DESCRIPTION & EVALUATION OF THE ALASKA PAVEMENT RATING PROCEDURE

by

Robert L. McHattie, P.E., Senior Research Engineer
Billy G. Connor, P.E., Senior Research Engineer

February 1982

STATE OF ALASKA
DEPARTMENT OF TRANSPORTATION AND PUBLIC FACILITIES
DIVISION OF PLANNING AND PROGRAMMING
RESEARCH SECTION
2301 Peger Road
Fairbanks, Alaska 99701

in cooperation with

U.S. Department of Transportation
Federal Highway Administration

ABSTRACT

This report describes and examines pavement condition rating methods used on Alaska's roadways since 1978. The methods were intended to provide the specific performance data necessary to optimize construction/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 but are also combined with traffic data to indicate more general roadway serviceability levels.

Field evidence shows that a high degree of variability presently exists in the measurement of cracking, patching and rutting. Coefficients of variations above 20% 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 which should improve manual rating methods; although mechanized or electronic data acquisition techniques must be developed to eliminate human error.

Introduction:

The Alaska Department of Transportation and Public Facilities (DOTPF) 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 which comprise it. The current Alaskan rating attempts to quantify surface fatigue cracking, patching and wheelpath rutting as an aid to design, construction and maintenance planning. This report examines the amount of error associated with measurements of pavement of distress, and suggests improvements which can be incorporated into future inventory work.

Principal topics addressed are:

- the development of Alaska's rating method for flexible pavements
- accuracy and reproducibility of field measurements for pavement ratings conducted by a two-man crew
- ways of improving observation and sampling methods

The research data base utilized in this study consisted of data and experience accumulated from two complete inventories of the Alaskan paved highway system conducted over the past three years. The study also examines results of repetitive sampling conducted specifically for this project on five typical pavement sections located near Fairbanks, Alaska.

DEVELOPING AN ALASKAN
PAVEMENT RATING PHILOSOPHY

Background and Literature:

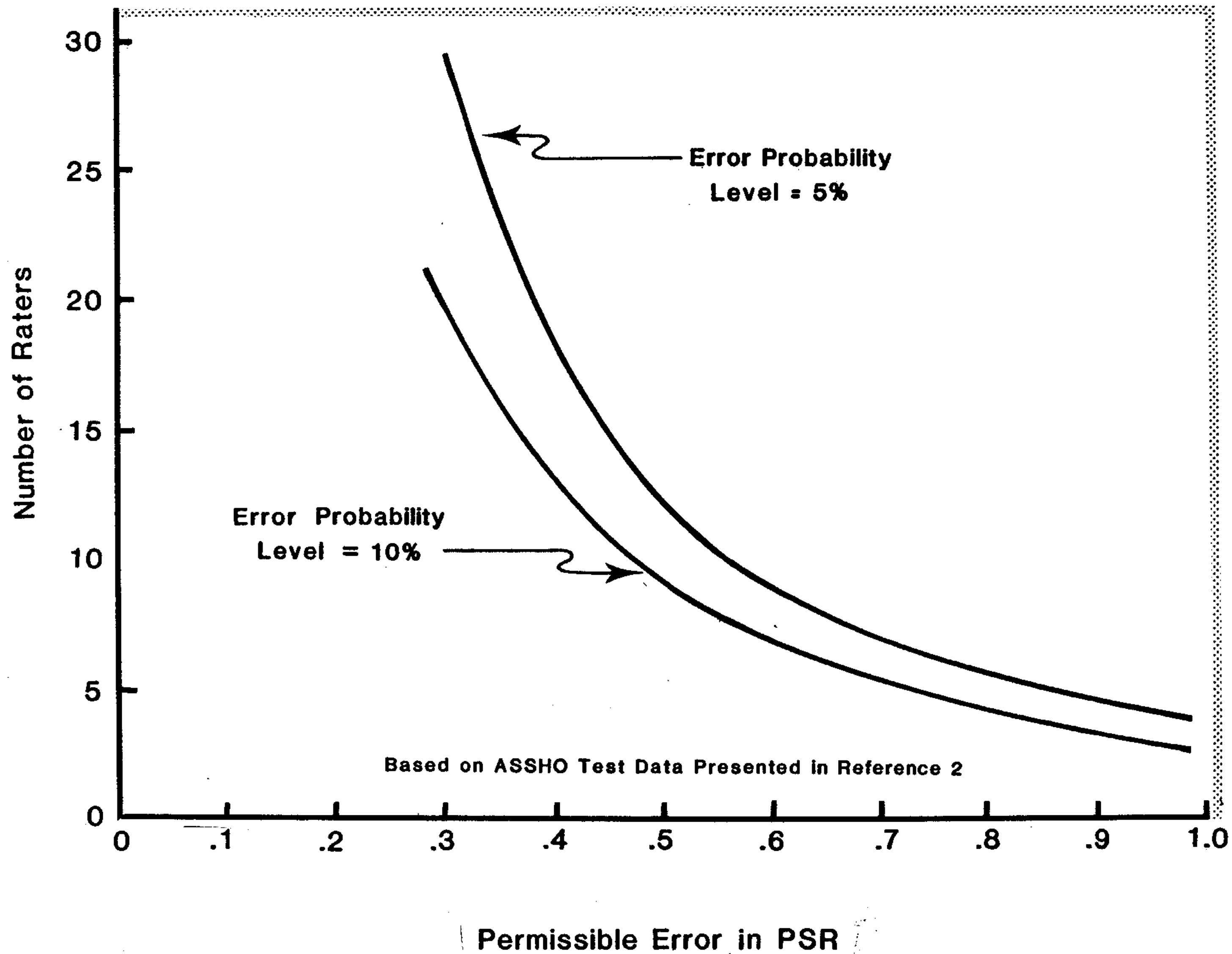
During the winter of 1977-78 the Planning Division of Alaska's DOTPF decided to revise its existing highway inventory procedure to fill the need for accurate, quantitative data for programming highway maintenance and construction funds. The Department's Research Section was commissioned to produce a practical inventory which would stand the scrutiny of statistical evaluation.

Before attempting to assemble an Alaskan highway inventory procedure, 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 50's to early 60's era which numerically classified pavements 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 were; surface deterioration, ride roughness, rutting, cracking and maintenance patching. This rating technique produced a number termed "Present Serviceability Rating" (PSR) intended for classifying a given section of road. Figure 1 indicates the number of individual raters necessary, statistically, to estimate the "true" value of PSR using the completely subjective AASHO method. This figure indicates that for 1 or 2 raters the error associated with estimation of PSR is greater than 1. Since the error can range either + or - from the true value, the full range of possible estimation is therefore 2 which represents 1/3 of the total 0-5 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 closely matching PSR panel scores. Regression equation independent variables consisted of standardized measurements of fatigue cracking area, maintenance patch area, wheelpath rut depth and longi-

Figure 1
ESTIMATING PSR



tudinal surface variation (roughness). The road surface condition values calculated by the regression equation are termed "Present Serviceability Index" (PSI).

- AASHO Equation -

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

Where:

SV = mean slope variance in the two wheelpaths as measured absolutely by a longitudinal profilometer (inches per mile x 10^6)

RD = mean rut depth (inches)

C+P = cracking + patching (square feet/1,000 square feet 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 then recently complied summary and critique of highway agency pavement management practices. A federally sponsored workshop was held in Tumwater, Washington, in November of 1977 which examined the existing state-of-the-art in the field of Pavement Management Systems (PMS). U. S. and Canadian representatives were selectively invited providing they were actively implementing, and therefore experienced in a PMS program. At the time the 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 critically compared them. Rating system elements were suggested which provided the best input to the overall PMS.

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

1. Ride roughness should be objectively rated.

2. Structural capacity should be rated but it was not clearly decided whether to rate structural capacity on the basis of deflection tests or surface distress measurements.

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 Present Serviceability Index (PSI) was said to provide a valid measure of pavement condition.

6. There is very little standardization of terminology and measurement technique among the available systems of pavement rating when these systems are examined in detail.

Each of the above points was seriously considered prior to development of the Alaskan rating system. Table 1, taken from the Tumwater Conference Report, indicates the salient features of the road rating methods of U.S. states and Canadian provinces in attendance.

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

TABLE 1 - Pavement Monitoring Features and Evaluation

FEATURE / AGENCY	Surface Condition	Roughness of Ride	Skid Resistance	Structural Capacity	Rating System	Primary Decision Criteria
Arizona	Crack survey	Mays Ride Meter on annual basis	MU Meter-500 ft at each mile post	Dynaflect - 3 locations per mile	Pavement Management Information System PMIS	Compares major maintenance alternatives
California	Pavement condition survey based on alpha-numeric rating	Ride score not part of pavement distress	Measured periodically		Alpha-numeric rating combines severity and extent of defects	Defects compared to repair strategies and costs
Florida	Structural defects of cracking, rutting and patching	Mays Ride Meter 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, RI, 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, PSS. Based on correlation w/known serviceability levels	Aimed at identifying budget needs, failed pavements, effectiveness of expenditures
Pennsylvania		Mays Meter used to develop serviceability	ASTM Skid trailer	Road Rater		
Texas	Distress survey based on vehicle mounted camera-visual distress rating	Mays Meter 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, ½-mile sections tested every 2 miles	Dynaflect. For predicting remaining life	Present Serviceability Index, PSI	Overall priority ranking for preventative rehabilitation
Washington ✩	Pavement condition survey every 2 years covering entire network	PCA Roadmeter on all sections	ASTM Skid trailer for high accident locations-considered separataly	Limited use of Benkelman Beam	Combined structural rating and ride score	Tabulate rehabilitation strategies and costs based on pavement condition
Ontario	Pavement Condition Rating, PCR 1-2 year cycle	Subjective. Riding Comfort Index, RCI		Dynaflect. Random sample locations in need of rehabilitation	Subjective. Pavement Condition Rating, PCR	Required overlay prediction based on expected performance
Saskatchewan	Annual surface condition rating	PCA Roadmeter on intervals of 1 month to 1 year		Benkelman Beam data used for overlay design	Condition Rating system used to priortize projects for overlay or scaling, etc.	Preventative maintenance is primary goal

✩ Photo logging of entire system

It was decided that a pavement condition (rating) must:

1. Provide information for planning/prioritizing 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.

3. Provide design feedback information.

Alaska's Existing Rating Method - Principal Elements:

Literature review plus common sense pointed to the need for a rating method which would adequately characterize the road condition while allowing a high degree of reproducibility at a minimal cost. The data must provide true reproducable characterization of pavement condition changing from year to year in a rational manner, e.g., pavements should not anomalously appear to heal 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 (alligatoring)
2) major patching (at least full lane width)
3) wheelpath rut depth
4) ride roughness as measured by the Mays Ridemeter

Fatigue cracking was selected as a rating parameter because it serves as an excellent indicator of structural condition and load-life potential. The pavements design-life vehicle load capacity is said to be reached when significant alligatoring becomes apparent. Fatigue cracking is also often associated with unacceptable rutting, vehicle ride roughness and pavement surface disintegration.

Major patching, necessitated to repair a host of problems including fatigue cracking, embankment settlement, rutting, etc., gives a general picture of the maintenance effort required on a given road section. It 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 1/2 inch is a safety hazard which 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 and/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 that characteristic of the pavement which is of primary concern to the driving public. The combination of differential settlement and leveling patches are common to all parts of Alaska and together are the major cause of roughness felt by the driving public. Ride roughness is objectively measured on a continuous basis using available technology such as the Mays Ridemeter.

Some recognized surface distress features were disregarded in order to simplify the rating process. These include ravelling, longitudinal cracks, thermal cracking, shoving and bleeding, potholes and deflection.

Recent skid measurements (Statewide in 1975) have 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.

Statewide deflection testing will utimately become part of the normal inventory process. This process began in 1982 and will require approximately 5 years per statewide cycle. The falling weight deflectometer is presently being used to collect inventory data.

RATING AND SCORING PROCEDURES USED SINCE 1978

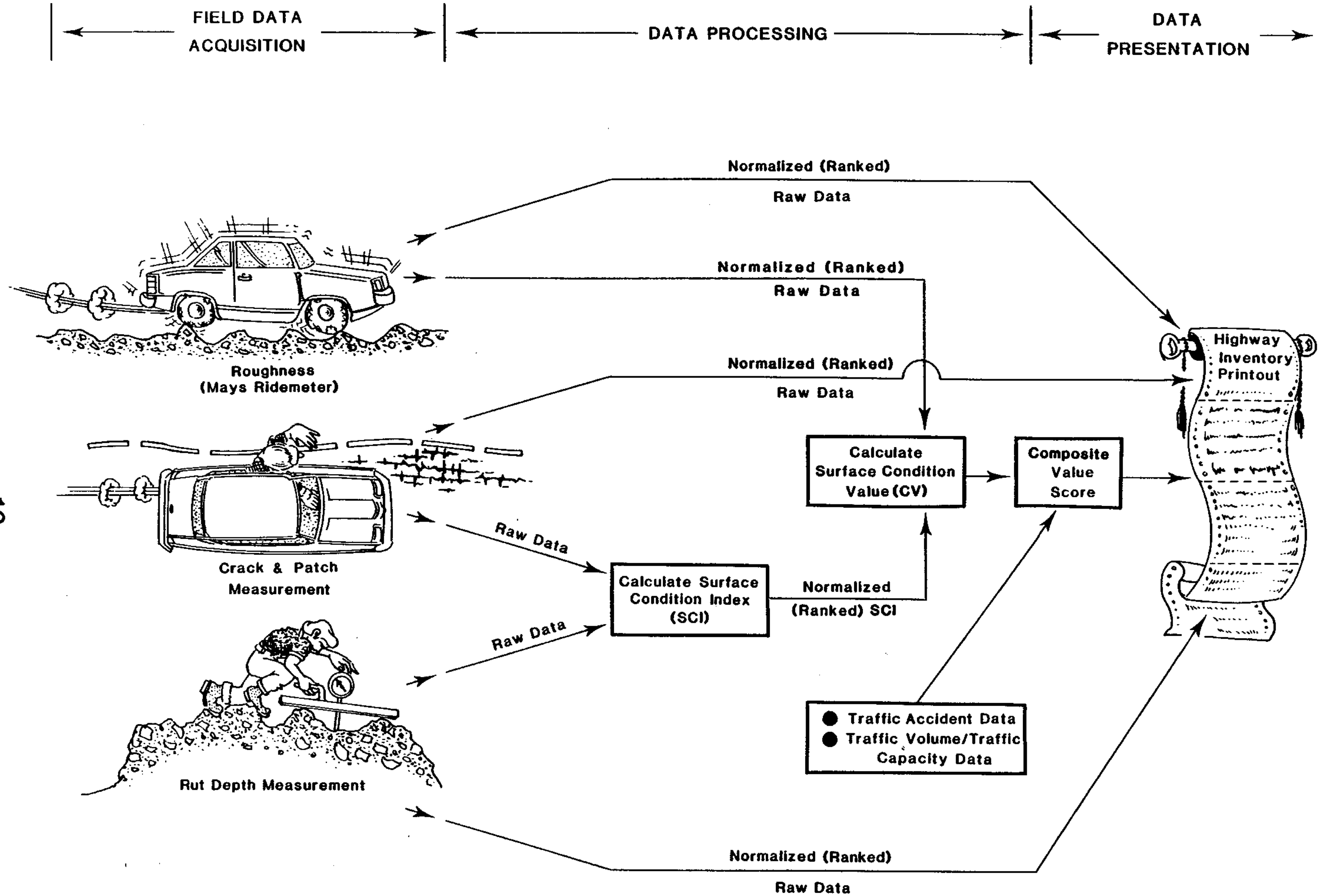
This section discusses pavement rating methods used by Alaska DOTPF since 1978 and shows how field data are manipulated for purposes of scoring and reporting. Figure 2 illustrates the manner in which raw field data are transformed into a useful pavement inventory report.

Development of Field Methods:

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

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

Methods for measuring alligator cracking and major patching were initially evaluated on seven sections of roadway near Fairbanks. Each section was divided into tenth mile subsections which were independently rated. Full width patching was characterized on the basis of total length (density), while 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 that which is visible while driving 7-10 mph. It is measured as total percent of road section length exhibiting cracking, regardless of wheel path location. Histograms were constructed from field data (Figure 3) to show the frequency distribution of fatigue cracking for the subsections within each

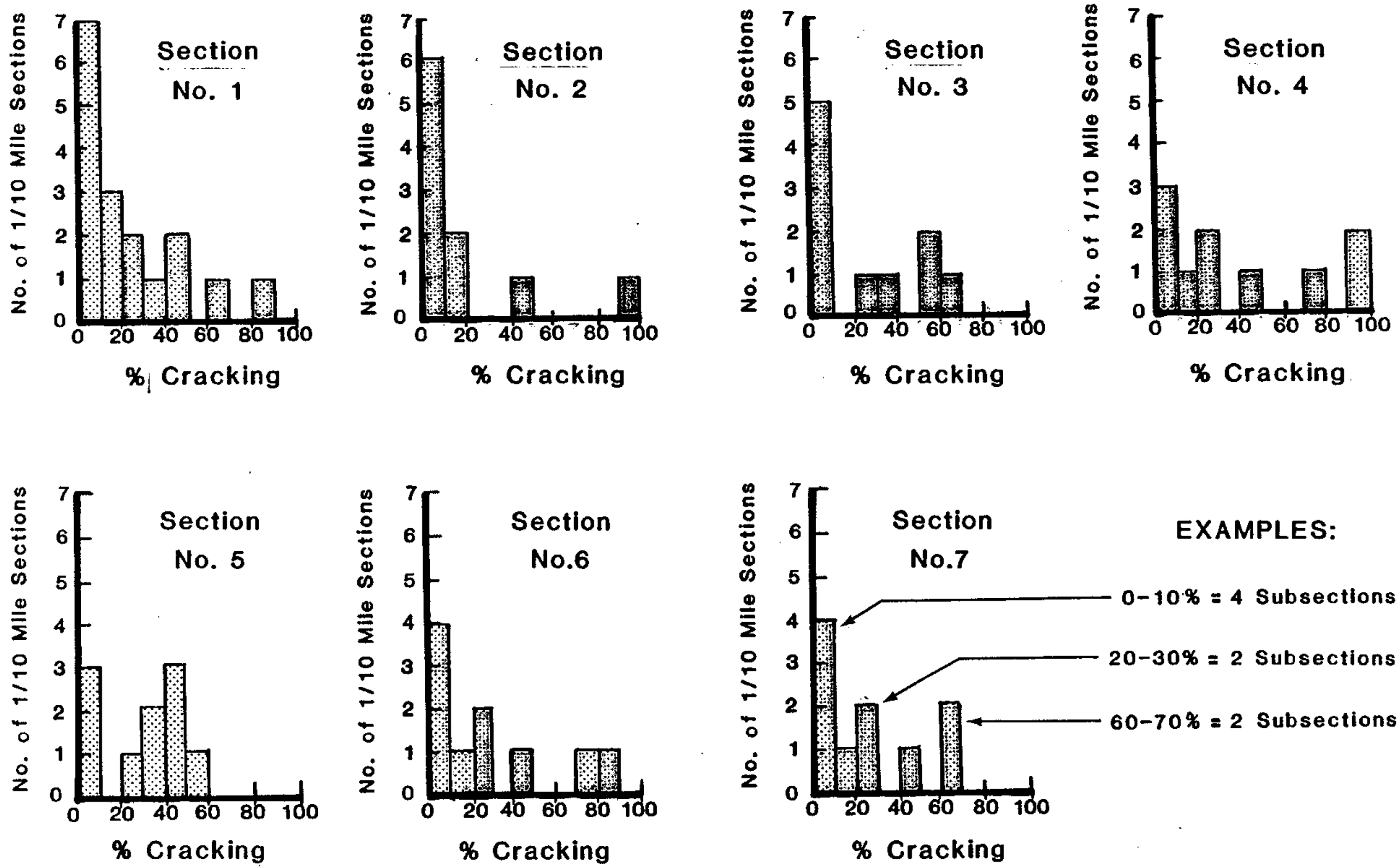


Elements Of The Alaskan Pavement Inventory

Figure 2

ALLIGATOR CRACK FREQUENCY DISTRIBUTION

(variation in amount cracking in subsection measurements within 7 test sections)



NOTE: Except for section No. 1, all were 1 mile in length and included 10-1/10 mile subsections.
Section No. 1 was 1.7 miles in length and included 17-1/10 mile subsections.

Figure 3

mile. The distribution of cracking is strongly poly-modal (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 shows the probability of randomly selecting a 0.1 mile sample which will predict the true mean condition of each section of roadway. The probability is obviously small in all cases. In view of these data it was apparent that a one, two, three or more mile length of paved road could not accurately be rated for fatigue cracking based on measurements in a randomly selected subsection several hundred feet long. The normal assumption of a 10-20 percent sampling density is of no value in this case. As a result of this initial trial, it was apparent that fatigue cracking must be measured by continuous observation through each mile of roadway. All data collected subsequent to the initial trial has supported this decision.

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

Rut depth measurement frequency was also briefly examined prior to development of the rating method through multiple readings taken on each of eight one-mile long pavement sections near Fairbanks. Rut depth averages ranged from 0.185 inch to 0.244 inch. The sample standard deviations 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. It was assumed from these trials 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:

$$\mu_o = \bar{x} \pm S\,T\sqrt{N}$$

where: μ_o = true population average, i.e., true average rut depth

$\bar{x}$ = average rut depth as determined from sample

S = standard deviation of sample

N = number of measurements constituting the sample

t = "students t" value for a given confidence level and N

Table 2

Probability of Sampling True Mean Performance

Section	% Total Area Cracked	Range Acceptable	Probability of Selecting 0.1 Mile Section With Correct % 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

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:

$$(\mu_o - \bar{X}) / s = T / \sqrt{N}$$

The error in estimating true rut depth average, i.e., $\mu_o - \bar{X}$ 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 a least 4 locations would be necessary to insure 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. Since S of the trial road sections averaged approximately 0.05 inches, it was expected that errors in estimating rut depth during inventory work would be no larger than + or - 1.6 X 0.05, i.e., 0.08 inches. This accuracy was considered good enough for beginning the pavement inventory process. Less than 4 readings per mile were required in the 1978 rating method if rutting was generally observed to be less than 0.25 inch.

Summary of Required Measurement Frequencies:

It was decided on the basis of very limited field trials previously described 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 absolutely reinforced the idea of utilizing a 100% sampling.

The frequency of measurements necessary to adequately determine average rut depth was calculated from a preliminary statistical assessment. Measurement of ruts was known to be a disproportionately time consuming job when compared to other distress observations. It was hoped that accumulating experience would show that no more than 4 sets of readings would be required per mile of road.

PLOT OF RELATIVE ERROR FACTOR ($t/\sqrt{n}$) VERSUS SAMPLE NUMBER (n)

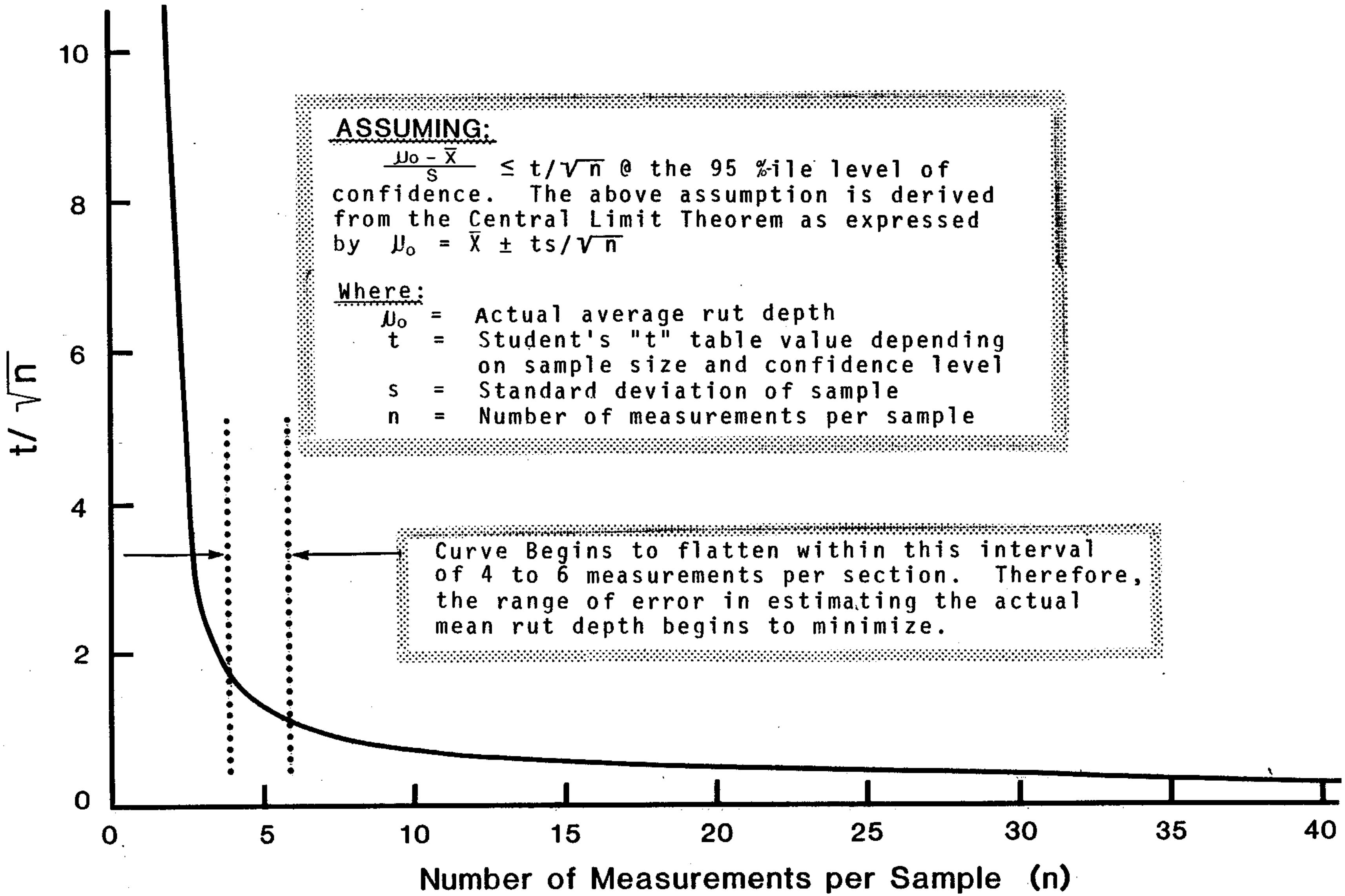


Figure 4

Road Condition Scoring Utilizing Alaska's Pavement Rating:

Alaska utilizes its pavement inventory data to construct a mile by mile summary report listing individual condition scores (% cracking, rut depth, % 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:

$$CV=\frac{\text{Mays Ridemeter Score(Ranked)}+\text{Surface Condition Index(Ranked)}}{2.0}$$

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

$$\%\text{-worse-than}=[(1/2E+L)/N]X100$$

where: E=number of statewide sections rated the same

L=number of statewide sections rated worse

N=total number of statewide sections rated

Mays Ridemeter score as shown in the CV equation is derived directly through the percentile ranking equation from raw Mays Ridemeter data. Surface Condition Index (SCI) is calculated by means of the following equation and then transformed to a %-worse-than ranking through the ranking formula.

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

where: R=average rut depth (inches)

A=% of road section which is alligator cracked

P=% of road section which 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 describing 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:

Composite Value Score =

$$[\text{Condition value} \times \text{Capacity (ranked score)} \times \text{accidents (ranked score)}]^{1/3}$$

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

Figure 5

Example Of Pavement Inventory Printout

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		312		24	0	0	72 ☼	85			
(State Route 170000)		313		63	5	0	83 ☼	52			
		314		35	0	0	62 ☼	82			
JCT Old Nenana Hwy. Ester JCT	20	315									
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	

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

Introduction:

After the Alaskan pavement rating method had been in use for 2 years it was decided that a more detailed evaluation of its constituent measurements was needed. 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, to reflect the average range of road surface conditions commonly encountered. Each of the sections were rated, using the current standard Alaskan procedure, by 15 different 2-member crews. 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 had prior 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. It was considered important that the sequence of sections not change because this assured that the sun angle relative to the viewer remained consistant for each crew, for each section. Sun illumination was known through accumulated field experience to greatly affect pavement crack visibiltiy. In order to maximize the observational abilities of each rating crew, all ratings were performed from a vehicle of light truck "van" configuration. A nearly vertical windshield combined with a relatively high seating position allowed the most advantageous pavement surface view of any standard vehicle type. 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 two-tenths mile. Distances were measured with an electronic odometer capable of one (1) foot resolution.

Analysis of the 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=(std. deviation/mean value) X 100

In general a small C_v of around 5-10% 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 rating method's simplicity, reproducibility of ratings between crews could be taken for granted. The following estimates of C_v were calculated from project data:

Type I Alligatoring	C_v average = 43%
Rut Depth (calculated average)	C_v average = 25%
Rut Depth (calculated std. deviation)	C_v average = 40%

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

Type II (severe) alligatoring 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 between rating crews. Based on these limited observations it is fairly apparent, however, that the variability in measuring patching length is somewhat lower than for alligatoring with a C_v of perhaps 10-20%. A clear distinction between type I and type II alligatoring was not easily made by the rating crews. A tendency, except in the most obviously severe cases, was to place all cracking into the type I 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 critically rated.

The large amount of variability observed in the collected data is shown in Table 3. Considering the similarity in training and background between these "experimental" raters and previous inventory crews, these variabilities could be expected on pavement sections throughout the State.

Table 3 summaries 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%. It can be argued that the alligator cracking expressed as a percentage of section length need only be determined to be within 10% of the true percentage for inventory purposes. If it is assumed that the mean is the true value, then all 5 sections meet this criterion. More detailed measurements may be necessary for design processes.

The overall effect of crew measurement variations on rut depth determinations is magnified by the fact that the DOTPF usually reports "maximum" rut depth in terms of average plus 2 standard deviations.

Examples: The mean, mean + 1 standard deviation and mean + 2 standard deviations are given for the following sections:

*Section 1	mean = 0.090"	mean + s = 0.145"	mean + 2s = 0.200"
Section 2	0.240"	0.480"	0.720"
Section 3	0.150"	0.240"	0.330"
Section 4	0.180"	0.260"	0.340"

*Note--Using Table 3, outer wheel path data.

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

Table 3

Observed Variation in Pavement Distress Measurements

	Section #	Range	Average	Range as % of Section Length	Average as % of Section Length
Type I					
Alligator Cracking	1	29-187 ft.	140'	1-4 %	3
	2	35-1,434	820'	0-14 %	8
	3	69-700	300'	1-13 %	6
	4	54-218	120'	1-4 %	2
	5	190-505	340'	4-10 %	6
Type II					
Alligator Cracking	1	none detected	0'	0 %	0
	2	4-19 ft.	0'	0 %	0
	3	0-13	0'	0 %	0
	4	none detected	0'	0 %	0
	5	none detected	0'	0 %	0
Full Width Patching	1	350-382 ft.	360'	7-7.5 %	--
	2	439-1,042	820'	8-20 %	--
	3	91-197	100'	2-4 %	--
	4	none detected	0'	0 %	--
	5	none detected	0'	0 %	--

Table 3 (continued)

Observed Variation in Pavement Distress Measurements

	Section #	Range	Average	Range as % of Section Length	Average as % of Section Length
Rut Depth Average Inner Wheelpath	1	.016-.059 inch	.040"	--	--
	2	.110-.393	.210'	--	--
	3	.114-.289	.180"	--	--
	4	.100-.257	.170"	--	--
	5	.134-.271	.210"	--	--
Rut Depth Std. Deviation Inner Wheelpath	1	.005-.055 inch	.020"	--	--
	2	.051-.601	.160"	--	--
	3	.040-.198	.080"	--	--
	4	.023-.263	.080"	--	--
	5	.060-.241	.110"	--	--
Rut Depth Average Outer Wheelpath	1	.050-.167 inch	.090"	--	--
	2	.116-410	.240"	--	--
	3	.089-.248	.150"	--	--
	4	.062-.272	.180"	--	--
	5	.172-.445	.270"	--	--
Rut Depth Std. Deviation Outer Wheelpath	1	.022-.105 inch	.055"	--	--
	2	.069-.596	.240"	--	--
	3	.040-.322	.090"	--	--
	4	.032-.184	.080"	--	--
	5	.098-.257	.160"	--	--

Discussion of Alligator Cracking Measurements:

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 directly compared even though each has a different mean rut depth or length of alligatoring. Normalization of scoring, e.g., % alligatoring, average rut depth, etc., is accomplished as follows:

Normalized % alligatoring = (A-B)/C

where: A = % alligatoring as measured by an individual crew on a specific road section.

B = average % alligatoring calculated from the measurements of all crews on the above section.

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 5 pavement sections. This plot indicates the ability of certain crews, e.g., 7 and 8 to see more damage than others. Conversely, crew number 14 saw much less cracking in all 5 pavement sections than the calculated average. Figure 6 includes the instructor's subjective assessment of each crew in terms of: communication between crew members [rated: low (L), moderate (M) and high (H)], and initial impression of rating ability (rated: fair, good, expert). It should be noted 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 particularly accounted for in obvious attitudes or abilities. It is interesting to note, however, that crew 8, which saw much more pavement damage than crew 14, also rated higher in the instructor's opinion. It is suggested that best results are obtained when active conversation concerning the rating process is encouraged between crew members, especially during the first few days of inventory.

RANGE OF VARIATION IN EACH CREW'S MEASUREMENT OF TYPE I ALLIGATOR CRACKING (In Units of Standard Deviation)

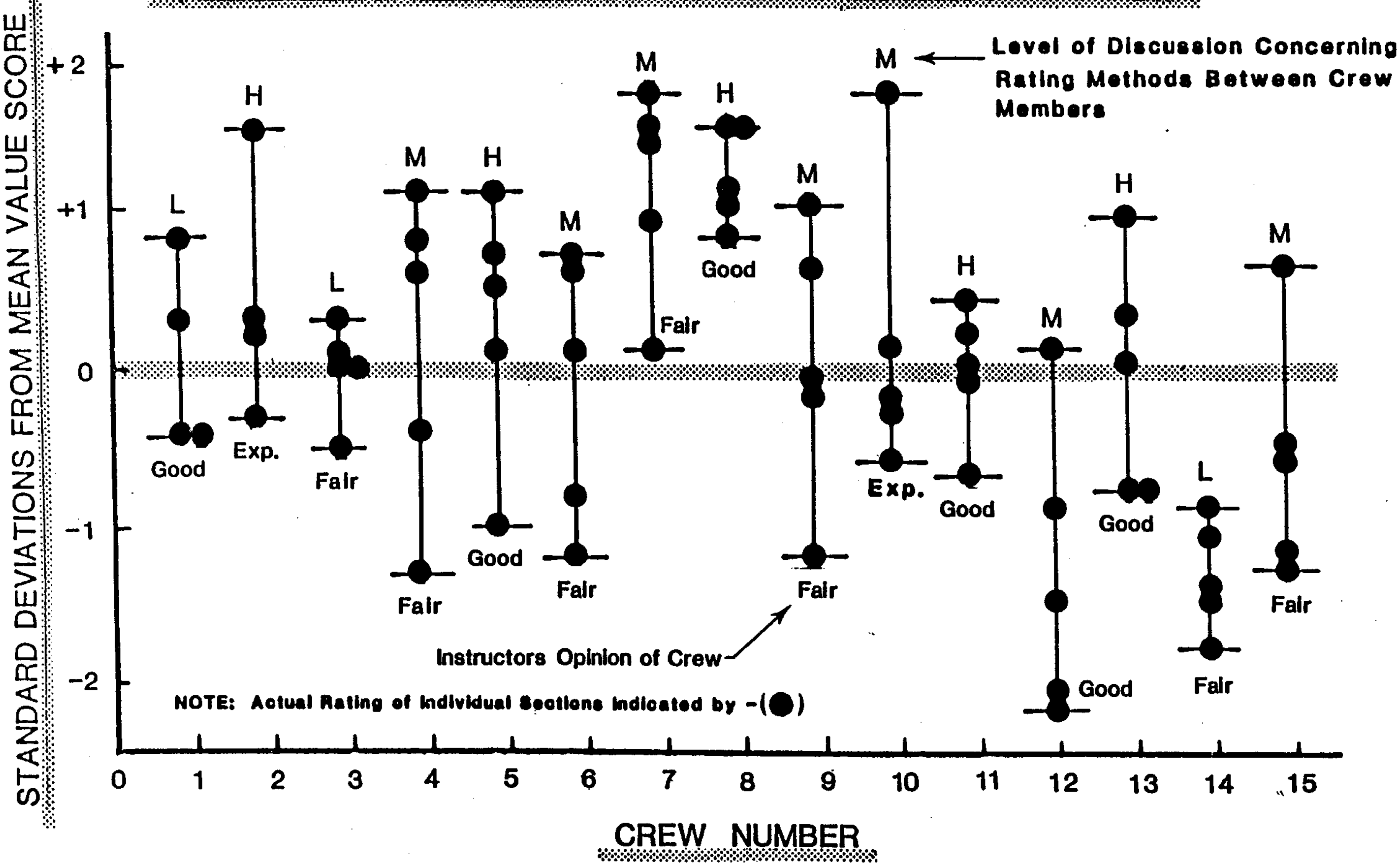


Figure 6

Table 4 attempts to delineate reasons for differences between crew ratings. The samples have been broken down into a stratified format and cross indexed in terms of crew communications and weather/pavement surface condition at the time of rating. The numbers shown in table 4 as X (characteristic sample average) and S (characteristic sample standard deviation) have been "normalized" as previously described, thus allowing all 5 pavement sections to be considered in the same analysis. The combination of a SW (slightly wet) pavement surface and a highly communicative crew resulted in more visible cracking with a characteristic average of +0.8 std. deviations 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 the most hairline alligatoring 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. It is strongly suggested that observations of cracking 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 a rain will keep the adjacent pavement wet longer than in areas of no cracking.

From the previous discussion, it would appear that pavement ratings could best be done shortly after a rainstorm. However, although a slightly wet surface allows highest crack visibility, a dry road condition represents the more normally encountered situation. Because of the need for a standard rating procedure it is strongly suggested that crack measurements be made only on dry pavement.

Sun Angle: Illumination effects due to variations in vertical and horizontal sun angle are known to strongly affect crack visibility. Experience indicates that optimal lighting conditions are provided by a more or less head-on sun incidence. Frontal light tends to shade and

ANALYSIS OF
TYPE I ALLIGATOR CRACKING

	S	C	SW	VW	WEIGHTED AVERAGE OF ROWS
H	$\bar{x}$ = -0.1 s = 0.5 n = 7	0.5 0.7 5	0.8 0.8 8	0.2 0.9 4	$\bar{x}$ = 0.4 s = 0.7
M	$\bar{x}$ = 0 s = 0.8 n = 7	0.3 1.1 19	-0.3 0.6 5	-1.7 0.5 4	$\bar{x}$ = -0.1 s = 0.9
L	$\bar{x}$ = 0.1 s = 0.6 n = 4	-0.7 0.8 10	No Samples	No Samples	$\bar{x}$ = -0.5 s = 0.7
WEIGHTED AVERAGE OF COLUMNS	$\bar{x}$ = 0 s = 0.6	$\bar{x}$ = 0 s = 1.0	$\bar{x}$ = 0.4 s = 0.7	$\bar{x}$ =-0.8 s = 0.7	

COMMUNICATIONS BETWEEN CREW MEMBERS	WEATHER CONDITIONS:
H = Active discussion of rating methods M = Moderate discussion of rating methods L = Little discussion of rating methods	S = Sunny C = Cloudy SW = Cloudy, Road Surface Slightly Wet VW = Rain, Road Surface Very Wet

Table 4

therefore darken the visible side of crack segments which 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 4 of the 5 pavements sections which is usually considered a worst-case viewing condition. However, each test section was examined at approximately the same time of day by each crew to assure a consistent sun angle.

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 patching measurement differences between crews may be about half those expected from observations of cracking. The distribution of normalized scores indicated in Figure 7 represent only the 3 test sections which actually contained patching. The variation between crews is markedly less pronounced than for alligator cracking.

It is concluded that patching is 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 quite easily seen. Observation conditions which 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 between crews, while 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 initially taken to determine a sample number (as indicated in Figure 4) was a rough attempt to limit the possibility of gross errors. Sufficient field data has since been collected to allow a much more valid estimation of rut depth.

As stated previously, 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?
2. Must rut depth measurements be taken in both inner and outer wheelpaths?

Sampling Frequency: Without becoming involved in a detailed explanation of statistical sampling methods and hypothesis testing, it can be stated that the sampling frequency must be high enough to insure (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 which would be measured from an infinitely large sampling. Sampling tables which are available in references such as the CRC 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 inch would be allowable.

In most sampling situations, little concern is expressed over type 2 errors. This philosophy leads to 50% level of type 2 error control, i.e., no control and a significantly reduced sample size.

Since determination of sample size is dependent upon expected standard deviation, it is important to consider the magnitude of values which might commonly be encountered. Rut measurements made on the 5 Fairbanks test sections indicated standard deviations ranging from about 0.02 to more than 0.35 inches associated with average rut depths between 0.02 and 0.40 inches. Indications of rut measurement variability derived from the Fairbanks test section data suggest that minimum sampling be based on a standard deviation perhaps as high as 0.30-0.35. This magnitude of deviation plus 90-95% confidence level against error results in a minimum sample size in excess of one hundred. 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 between 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 very large sampling. Problem rut depths on the order of 0.4-0.5 inch or larger would require an assumed standard deviation of at least 0.3 inch. Reasonable error confidence levels indicate a sampling obviously greater than 100 per 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 which appear from general observation to represent worst case conditions,
3. purchase or build an automatic rut measuring device as described in reference 5.

Since deeply rutted sections are usually associated with severe alligatoring on most Alaskan road sections, it would seem reasonable to suggest that the measurement of both is not necessary. It is further known that rutting within the state is rarely as deep as 1/2 inch 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. However, the numbers generated from alternative 2 have no basic statistical validity and might be thought of as inventory garbage. Alternative 3 is preferred if departmental policy requires an accurate rut depth determination. A 1981 cost estimate for the purchase of an automatic rut measuring device was 150,000-200,000 dollars.

A COMPARATIVE LOOK AT PREVIOUS INVENTORY DATA

This section looks at actual pavement inventory data in view of the preceeding findings of this report. Mindful 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 0-100 (worst-best). The reader should note that data at coordinates (0%, 0%) and (100%, 100%) are often repeated in figure 15 accounting for the appearance of fewer than expected individual points on plots of cracking and patching.

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

1) a very high degree of overall scatter

2) an unusually large number of data points lie 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 representing 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 scoring better than average in 1978 show a very high rate of apparent improvement with time. The likelihood of initially good pavements (scoring 50-100) being substantially improved within a period of three years through maintenance, etc., is slight.

COMPARISON OF 1978 PAVEMENT INVENTORY DATA WITH 1980/81 DATA

NOTE: All numbers have been normalized to a 0–100 (worst–best) scoring system

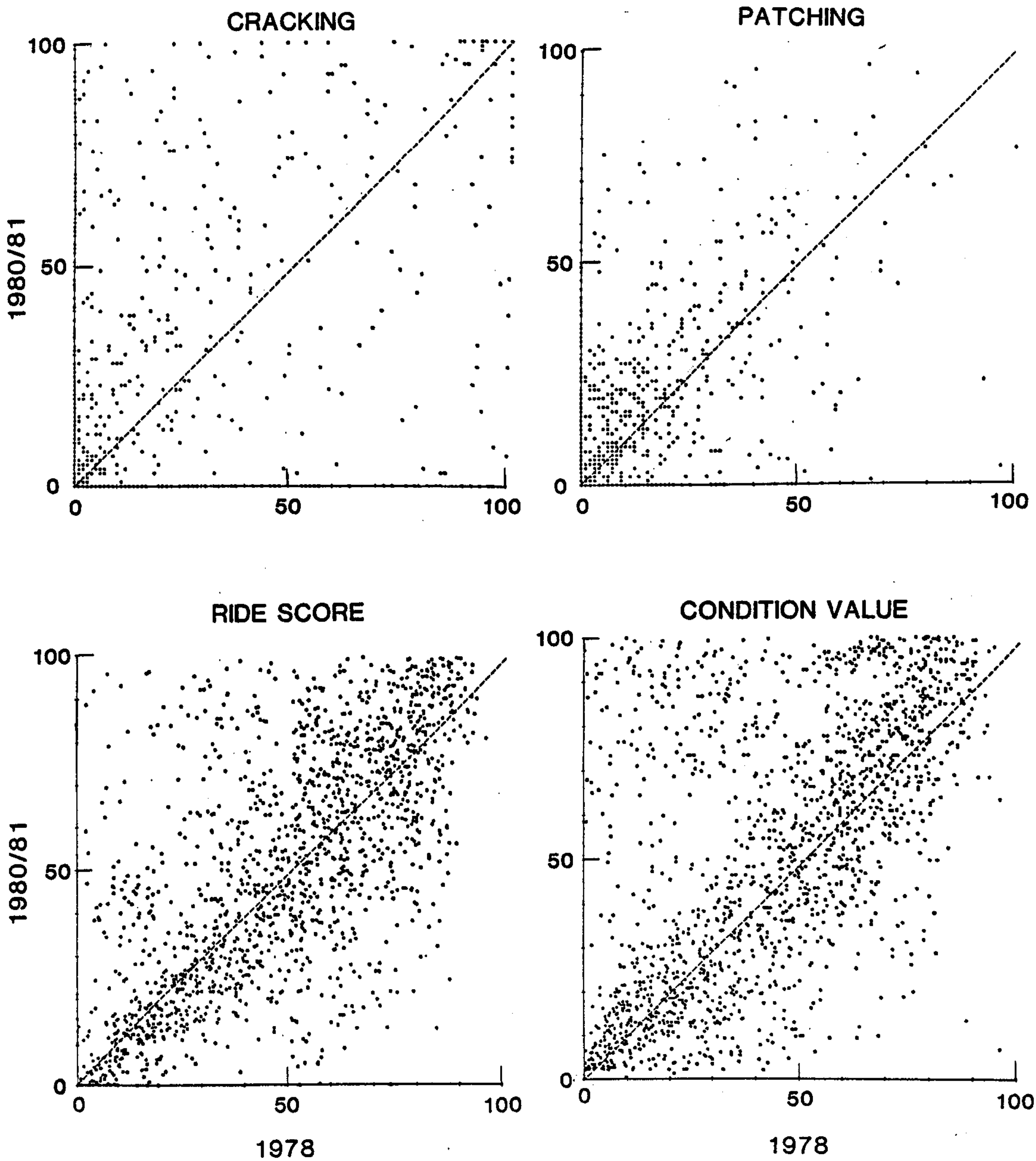


Figure 8

SUMMARY & CONCLUSIONS

This report describes the development and evaluation of Alaska's inventory rating procedure for flexible pavements.

Development of the system was based on the generally accepted principals 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
2) the highway engineer ---
   * fatigue (alligator) cracking
   * major (full lane width) patching
   * wheelpath rutting

These rating features are reported on a mile-by-mile summary both individually and in terms of a composite servicability score. As the rating method was being developed, a concerted effort was made to keep all distress measurements as simple as possible while providing adequate information for pavement management needs.

The rating method was evaluated through a special field study and experience accumulated during the three 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 task-interest expressed by each crew and weather factors controlling 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 is suggested which is capable of more than 100 measurements per section in both inner and outer wheelpaths. 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 insure a high degree of reproducibility is not demonstrated with 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 data collected from previous inventory work.
* The use of machine measurements is suggested wherever possible in all phases of the rating process.
* It is evident that pavement rating by "eyeball" methods is a difficult process requiring careful and rigorously standardized technique. Pavement rating instructions must be formalized to include guidelines for training rating crews and insuring acceptable performance. Specifications are necessary for standardization of viewing height, acceptable lighting conditions and vehicle speed.

RECOMMENDATIONS

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

Recommendations for improvement include:

1) Phase out human "eyeball" measurements of pavement distress as reliable machine methods become available.

2) Except for very rough classification purposes, rut measurements should be discontinued until sampling rates of more than 100 per mile can be achieved.

3) Continue existing approach but with greatly increased/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 performed by experienced personnel. The authors suggest a five-day "tuning" period for new rating crews. Ratings performed during this first week would not be included in the inventory summary before being verified by repeated observation.

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

1) vehicle speed at 6 mph or less
2) rating only completely dry road surfaces
3) utilization of optimal sun incidence whenever possible
    best illumination
    * horizontal sun angle of +-70 degrees from head-on
    * 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 at 5 1/2 feet $\pm$ 1/2 foot
5) use of utility van-type vehicle with nearly vertical windshield

REFERENCES

1 Highway Research Board: "The AASHO Road Test -- Report 5: Pavement Research", HRB, Special Report 61E, 1962.

2 Yoder, E.J. and M.W. Witczak: "Principles of Pavement Design", 2nd Edition, Wiley, 1975.

3 TRB: "Pavement Management Workshop", Committee Draft Reports, Transportation Research Board National Workshop on Pavement Management, "TUMWATER 77", Tumwater, Washington, November 1977.

4 The Chemical Rubber Company: "Handbook of Tables for Probability and Statistics", CRC Press Inc., 1, W.H. Beyer Editor, 2nd Edition, Cleveland, Ohio, 1968.

5 Jurick, R.: "Automated Pavement Rut Depth Measuring System", Alaska Department of Transportation and Public Facilities, Report No. AK-RD-80-1, 1981.

ACKNOWLEDGEMENT

On behalf of the Alaska Department of Transportation, the authors wish to acknowledge technical assistance and funding support provided by the Federal Highway Administration. The ADOT&PF 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.

The contents of this report reflect the views of the authors and not necessarily those of the State of Alaska or the Federal Highway Administration.