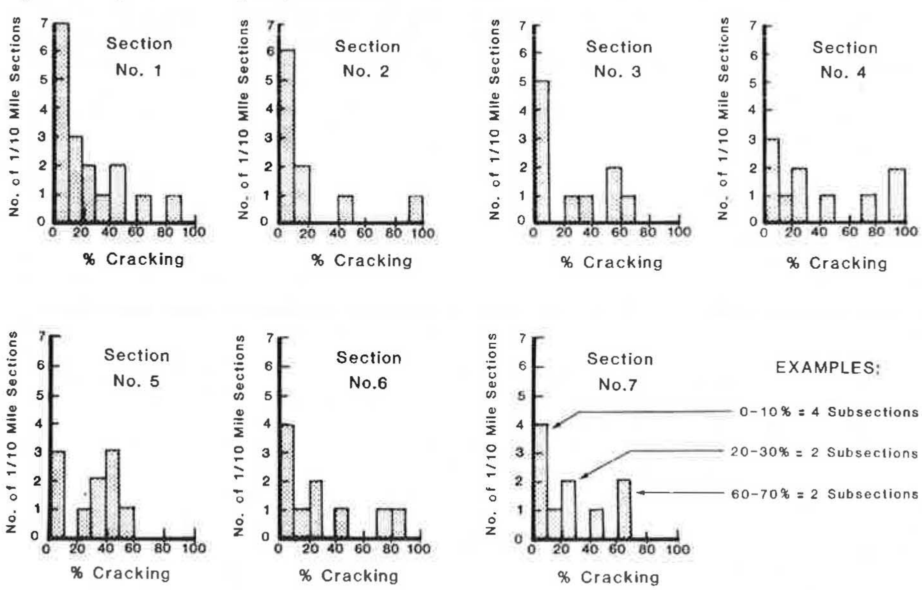


Figure 3. Alligator crack frequency distribution.



Note: Figure shows variation in amount of cracking in subsection measurements within seven sections. Except for section 1, all were 1 mile long and included 10-1/10-mile subsections. Section 1 was 1.7 miles long and included 17-1/10-mile subsections.

This equation is an expression of the central limit theorem, which describes the distribution of sample means about a true mean. In modified form the equation can be expressed as follows:

$$(\mu_o - \overline{X})/S = T\sqrt{N} \quad (3)$$

The error in estimating true rut depth average (i.e., $\mu_o - \overline{X}$) is small in relation to the sample standard deviation (at a given level of confidence) when the term $T/\sqrt{N}$ is minimized. Figure 4 is a plot of N versus $T/\sqrt{N}$ used to select sampling frequency for the initial inventory runs in 1978. Flattening of the curve beginning between N = 4 and N = 7 suggested that a sampling of at least four locations would be necessary to ensure that the error in estimating true mean rut depth would be less than 2 standard deviations of the sample. Figure 4 indicates that the error of estimating true mean rut depth is about 1.6 x S for N = 4. Because S of the trial road sections averaged approximately 0.05 in., errors in estimating rut depth during inventory work would be expected to be no larger than ±1.6 x 0.05 (i.e., 0.08 in.). This accuracy was considered good enough for beginning the pavement inventory process. Fewer than 4 readings per mile were required in the 1978 rating method if rutting was generally observed to be less than 0.25 in.

Summary of Required Measurement Frequencies

On the basis of limited field trials it was decided that pavement distress, except for rut depth measurement, should be characterized by continuous observation of the entire road. Three field seasons of field data collection have reinforced the idea of using 100 percent sampling.

The frequency of measurements necessary to determine average rut depth adequately was calculated from a preliminary statistical assessment. Measurement of ruts was known to be a disproportionately time-consuming job when compared with other distress observations. The hope was that the experience in accumulation would show that no more than four sets of readings would be required per mile of road.

Table 2. Proability of sampling true mean performance.

Section	Percentage of Total Area Cracked	Range Acceptable From	Range Acceptable To	Probability of Selecting 0.1-Mile Section with Correct Percentage of Crack
1	23	13	33	0.29
2	20	10	30	0.20
3	24	14	34	0.10
4	39	29	49	0.10
5	31	21	41	0.20
6	28	18	38	0.30
7	27	17	37	0.20

Road Condition Scoring Using Alaska's Pavement Rating

Alaska uses its pavement inventory data to construct a mile-by-mile summary report listing individual condition scores (percentage of cracking, rut depth, percentage patching, and ride roughness) and also a combined condition value (CV) score. The CV is analogous to the AASHO PSI and provides a single numerical descriptor of a given road section.

The CV is calculated from the inventory data in the following way:

$$CV = [\text{Mays ridemeter score (ranked)} + \text{Surface condition index (ranked)}] \div 2.0 \quad (4)$$

Ranked data indicate that the data have been transformed mathematically into a percent-worse-than score before calculation of CV. This provides a normalizing of raw scores on a 0 to 100 (worst to best) scale.

$$\text{Percentage worse than} = [(1/2E + L)/N] \times 100 \quad (5)$$

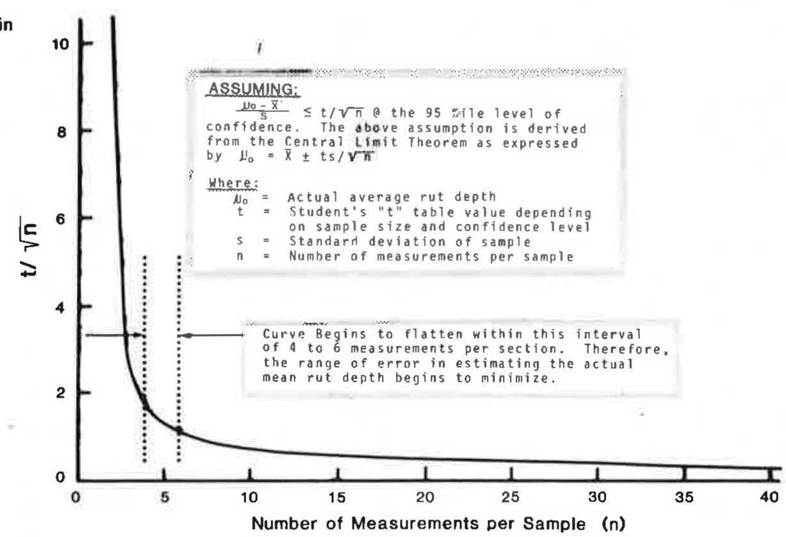
where

E = number of statewide sections rated the same,
L = number of statewide sections rated worse, and
N = total number of statewide sections rated.

The Mays ridemeter score shown in Equation 4 is derived directly through the percentile ranking equation from raw Mays ridemeter data. Surface condition index (SCI) is calculated by means of Equation 6 and then transformed to a percentage-worse-than ranking through the ranking formula.

$$SCI = 1.38R^2 + 0.01(A+P) \quad (6)$$

Figure 4. Plot of relative error factor ($t/\sqrt{n}$) versus number in sample.



where

R = average rut depth (in.),
A = percentage of road section that is alligator cracked, and
P = percentage of road section that is covered by full-width patching.

In addition to reporting a summary of the previously discussed information, the pavement inventory report also includes, for multiple mile sections, the ranked scores of a volume/capacity ratio and the section's accident rating value. Finally, the condition value plus capacity and accident scores are combined in the form of a geometric mean to produce a composite value calculated as

$$\text{Composite value score} = [\text{Condition value} \times \text{Capacity (ranked score)} \times \text{Accidents (ranked score)}]^{1/3} \quad (7)$$

This composite score, like CV, is used mostly for generalized administrative planning and programming purposes. By combining the three parameters in this manner the lowest value may affect the calculation significantly. Thus, if any of the three values has a very low score, it caused that mile to be flagged. Figure 5 is a sample page from the 1979 inventory summary.

REVIEW OF ALASKA'S PAVEMENT RATING METHODS BASED ON RECENT FIELD STUDIES

After the Alaskan pavement rating method had been in use for 2 years a more detailed evaluation of its constituent measurements was thought necessary. A field study was begun in 1980 to investigate the repeatability of cracking and patching measurements made by different rating crews. Frequency of measurements necessary to estimate a true mean rut depth was also reviewed.

Method of Study and Data Acquisition

Five roadway sections were selected near Fairbanks that reflect the average range of road surface conditions commonly encountered. Each of the sections was rated by 15 different two-member crews using the current standard Alaskan procedure. Members were drawn mostly from the middle-level professional and technical ranks of road design, maintenance, right-of-way, and materials sections, but only four raters had previous pavement rating experience. Raters with previous experience were drawn from the department's research and development section.

Each crew of raters was given the same introduction to pavement rating and directed from one pavement section to another by the instructor. Ratings by each crew required a full day and the sequence of pavement sections remained constant throughout the duration of the study. Considered important was that the sequence of sections not change because this assured that the sun angle relative to the viewer remained consistent for each crew for each section. Sun illumination was known through accumulated field experience to greatly affect pavement crack visibility. To maximize the observational abilities of each rating crew all ratings were performed from a light truck or van. A nearly vertical windshield combined with a relatively high seating position allowed the most advantageous view of pavement surface of any standard type of vehicle. Each section was inspected at under 10 mph in order to identify and measure cracking. Rut depths were measured in each of the four wheelpaths every 0.2 mile. Distances were measured with an electronic odometer capable of 1-ft resolution.

Analysis of Field Data

An indication of measurement variabilities between crews is given through the coefficient of variation (C_v) associated with each distress type:

$$C_v = (\text{SD}/\text{Mean value}) \times 100 \quad (8)$$

where SD is the standard deviation.

In general, a small C_v of approximately 5 to 10 percent indicates that a good estimate of a true mean value is possible from relatively few individual measurements. C_v values associated with measurement of all pavement distress indicators were considered very high. This tends to contradict the initial hypothesis that, because of the simplicity of the rating method, reproducibility of ratings among crews could be taken for granted. The following estimates of C_v were calculated from project data.

Item	Avg. C_v (%)
Type 1 alligatoring	43
Rut depth, calculated average	25
Rut depth, calculated standard deviation	40

The significance of the foregoing listing should not be understated as the uniformity of C_v from section to section indicated.

Type 2 (severe) alligator cracking and full-width patching are not listed because their infrequent occurrence within the test sections did not provide an adequate sampling to allow a good evaluation of differences among rating crews. Based on these limited observations, however, the variability in measuring patching length is somewhat lower than for alligator cracking that has C_v of perhaps 10 to 20 percent. A clear distinction between type 1 and type 2 alligator cracking was not easily made by the rating crews. The tendency, except in the most obviously severe cases, was to place all cracking into the type 1 category. Most crews apparently selected a lower severity classification whenever the question of degree of damage arose. This problem can probably be remedied to some extent during the instruction process by specifically advising that pavements be rated critically.

The large amount of variability observed in the collected data is given in Table 3. In view of the similarity in training and background among these experimental raters and previous inventory crews, these variabilities could be expected on pavement sections throughout the state.

Table 3 summarizes the variation in data for the sections tested. The variation in all the pavement distress measurements is large when considering the range in cracked length. When considering the variation as a percentage of section length, the variation is less. The maximum variation between the mean and maximum values is 7 percent. It can be argued that the alligator cracking expressed as a percentage of section length need only be determined to be within 10 percent of the true percentage for inventory purposes. If it is assumed that the mean is the true value, then all five sections meet this criterion. More detailed measurements may be necessary for design processes.

The overall effect of variations in crew measurements on determinations of rut depth is magnified because ADOTPF usually reports maximum rut depth in terms of average plus two standard deviations. For example, the mean, mean + 1 standard deviation, and mean + 2 standard deviations are given for the fol-

Figure 5. Sample pavement inventory.

LOCATION			CONDITION ELEMENTS					PERFORMANCE VALUES			
TERMINI	SECTION LENGTH	CDS MILE	ADT	RIDE (in/mi)	CRACKING (%/mi)	PATCHING (%/mi)	RUTTING (in/1000)	CONDITION VALUE	SERVICE VALUE	ACCIDENT VALUE	COMPOSITE VALUE
FAP 35 Parks Highway (State Route 170000)		312									
		313		24	0	0	72 ☼	85			
		314		63	5	0	83 ☼	52			
JCT Old Nenana Hwy. Ester JCT	20	315		35	0	0	62 ☼	82			
SECTION AVERAGES			806	29	1	1	87	79	82	74	78
		316		132	41	15	137	16			
		317		154	16	29	119	15			
		318		129	4	14	188	26			
		319		166	27	21	169	13			
JCT FAS 649 GEIST ROAD	5	320		88	0	36	29	29			
SECTION AVERAGES			2362	134	18	23	128	20	1	60	11
JCT FAU AIRPORT SPUR	1	321		61	0	1	65	67			
FAP 35 PARKS HIGHWAY (STATE ROUTE 170000) SECTION AVERAGES			4090	61	0	1	65	67	60	72	66
		322		56	0	1	116	63			
		323		90	0	3	220	39			
		324		112	0	0	138	49			
JCT STEESE AND RICHARDSON HWYS	4	325		60	0	0	340	45			
SECTION AVERAGES			17919	80	0	1	204	49		42.6	

Table 3. Observed variation in pavement distress measurements.

Pavement	Section No.	Range	Average	Range as Percentage of Section Length	Avg as Percentage of Section Length
Type 1, alligator cracking	1	29–187 ft	140 ft	1–4	3
	2	35–1,434 ft	820 ft	0–14	8
	3	69–700 ft	300 ft	1–13	6
	4	54–218 ft	120 ft	1–4	2
	5	190–505 ft	340 ft	4–10	6
Type 2, alligator cracking	1	None detected	0 ft	0	0
	2	4–19 ft	0 ft	0	0
	3	0–13 ft	0 ft	0	0
	4	None detected	0 ft	0	0
	5	None detected	0 ft	0	0
Full-width patching	1	350–382 ft	360 ft	7–7.5	
	2	439–1,042 ft	820 ft	8–20	
	3	91 197 ft	100 ft	2–4	
	4	None detected	0 ft	0	
	5	None detected	0 ft	0	
Rut depth, average inner wheelpath	1	0.016–0.059 in.	0.040 in.		
	2	0.110–0.393 in.	0.210 in.		
	3	0.114–0.289 in.	0.180 in.		
	4	0.100–0.257 in.	0.170 in.		
	5	0.134–0.271 in.	0.210 in.		
Rut depth, SD inner wheelpath	1	0.005–0.055 in.	0.020 in.		
	2	0.051–0.601 in.	0.160 in.		
	3	0.040–0.198 in.	0.080 in.		
	4	0.023–0.263 in.	0.080 in.		
	5	0.060–0.241 in.	0.110 in.		
Rut depth, average outer wheelpath	1	0.050–0.167 in.	0.090 in.		
	2	0.116–0.410 in.	0.240 in.		
	3	0.089–0.248 in.	0.150 in.		
	4	0.062–0.272 in.	0.180 in.		
	5	0.172–0.445 in.	0.270 in.		
Rut depth, SD outer wheelpath	1	0.022–0.105 in.	0.055 in.		
	2	0.069–0.596 in.	0.240 in.		
	3	0.040–0.322 in.	0.090 in.		
	4	0.032–0.184 in.	0.080 in.		
	5	0.098–0.257 in.	0.160 in.		

lowing sections by using Table 3 outer wheelpath data.

Section	Mean (in.)	Mean + SD (in.)	Mean + 2 SD (in.)
1	0.090	0.145	0.200
2	0.240	0.480	0.720
3	0.150	0.240	0.330
4	0.180	0.260	0.340

The foregoing examples demonstrate a wide range of uncertainty as to the measured depth of rutting even though calculated mean values are low.

Discussion of Measurements of Alligator Cracking

In several of the following figures the variations in measurements have been normalized. This normalization step is used so that various road sections can be compared directly even though each has a different mean rut depth or length of alligator cracking. Normalization of scoring, (e.g., percent of alligator cracking and average rut depth) is accomplished as follows:

$$\text{Normalized percentage of alligator cracking} = (A-B)/C \tag{9}$$

where

A = percentage of alligator cracking as measured by an individual crew on a specific road section,
B = average percentage of alligator cracking calculated from the measurements of all crews on the above section, and
C = standard deviation value calculated from the measurements of all crews on the above section.

Figure 6 shows how normalized scores of individual crews rank in relation to calculated average values on all five pavement sections. This plot indicates the ability of certain crews (e.g., 7 and 8) to see more damage than others. Conversely, crew 14 saw much less cracking in all five pavement sections than the calculated average. Figure 6 includes the instructor's subjective assessment of each crew in terms of (a) communication between crew members [rated low (L), moderate (M), and high (H)] and (b) initial impression of rating ability (rated fair, good, and expert). Note that crews 2 and 10, rated expert by the instructor, had at least a full season's rating experience and were included for purposes of comparison with the other crews. Although Figure 6 indicates that some crews could apparently see more pavement damage than others, this difference was not accounted for in obvious attitudes or abilities. Note, however, that crew 8, which saw much more pavement damage than crew 14, also rated higher in the instructor's opinion. Best results are obtained when conversation concerning the rating process is encouraged between crew members, especially during the first few days of inventory.

The data in Table 4 attempt to delineate reasons for differences among crew ratings. The samples have been broken down into a stratified format and cross indexed in terms of crew communications and weather and pavement surface condition at the time of rating. The numbers given in Table 4 as $\bar{X}$ (characteristic sample average) and SD (characteristic sample standard deviation) have been normalized, as

Figure 6. Range of variation in each crew's measurement of type 1 alligator cracking.

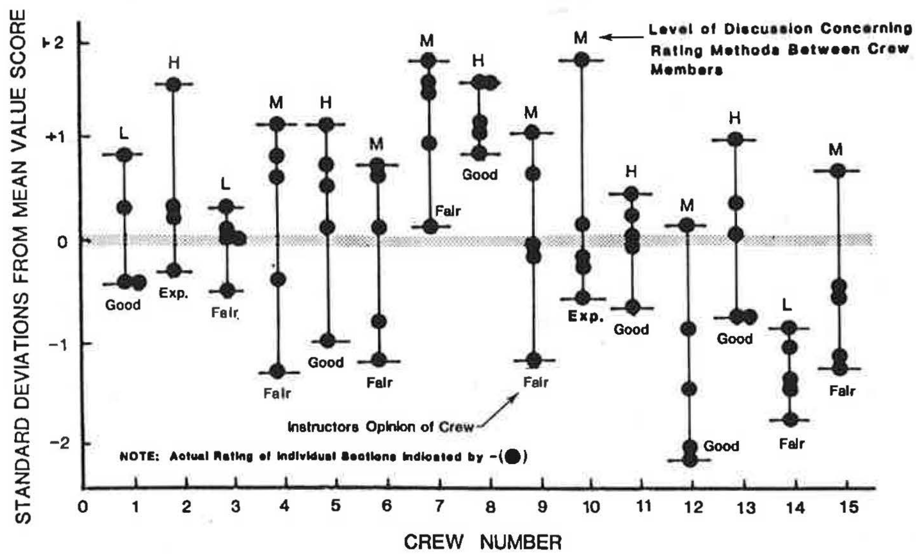


Table 4. Analysis of type 1 alligator cracking.

Discussion of Rating Methods	Statistic	Sunny	Cloudy	Cloudy, Slightly Wet Road Surface	Rain, Very Wet Road Surface	Weighted Avg of Rows
Active	$\bar{X}$	-0.1	0.5	0.8	0.2	0.4
	SD	0.5	0.7	0.8	0.9	0.7
	N	7	5	8	4	
Moderate	$\bar{X}$	0	0.3	-0.3	-1.7	-0.1
	SD	0.8	1.1	0.6	0.5	0.9
	N	7	19	5	4	
Little	$\bar{X}$	0.1	-0.7			
	SD	0.6	0.8	No samples	No samples	0.5
	N	4	10			0.7
Weighted avg of columns	$\bar{X}$	0	0	0.4	0.8	
	SD	0.6	1.0	0.7	0.7	

previously described, thus allowing all five pavement sections to be considered in the same analysis. The combination of a slightly wet (SW) pavement surface and a highly communicative crew resulted in more visible cracking and a characteristic average of +0.8 SD above the overall sample average. Also, in examining the weighted (for sample number) averages of both rows and columns, good crew communication and a slightly wet road surface are individually associated with increased damage observation.

Surface Wetness

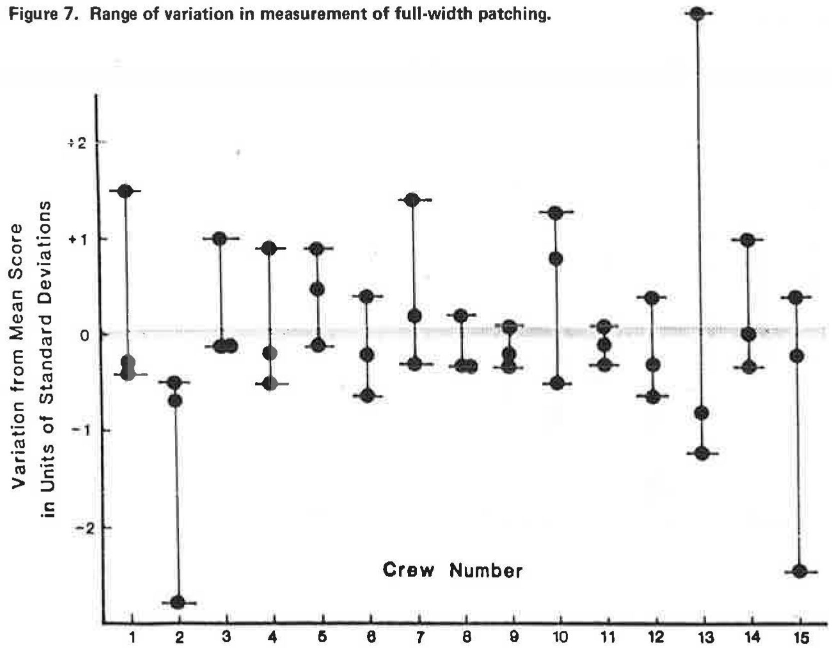
The effect of a slightly wet surface in optimizing the visibility of alligator cracking is fairly obvious to even the casual observer and can often cause hairline alligator cracking to stand out in vivid detail. On the other hand a very wet road surface, such as obtained during or shortly after a rainstorm, camouflages all but severe cracking. Observations of cracking should be discontinued during rainstorms or other periods when the pavement surface is covered by free water. Table 4 generally associates the least observed cracking with a very wet (VW) surface condition. The ideal, slightly wet surface condition is created when the road surface is dry except in and around individual cracks. In this case, water stored in the cracks during rainfall will keep the adjacent pavement wet longer than in areas of no cracking.

The previous discussion leads to the conclusion that pavement ratings could be done best shortly after a rainstorm; however, a dry road condition represents the more normally encountered situation. Because of the need for a standard rating procedure crack measurements should be made only on dry pavement.

Sun Angle

Illumination effects due to variations in vertical and horizontal sun angle are known to affect crack visibility strongly. Experience indicates that optimal lighting conditions are provided by a more or less head-on sun incidence. Frontal light tends to shade and, therefore, darken the visible side of crack segments that are perpendicular to the observer and most easily viewed. This has the net effect of maximizing apparent tone and texture differences between cracked and uncracked pavement. The travel direction chosen for the experimental ratings produced over-the-shoulder lighting on four of the five pavement sections, which is usually considered a worst-case viewing condition. Each test section was examined at approximately the same time of day by each crew to ensure a consistent sun angle.

Figure 7. Range of variation in measurement of full-width patching.



Discussion of Full-Width Patching Measurements

The occurrence of full-width patching within the test sections was somewhat limited. Data from section 3 indicate that differences in patching measurements among crews may be about half those expected from observations of cracking. The distribution of normalized scores indicated in Figure 7 represents only the three test sections that actually contained patching. The variation among crews is markedly less pronounced than for alligator cracking.

Patching appears to be more easily measured than alligator cracking even though both are evaluated in a similar way. In most cases patching, at least new patching, is actually seen quite easily. Observation conditions that provide the best view of alligator cracking also tend to make patched areas stand out. Again, very wet surfaced roads resulted in the most variable measurements among crews, and cracks are most easily seen on a slightly wet pavement. Regardless of the better viewing condition afforded a slightly wet surface, the dry road condition is most commonly encountered in field work and is, therefore, suggested as the standard for inventory purposes.

Discussion of Rut Depth Measurement

The approach taken initially to determine a sample number (as indicated in Figure 4) was a rough attempt to limit the possibility of gross errors. Sufficient field data have since been collected to allow a more valid estimation of rut depth. The problem of rut depth measurement can be addressed by normal statistical methods. The principal questions asked are

1. How frequently must rut depth measurements be taken? and
2. Must rut depth measurements be taken in both inner and outer wheelpaths?

Sampling Frequency

The frequency of sampling must be high enough to ensure (to some specified confidence level) that a calculated mean rut depth is reasonably close to the actual mean rut depth. Actual or population average in this case is that value that would be measured from an infinitely large sampling. Sampling tables available in references such as the Chemical Rubber Company statistical handbook (4) indicate minimum sample numbers necessary to attain specific levels of confidence against either a type-1 or type-2 error being committed. A type-1 error occurs if statistical calculations indicate that the sample mean is not representative of the population mean, when in fact it is. Conversely, a type-2 error occurs when statistics indicate that the sample mean is representative of a population mean when it is not.

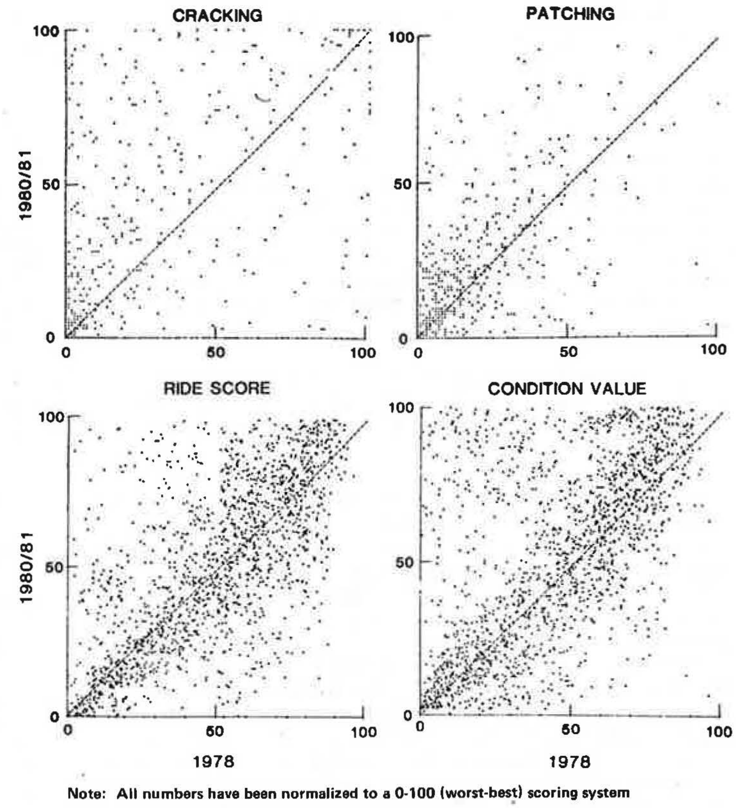
It was assumed that, for predicting the actual rut depth average from sample data, an error of no more than ±0.05 in. would be allowable. In most sampling situations, little concern is expressed over type-2 errors. This philosophy leads to 50 percent level of type-2 error control (i.e., no control and a significantly reduced sample size).

Determination of sample size is dependent on expected standard deviation; therefore, it is important to consider the magnitude of values that might commonly be encountered. Rut measurements made on the five Fairbanks test sections indicated a range of standard deviations from about 0.02 to more than 0.35 in. associated with average rut depths between 0.02 and 0.40 in. Indications of variability in rut measurement derived from the Fairbanks test section data suggest that minimum sampling be based on a standard deviation perhaps as high as 0.30 to 0.35 in. This magnitude of deviation plus 90 to 95 percent confidence level against error results in a minimum sample size in excess of 100. Rut depth measurement, therefore, begins to appear impossible except through an automatic rut-measuring device capable of high-density sampling.

Alternatives

Several sources of rut measurement data were used to construct functional relationships among average rut depth, calculated standard deviation, and required number of sampling points. This report substantiates previous contentions (5) that true mean rut depth can be accurately characterized only through a

Figure 8. Comparison of 1978 with 1980/81 pavement inventory data.



large sampling. Problem rut depths on the order of 0.4 to 0.5 in. or larger would require an assumed standard deviation of at least 0.3 in. Reasonable error confidence levels indicate a sampling obviously greater than 100/section. Furthermore, the inability to predict whether inner or outer wheelpath represents the worst-case condition would require doubling of the sampling effort. In dealing with this question the choices are

1. Assume rutting to not be a problem and cease measurement,
2. Perform a few random measurements per mile at locations that appear from general observation to represent worst-case conditions, or
3. Purchase or build an automatic device for rut measurement as described by Jurick (5).

Deeply rutted sections are usually associated with severe alligator cracking on most Alaskan road sections; therefore the measurement of both is unnecessary. Rutting within the state is rarely as deep as 0.5 in., which is considered critical in most literature sources. Alternative 1 appears to be a reasonable course of action at present. Alternative 2 provides numbers and the numbers can, of course, be included in subsequent discussions of pavement condition. The numbers generated from alternative 2 have no basic statistical validity, however, and might be thought of as inventory garbage. Alternative 3 is preferred if departmental policy requires an accurate determination of rut depth. A 1981 cost estimate for the purchase of an automatic rut-measuring device was $150,000 to $200,000.

COMPARATIVE LOOK AT PREVIOUS INVENTORY DATA

This section looks at actual pavement inventory data in view of the findings of this paper. Because of the rather gross variability evident in the experimental measurement of cracking and patching, a direct mile-by-mile comparison between two previous inventories is made.

Figure 8 shows the apparent variation in pavement distress between 1978 and 1980/81. As shown, these data have been normalized to provide a total scoring range of 0 to 100 (worst-best). Note that data at coordinates (0 percent, 0 percent) and (100 percent, 100 percent) are often clustered in the graphs for cracking and patching, which accounts for the appearance of fewer than expected individual points on these plots. The same number of data points was recorded for all of the graphs.

A line of x = y has been included in each plot and differentiates pavement sections that apparently or actually improved with time (points above the line) from those that became worse (points below the line). Examination of plotted data indicates

1. A very high degree of overall scatter and
2. An unusually large number of data points above the line of x = y (i.e., performance improvement with time).

Taken together, these findings demonstrate a marked degree of randomness inherent in the rating process. The implication of point number two is especially significant in view of the common-sense assumption that pavement condition deteriorates with time. This assumed generality could, of course, be altered by reconstruction, overlay, or careful patching, and no attempt was made to remove specific points that represent reconditioned pavement sections from the plots. This should, however, account for only a small percentage of total rated mileage. A significant degree of randomness is suggested because even sections that scored better than average in 1978 showed a very high rate of apparent improvement with time. The likelihood that initially good pavements (scoring 50 to 100) will be improved substantially within a period of 3 years through maintenance is slight.

SUMMARY AND CONCLUSIONS

The development and evaluation of Alaska's inventory rating procedure for flexible pavements has been described. Development of the system was based on the generally accepted principles of pavement rating practice as outlined in recent literature. The Alaskan rating method attempts to measure basic elements of road quality from two important viewpoints:

1. The highway user--ride roughness and
2. The highway engineer--fatigue (alligator) cracking, major (full lane width) patching, and wheelpath rutting.

These rating features are reported on a mile-by-mile summary both individually and in terms of a composite serviceability score. A concerted effort was made during the development of the rating method to keep all distress measurements as simple as possible but still provide adequate information for pavement management needs.

The rating method was evaluated through a special field study and experience accumulated during the 3 years since its implementation. Findings indicate a large variation in the abilities of different rating crews to characterize the extent and severity of patching and cracking. The range of variation in crack and patch measurements obtained by 15 crews on 5 selected pavement sections was found to be as much as twice the mean measured value. These differences are apparently associated with the level of interest in the task expressed by each crew and weather factors that control visibility of pavement surface features. Examination of previous inventory ratings confirmed the data scatter indicated by the experimental pavement sections.

The variation in rut depth measurements was large enough to require very high sampling frequencies. A mechanized form of rut-measuring device, capable of more than 100 measurements per section in both inner and outer wheelpaths, is suggested. Marked differences between average depths of inner and outer wheelpaths require data from both locations in order to define the worst-case condition.

Conclusions

The assumption that Alaska's pavement rating methods are simple enough to ensure a high degree of reproducibility is not demonstrated by the available data. A great deal of variation is apparent in the field measurement of cracking, patching, and rutting. This is indicated through examination of experimental data as well as from data collected from previous inventory work. The use of machine measurements is suggested wherever possible in all phases of the rating process.

The visual rating of pavements is a difficult process that requires careful and rigorous standardized technique. Pavement rating instructions must be formalized to include guidelines for training rating crews and ensuring acceptable performance. Specifications are necessary for standardization of viewing height, acceptable lighting conditions, and vehicle speed.

Recommendations

The ability to quantify pavement performance is a requirement of almost any approach to pavement management. Alaska's pavement rating method should therefore be viewed as a tool to be improved rather than discarded.

Recommendations for improvement include

1. Phase out visual measurements of pavement distress as reliable machine methods become available;
2. Except for very rough classification purposes, discontinue rut measurements until sampling rates of more than 100/mile can be achieved; and
3. Continue existing approach but with greatly increased and improved crew training and a strict standardization of observation technique.

An ideal form of instruction would include the use of standard road sections. On these sections the crew would attempt to match the ratings assigned by experienced personnel. A five-day tuning period is suggested for new rating crews. Ratings performed during this first week would not be included in the inventory summary before verification by repeated observation.

Observation conditions for the inventory measurement of cracking and patching should be standardized:

1. Vehicle speed of 6 mph or less,
2. Rating of only completely dry road surfaces,
3. Use of optimal sun incidence whenever possible for best illumination--a horizontal sun angle of ±70 degrees from head-on or a vertical sun angle of more than 10 and less than 60 degrees from the horizontal (this point should be emphasized even if it requires that the direction of travel, i.e., direction of the rater's view, be changed),
4. Standardized viewing height of 5.5 ft ±0.5 ft, and
5. Use of utility van-type vehicle that has a nearly vertical windshield.

ACKNOWLEDGMENT

On behalf of the Alaska Department of Transportation, we wish to acknowledge technical assistance and funding support provided by FHWA. ADOTPF Interior Region Division of Planning and Programming, Design and Construction, and Maintenance and Operations are offered special thanks for providing temporary use of their employees as rating crew members.

REFERENCES

1. The AASHO Road Test: Report 5, Pavement Research. HRB, Special Rept. 61E, 1962, 352 pp.
2. E.J. Yoder and M.W. Witczak. Principles of Pavement Design, 2nd ed. Wiley, New York, 1975.
3. R.L. Terrel and R.V. LeClerc; Washington Department of Transportation. Pavement Management Workshop: Tumwater, Washington. FHWA, FHWA-TS-79-206, Sept. 1978.
4. The Chemical Rubber Company: Handbook of Tables for Probability and Statistics, 2nd ed. (W.H. Beyer ed.). CRC Press Inc., Vol. 1, Cleveland, Ohio, 1968.
5. R. Jurick. Automated Pavement Rut Depth Measuring System. Alaska Department of Transportation and Public Facilities, Fairbanks, Rept. AK-RD-80-1, 1981.

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