Factor Analysis of Pavement Distresses for Surface Condition Predictions

J. J. HAJEK AND R. C. G. HAAS

Pavement distress information is needed to assess maintenance requirements and to plan rehabilitation. For immediate maintenance requirements, it is necessary that the details of individual distress types, severity, and density be known. However, for pavement design and long-range rehabilitation planning, more approximate and aggregated data are sufficient. Furthermore, due to correlation between individual distresses, it is only practical to predict aggregated rather than individual detailed distresses. The Ontario Ministry of Transportation and Communications currently uses 15 pavement surface distress types (characterized by five levels of severity and five levels of density) in their condition survey procedures for flexible pavements. In this paper, the distress types are aggregated into five fundamental uncorrelated categories (factors) using factor analysis techniques. The five factors are thermal cracking, edge cracking, surface instability, fatigue cracking, and random cracking. Analyses are based on data observed on about 350 pavement sections. In general, the pavement structure of these sections consisted of asphalt concrete on top of granular materials. The results show that of the variance associated with the original 15 distress types about 60 percent could be explained by the five fundamental factors. Although these five basic uncorrelated factors should be used for future surface distress predictions, the 15 individual distress types will still be required for the selection of specific maintenance treatments and for establishing the existing values of the five factors.

Because pavements deteriorate with time, traffic, and climate, not only is investment in preservation through timely maintenance and rehabilitation important today, but also for the future. Surface distress manifestations and their quantification, and measurements of surface roughness, structural adequacy, and friction provide much of the information for determining present and future needs and for planning maintenance and rehabilitation.

Distress manifestations are defined as visible consequences of various mechanisms that usually lead to a reduction in pavement performance *(1)*. At present, the Ontario Ministry of Transportation and Communications (MTC) uses the 15 distress manifestations presented in Table 1 to visually characterize pavement condition and to calculate the distress manifestation index as a measure of pavement structural performance *(2)*.

Although each of the 15 distress manifestations describes a unique, or at least different, visual pattern or characteristic, all 15 distresses still describe the same general phenomenon, that of pavement deterioration. As subsequently shown, the distress manifestations are interrelated and many are statistically highly correlated.

The principal objective of the research reported herein was to investigate if it is possible to identify some fundamental categories of distress manifestations and thus to simplify the existing method of describing and analyzing pavement distresses. The original impetus for this work was a need to develop a simplified method for describing and predicting distress manifestations, and for identification of pavement failure modes that would provide an expert system for selection of pavement preservation treatments *(3)*. However, this paper is mainly concerned with how the simplified method of describing pavement distress manifestations can be used for prediction of pavement performance.

INTERDEPENDENCY OF DISTRESSES

In order to illustrate statistical dependence among the 15 distress manifestation characteristics, a correlation matrix of these distresses based on 347 observations representing nearly all pavement management sections in three MTC districts (Huntsville, Kingston, and Stratford) is presented in Table 1. The section lengths ranged from 0.3 to 25.7 km, with an average of 9.9 km. The sections were selected to exhibit a uniform pavement performance. The distresses were identified and rated by using the procedures described by Hajek et al. *(2)*, and were measured in terms of their density and severity on interval and ordinal scales, respectively, both scales ranging from 0 to 5 *(4)*. Final values of the distress manifestation variables were obtained by adding scaled values of density and severity. For example, if the severity of wheel track rutting was moderate (i.e., rutting depth was in the range of 12 to 19 mm) and its density was throughout (i.e., moderate rutting occurred on 80 to 100 percent of the section length), the severity was assigned a value of 3, the density a value of 5, and the final value was 8. This procedure was originally developed for calculation of the distress manifestation index *(2)*.

The coefficients of the correlation matrix for the 15 variables (Table 1) appear reasonable and as expected in both sign and magnitude. For example, variable B, flushing, is positively correlated with variable E, distortion ($r = 0.316$). This correlation suggests that excess asphalt, which results in flushing, also contributes to distortion. On the other hand, the lack of significant correlation between flushing and variable L (single and multiple transverse cracking), $r = -0.002$, suggests that flush-

---

J. J. Hajek, Ministry of Transportation and Communications, 1201 Wilson Ave., Downsview, Ontario, Canada M3M 1J8. R. C. G. Haas, Department of Civil Engineering, University of Waterloo, Waterloo, Ontario, Canada N2L 3G1.

TABLE 1 SIMPLE CORRELATION MATRIX OF MTC'S 15 DISTRESS MANIFESTATIONS

VARIABLE NAME	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
A Ravelling & Coarse Aggregate Loss	1.000														
B Flushing	-0.042	1.000													
C Rippling & Shoving	-0.113	0.132	1.000												
D Wheel Track Rutting	0.238	0.084	0.040	1.000											
E Distortion	0.066	0.316	0.310	0.313	1.000										
F Long. Wheel Track- Single & Multiple Cracking	0.232	0.189	0.059	0.459	0.426	1.000									
G - Alligator Cracking	0.010	0.025	0.096	0.204	0.194	0.319	1.000								
H Centreline - Single & Multiple Cracking	0.272	0.062	-0.085	0.343	0.243	0.468	0.145	1.000							
I - Alligator Cracking	0.155	0.032	0.095	0.153	0.207	0.245	0.291	0.224	1.000						
J Pavement Edge - Single & Multiple Cracking	0.218	0.099	0.137	0.300	0.256	0.302	0.225	0.334	0.190	1.000					
K - Alligator Cracking	0.120	0.060	0.220	0.202	0.321	0.252	0.179	0.141	0.081	0.340	1.000				
L Transverse - Full, Half & Multiple Cracking	0.251	-0.002	-0.077	0.400	0.101	0.419	0.101	0.586	0.122	0.292	0.059	1.000			
M - Alligator Cracking	-0.006	-0.029	0.021	0.041	-0.023	0.056	0.158	0.081	0.139	0.146	0.105	0.096	1.000		
N Longitudinal Meander & Midlane Cracking	0.096	0.204	0.177	0.331	0.404	0.537	0.213	0.464	0.249	0.366	0.178	0.430	0.203	1.000	
O Random Cracking	0.048	0.047	0.079	0.148	-0.006	0.239	0.149	0.090	0.085	0.185	0.176	0.159	0.222	0.173	1.000

Note: Correlation coefficients are based on 347 observations and are significant at the 1% level if their value exceeds approximately |0.15|.

ing does not have any effect on the formation of transverse cracks.

Detailed examination of the correlation matrix suggests that the distresses are highly correlated in many complex ways. For example, half, full, and multiple transverse cracking have statistically significant correlations with, among other variables:

- Wheel track rutting ($r = 0.40$),
- Single and multiple longitudinal wheel track cracking ($r = 0.42$),
- Single and multiple centerline cracking ($r = 0.59$), and
- Longitudinal meander and midlane cracking ($r = 0.43$).

Wheel track rutting has, in turn, statistically significant correlation with yet other distresses such as distortion ($r = 0.31$).

Variables that are highly correlated may actually measure similar characteristics and may be interchangeable to a certain degree. Also, correlation between variables masks specific influences of individual variables and often prevents their use in multiple regression models, particularly those obtained by stepwise regression.

A question then arises: What is the minimum number of distresses (variables) capable of describing distress manifestations in a concise, elucidative manner? Preferably, these variables should be uncorrelated and should attempt to identify fundamental distress categories.

The task of summarizing interrelationships among many variables in a concise basic manner can be tackled very effectively by factor analysis techniques *(5, 6)*. These techniques were first applied in the area of psychology in the early 1900s. Since the 1940s, factor analysis has been used in many other fields such as sociology, medicine, business, and transportation planning *(7)*. However, it appears that this technique has realized limited, if any, application to pavement technology.

FACTOR ANALYSIS TECHNIQUES

The basic factor model represents variables as additive composites of several weighted factors or loadings using a set of linear equations called the total factor pattern:

$$\begin{aligned} Z_1 &= a_{11}F_1 + a_{12}F_2 + \ldots a_{1m}F_m + b_1S_1 + e_1E_1 \\ Z_2 &= a_{21}F_1 + a_{22}F_2 + \ldots a_{2m}F_m + b_2S_2 + e_2E_2 \\ &\ldots \\ Z_n &= a_{n1}F_1 + a_{n2}E_2 + \ldots a_{nm}F_m + b_nS_n + e_nE_n \end{aligned} \tag{1}$$

where, using index j to designate variables and index i to designate individuals (observations):

Z_j = Observed variable, total number of variables being n.
F_j = New uncorrelated components called common factors. The total number of common factors is m, which is usually much smaller than n. F_j is a factor common to all variables.
S_j = Specific factor.
E_j = Error factor; $S_j + E_j$ values are called unique factors.
a_{ji} = Common factor coefficients; $a_{ji}F_j$ is the contribution of factor F_j to the linear composite.
b_j = Specific factor coefficient.
e_j = Error factor coefficient.

The factor analysis model resembles regression analysis insofar as a variable is described as a linear combination of another set of variables plus a residual. However, in regression analysis this set of variables (i.e., the set of independent variables) are observable quantities, whereas in factor analysis they are hypothetical constructs that can only be estimated from the observed data *(6)*. The important property of factors as hypothetical constructs is that they are independent (and uncorrelated) even though the original variables themselves may be related.

The objective of factoring, which constitutes the basic part of factor analysis, is to find the coefficients of the factor pattern a_{ji}, b_j, and e_j. Because a system of orthogonal (uncorrelated) factors consistent with observed data and satisfying Equation 1 may be chosen in an infinite number of ways, the coefficients of the factor pattern cannot be uniquely determined. Consequently, many factor analysis techniques have been developed to extract factors and to transform the extracted factors in order to obtain the factor solutions that are most amenable to interpretation.

The factoring operation is usually done on dimensionless standardized values of variables Z_j for all individuals i. The standardized variables have zero means and variances equal to unity. Using the notation of Equation 1, the composition of the unit variance of variable Z_j is defined as

$$\sum_{i=1}^{m} a_{ji}^2 + b_j^2 + e_j^2 = 1 \qquad (2)$$

where

$\sum_{i=1}^{m} a_{ji}^2$ = communality of variable Z_j, defined as the proportion of the unit variance explained by m common factors;
b_j^2 = specificity of the observed variable Z_j; and
e_j^2 = error variance associated with variable Z_j.

Because e_j^2 is usually unknown, the sum $b_j^2 + e_j^2$ represents the uniqueness of the observed variable, that is, that proportion of the unit variance unexplained by common factors.

Factor analyses were performed using the SAS computer program package *(8)*. Three extraction techniques were investigated:

- Principal component analysis,
- Principal factor analysis, and
- Maximum-likelihood factor analysis.

Detailed results are reported only for the principal component analysis because the results obtained by this technique appeared to permit the best interpretation of the resulting factor solution. The factors obtained by this technique yield the best least-square estimates of the entire correlation matrix and each succeeding common factor accounts for the maximum obtainable amount of variation in the correlation matrix *(9)*. In other words, the factors are selected in a stepwise manner.

The indeterminacy of factor solutions may result in factors with loadings (i.e., a_{ji} coefficients) that may be difficult to interpret. This situation can be radically improved by rotating the axes of the reference frame on which the factors are measured. Because the objective of this research was to obtain simplified factors and to maintain orthogonality of the factors, the varimax method of rotation was used. The varimax method alters individual factor coefficients, as well as the variance explained by each factor, whereas the total variance explained by the rotated and the original factor patterns remains unchanged. The method also strives for a simplified factor solution by making the small-factor loadings approach zero and the large-factor loadings approach unity *(9, 10)*.

DATA ANALYSIS

Factor analysis was carried out on distress manifestation data obtained for three MTC districts. The procedure used for collecting the distress data was referenced previously; their correlation matrix was presented in Table 1. The three MTC districts were selected with the intention of obtaining a representative sample of distress manifestations associated with conventional flexible pavements (asphalt concrete on top of granular material) subjected to a variety of traffic and environmental exposures. Nearly all asphalt concrete pavements (about 3,000 km, centerline) on the King's highways in the three districts were included in the study.

The frequency of occurrence of distress manifestations in the three districts is illustrated in Figure 1 using the final values (sums of scaled values of density and severity) of the 15 distress variables. The most frequent distress was transverse cracking—half, full, and multiple—which occurred, in one form or another, on about 90 percent of all sections. On the other hand, alligator transverse cracking occurred on only 2 percent of the sections.

The factor solution obtained by principal component analysis and rotated using the varimax method is shown in Table 2. Table 2 (Part a) gives the rotated factor pattern for only five factors, or principal components, and thus gives results for a truncated component solution. The remaining 10 factors (there are 15 variables) were not included in the solution based on the Mineigan criterion, which states that the variance explained by a factor included in the solution must be at least equal to unity *(8)*. The number of factors required to represent the 15 variables was also tested using maximum-likelihood analysis. The probability level based on the chi-square test for the hypothesis of five factors being sufficient was 0.026; for that of six factors the probability was 0.500. This difference indicates that the five-factor model provides an appropriate representation of the data.

The entries of the factor pattern in Table 2 (Part a) are the a_{ji} coefficients of Equation 1. They are also regression coefficients between variables Z_j and factors F_m. These entries, which are referred to as factor loadings, theoretically can range from −1 to +1. The further the factor loading for a given variable is from 0, the more one can generalize from that factor to the variable.

Final communality estimates of the 15 variables explained by the five factors are given in Table 2 (Part b). These estimates, multiplied by 100, yield the percentage of the variance for a given variable, explained by the factors. With the exception of variables D and J (wheel track rutting and pavement edge single and multiple cracking), the factors explained more than 50 percent of the variance. The best results were obtained for distortion (68.9 percent) and for pavement edge alligator cracking (68.7 percent).

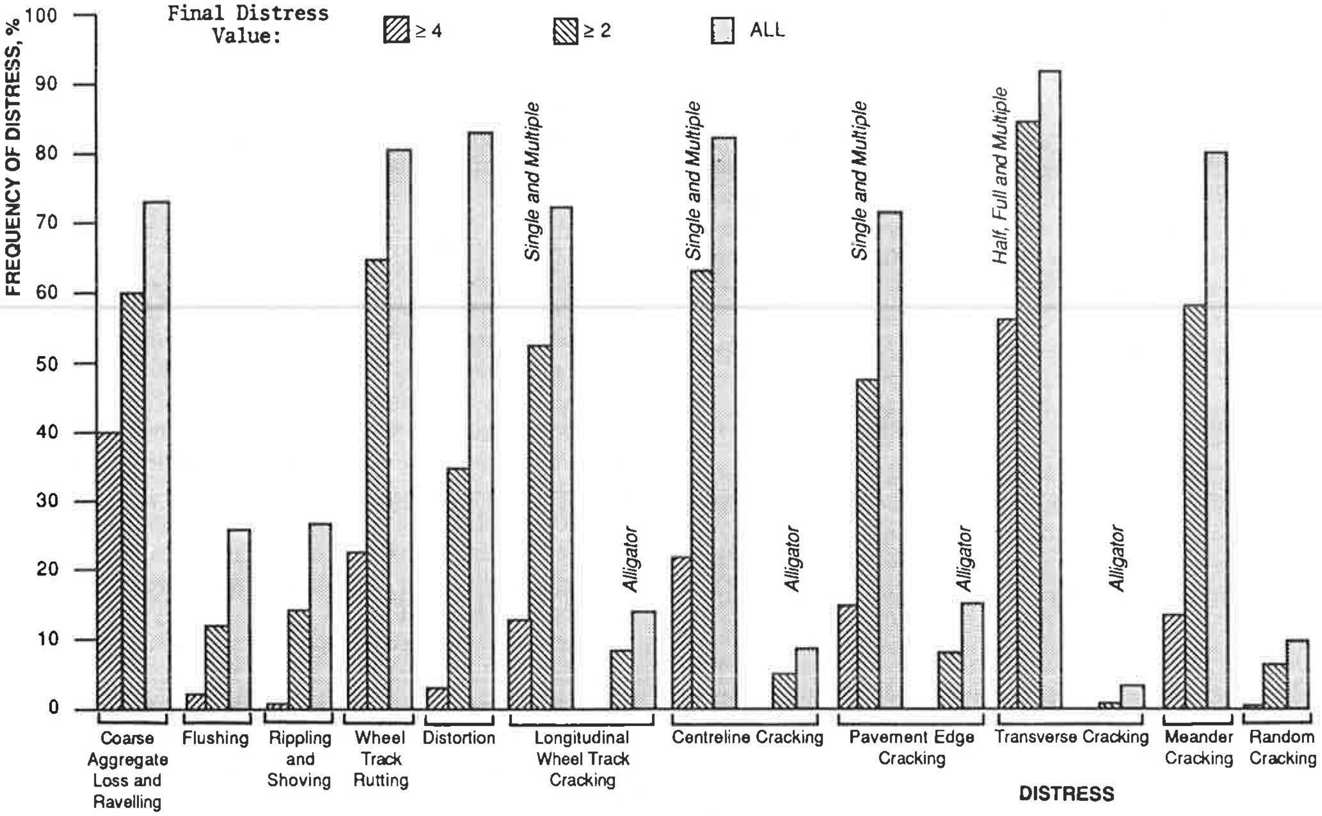


FIGURE 1 Occurrence of distress manifestations.

TABLE 2 VARIMAX SOLUTION FOR PRINCIPAL COMPONENT ANALYSIS

Part (a) Rotated Factor Pattern						Part (b) Final Communality Estimates[1]
	FACTORS					
VARIABLE NAME	1	2	3	4	5	
A Ravelling & Coarse Aggregate Loss	0.443	0.335	-0.455	0.036	-0.250	0.579
B Flushing	0.146	-0.018	0.719	-0.101	-0.049	0.551
C Rippling & Shoving	-0.214	0.448	0.492	0.143	0.062	0.553
D Wheel Track Rutting	0.595	0.284	0.054	0.095	-0.032	0.448
E Distortion	0.281	0.390	0.582	0.252	-0.236	0.689
F Long. Wheel Track - Single & Multiple Cracking	0.670	0.196	0.255	0.247	0.056	0.617
G - Alligator Cracking	0.094	0.143	0.049	0.708	0.184	0.567
H Centreline - Single & Multiple Cracking	0.796	-0.003	-0.008	0.115	0.018	0.648
I - Alligator Cracking	0.173	0.017	-0.019	0.802	-0.022	0.674
J Pavement Edge - Single & Multiple Cracking	0.399	0.527	0.009	0.131	0.180	0.486
K - Alligator Cracking	0.072	0.815	0.057	0.040	0.113	0.687
L Transverse - Full, Half & Multiple Cracking	0.805	-0.060	-0.080	-0.039	0.154	0.683
M - Alligator Cracking	0.038	-0.020	-0.048	0.231	0.731	0.592
N Longitudinal Meander Midlane Cracking	0.614	0.066	0.411	0.219	0.225	0.649
O Random Cracking	0.157	0.239	0.007	-0.069	0.703	0.581

Part (c) Variance Explained by Each Factor						
	FACTORS					
	1	2	3	4	5	Total[2]
Variance Explained (Eigenvalues)	3.035	1.613	1.556	1.448	1.311	8.963
Percent of Total Variance	20.2	10.8	10.4	9.7	8.7	59.8

[1] One minus communality gives variable uniqueness.

[2] Total unit variance for 15 observations is equal to 15.

Table 2 (Part c) presents the variance explained by each factor. These variances, which are equal to factor eigenvalues, can be calculated as

$$\sum_{j=1}^{15} (a_{jm})^2 \qquad (3)$$

For example, the standardized variance for Factor 1 was equal to 3.035 and represented about 20 percent of the total variance. (The total variance for the 15 standardized variables is equal to 15). The variance explained by the five factors of 8.93 accounted for about 60 percent of the total variance of the original 15 variables.

This 60 percent of the total variance explained by the five factors is reasonable, but not as large as may be desired, indicating the following:

1. Additional variables should be included in the analysis to better define the common factors.
2. The variables have a high degree of uniqueness.

Because uniqueness consists of a specific variance and an error variance in unknown proportions, it may be interpreted in many ways. For example, by attributing the main portion of uniqueness to the specific variance, the results suggest that many distress manifestation variables tend to measure unique specific pavement deterioration properties not common to other variables. On the other hand, by attributing the main portion of uniqueness to the error variance, the results indicate the presence of substantial measurement errors. The most likely interpretation is somewhere between these two extremes. The presence of these errors in the pavement distress rating process has been documented before *(11)*.

At any rate, the construction of five new, uncorrelated, hypothetical variables, which contain 60 percent of the information previously transported by the 15 original correlated variables, is a significant accomplishment, particularly in view of the purpose of the work, which was to provide an aggregated, approximate, and practical basis for predicting distresses.

The effect of varimax rotation is graphically illustrated in Figure 2, which shows a plot of the factor pattern in the common-factor space of two dimensions represented by Factors 1 and 2. The actual factor space is five-dimensional. The plot is an example of several possible projections of this space to a cartesian coordinate system.

The varimax rotation strives to simplify factor solution by concentrating variable loadings on as few factors as possible. Considering for example variable H (centerline single and multiple cracking), its loadings before rotation were 0.67 on Factor 1 and 0.40 on Factor 2. After the rotation, the corresponding loadings were 0.80 and 0.

INTERPRETATION OF FACTORS

An important step in factor analysis is interpretation of factor meanings. Because factors are unobservable hypothetical variables, their identification and interpretation are based upon observations of which variables are, and which are not, related to the factors and what these relationships may conceptually indicate. In other words, possible quantitative distinctions based on statistical results as well as qualitative distinctions based on intuitive evaluation of the underlying physical phenomena are sought.

The basic guide for factor interpretation is provided by the coefficients (factor loadings) of the factor pattern. These coefficients, as regression coefficients between factors and variables, reflect the importance of the factors in predicting the observed variables. To facilitate the interpretation, a list of salient factor loadings is given in Table 3. The salient (or prominent) factor loadings were defined as correlation coefficients equal to or greater than 0.4. This definition guarantees that the correlations are statistically significant ($P < 0.025$) even after making allowances for the possibility of increased errors of orthogonally rotated factor solutions *(9)*.

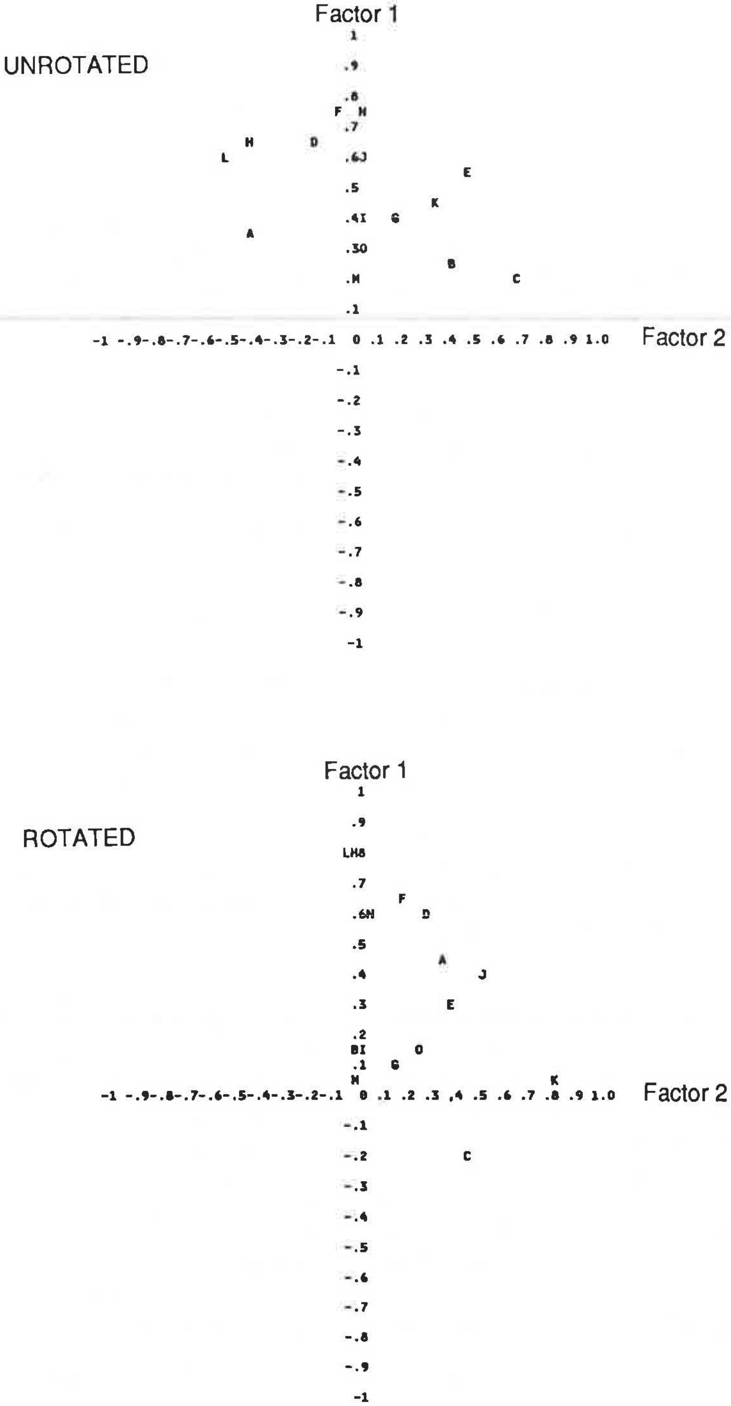
The main motivation for applying the technique to pavement distress data is its potential to explain the relationship among many variables in terms of more basic concepts. These concepts should, in turn, help to explain the overall problem. Pavement damage in Ontario occurs because of the following four basic damage attributes related to load and environment:

1. Traffic loads,
2. Temperature changes,
3. Moisture effects, and
4. Construction flaws due to materials and construction techniques.

Ideally, for explanatory purposes, one should strive to construct the factors so that each factor combines and isolates the effects of only one of the damage attributes. However, all damage attributes act on the pavement structure simultaneously in many complex ways. For example, a frost heave is the result of combined damage attributes of temperature changes, moisture effects, and construction flaws. On the top surface, the resulting pavement damage is aggravated by traffic loads. The frost heave itself would be described by the MTC distress survey in terms of distortion and perhaps also in terms of several different types of cracking. For these reasons, the ideal explanatory solution, based on the factor loadings of Table 3, is difficult to achieve.

By way of introduction to interpreting and naming factors, Factor 4 is considered first. This factor has only two salient (statistically significant) factor loadings and both of them are with variables describing alligator cracking: longitudinal wheel track alligator cracking (0.71) and centerline alligator cracking (0.80, Table 3). The remaining 13 variables do not have statistically significant loadings on Factor 4. Also, the two variables with salient loadings on Factor 4 do not have salient loadings on any other factor. Thus, Factor 4 can be considered a factor describing alligator cracking. However, the 15 original variables contain two additional variables describing other types of alligator cracking not included in Factor 4: pavement edge alligator cracking and transverse alligator cracking.

The first variable associated with alligator cracking—transverse alligator cracking—is rare (it occurs on less than 4 percent of the sections shown in Figure 1) and was included, as discussed later, in Factor 5. The second variable, pavement edge alligator cracking, loads strongly only on Factor 2 (0.81) and thus appears to be unrelated to the two types of alligator



LEGEND
A,B,C...O are variables defined in Tables 1 to 4.

FIGURE 2 Graphical representation of varimax rotation.

cracking associated with Factor 4. It may be hypothesized that Factor 4 describes alligator cracking caused by pavement structural fatigue, whereas Factor 2 describes alligator cracking caused by insufficient strength of the pavement edge typical for pavements with thin asphalt concrete surfaces. This assumption is supported by the next highest loading for Factor 2, which is single and multiple pavement edge cracking. The two types of pavement edge cracking (single and multiple, and alligator) are related to each other and with Factor 2, but are unrelated to fatigue cracking described by Factor 4. Factor 4 was thus named fatigue cracking and Factor 2 was named pavement edge cracking.

Factor 1 has high salient or statistically significant loadings on all four variables describing different types of single and multiple cracking (longitudinal wheel track, centerline, pavement edge, and transverse) and was named thermal cracking

TABLE 3 SALIENT FACTOR LOADINGS

VARIABLE NAME	Factor 1 Thermal Cracking	Factor 2 Edge Cracking	Factor 3 Surface Instability	Factor 4 Fatigue Cracking	Factor 5 Random Cracking
A Ravelling & Coarse Aggregate Loss	0.44		-0.45		
B Flushing			0.72		
C Rippling & Shoving		0.45	0.49		
D Wheel Track Rutting	0.60				
E Distortion		0.40	0.58		
F Long. Wheel Track- Single and Multiple Cracking	0.67				
G - Alligator Cracking				0.71	
H Centreline - Single and Multiple Cracking	0.80				
I - Alligator Cracking				0.80	
J Pavement Edge - Single and Multiple Cracking	0.40	0.53			
K - Alligator Cracking		0.81			
L Transverse - Full, Half & Multiple Cracking	0.81				
M - Alligator Cracking					0.73
N Longitudinal Meander Midlane Cracking	0.61		0.41		
O Random Cracking					0.70

[1] For factor loadings higher than 0.40.

factor. Evidently, single and multiple pavement edge cracking contributes both to the thermal cracking factor (Factor 1) with a loading of 0.40 and to the pavement edge cracking factor (Factor 2) with a loading of 0.53. The variance of this variable must then be subjectively divided for interpretive purposes. Because the two factors are orthogonal, it may be hypothesized that one portion of the single and multiple pavement edge cracking is related to the same causes as, for example, the transverse cracking and centerline cracking of Factor 1 that have been linked in the past to high asphalt concrete stiffness at low temperature and to other causes *(12)*. The other portion of the single and multiple pavement edge cracking would then be hypothesized to have the same causes as Factor 2, that is, insufficient strength of the pavement edge. The thermal cracking is also associated with raveling and coarse aggregate loss (loading of 0.44). This association appears to indicate that the low asphalt content and stripping, usually associated with raveling, also contributes to general cracking *(13)*.

Factor 3 has the highest positive loading on flushing (0.72) and the highest negative loading on raveling and coarse aggregate loss (–0.45). The two variables are related in the opposite directions. Flushing is associated with the absence of raveling and coarse aggregate loss, whereas the raveling and coarse aggregate loss tend to be associated with the thermal cracking factor (loading of 0.44). Other salient loadings of Factor 3 are for the variables rippling and shoving, distortion, and longitudinal meander and midlane cracking. For this reason, Factor 3 was named surface instability factor. Pavement damage caused by frost heaves is usually described in terms of distortion and longitudinal meander and midlane cracking. Both these variables are associated with Factor 3.

Longitudinal meander and midlane cracking is associated with both the surface instability factor (loading 0.41) and with the thermal cracking factor (loading 0.61). This association suggests, quantitatively, how the cause for this distress manifestation may be proportioned.

Factor 5 explained the lowest amount of variance (Table 2, Part c) and was the most difficult to interpret. Because it had high loadings from transverse alligator cracking (0.73) and from random cracking (0.70), it was named "random cracking" factor. The name was selected because the occurrence of transverse alligator cracking is low compared even to that of random cracking, as shown in Figure 1. Transverse alligator cracking usually develops from a single transverse crack by formation of additional parallel cracks alongside the original crack and by subsequent gradual formation of an alligator pattern. It is probably caused by all four basic damage attributes. Transverse alligator cracking was only marginally related to the fatigue cracking factor (0.231 in Table 2, Part a).

In summary, the following factors were identified. They are listed in order of their contribution in explaining the sample variance, together with their possible causes:

Factor	*Factor Name*	*Possible Principal Cause*
1	Thermal cracking	Temperature changes, traffic loads
2	Pavement edge cracking	Effect of load on thin AC pavement
3	Surface instability	Construction flaws (frost heaves)
4	Fatigue cracking	Traffic load
5	Random cracking	All causes

The factors can be viewed as uncorrelated fundamental operational representatives of all 15 distress manifestations. Because the factors are orthogonal, information provided by any given factor is uncorrelated and independent of that provided by any other factor or factors.

Factor analysis techniques lie somewhere between a science and an art *(14)*. Different analysts may use different techniques that yield somewhat different results and different researchers may interpret even the same results differently. Therefore, the results should be viewed as only one of the possible interpretations of statistical data. Nevertheless, they can also be explained in term of qualitative reasonableness of the basic physical phenomena involved.

FACTOR SCORES

The factors as hypothetical constructs can be also be expressed in terms of the 15 observed variables. It is thus possible to characterize, quantify, and predict pavement distress manifesta-

TABLE 4 SCORING COEFFICIENTS

VARIABLE NAME	Factor 1 Thermal Cracking	Factor 2 Edge Cracking	Factor 3 Surface Instability	Factor 4 Fatigue Cracking	Factor 5 Random Cracking
A Ravelling & Coarse Aggregate Loss	0.151	0.283	-0.396	-0.012	-0.260
B Flushing	0.061	-0.133	0.517	-0.161	-0.018
C Rippling & Shoving	-0.185	0.277	0.271	0.041	0.027
D Wheel Track Rutting	0.190	0.119	-0.030	-0.045	-0.082
E Distortion	0.021	0.153	0.316	0.093	-0.241
F Long Wheel Track- Single & Multiple Cracking	0.201	-0.019	0.113	0.057	-0.020
G - Alligator Cracking	-0.093	-0.017	-0.048	0.539	0.061
H Centreline - Single & Multiple Cracking	0.307	-0.129	-0.034	-0.018	-0.036
I - Alligator Cracking	-0.045	-0.115	-0.091	0.652	-0.112
J Pavement Edge - Single & Multiple Cracking	0.061	0.323	-0.095	-0.037	0.078
K - Alligator Cracking	-0.098	0.602	-0.093	-0.103	0.030
L Transverse - Full, Half & Multiple Cracking	0.336	-0.152	-0.060	-0.149	0.091
M - Alligator Cracking	-0.049	-0.093	-0.035	0.119	0.563
N Longitudinal Meander & Midlane Cracking	0.189	-0.145	0.252	0.028	0.134
O Random Cracking	0.006	0.125	-0.015	-0.187	0.551

tions in terms of the five fundamental uncorrelated factors rather than in terms of the 15 highly correlated original variables.

Table 4 gives a matrix of scoring coefficients that can be used to calculate factor scores S_{ki} for individual observations from the following formula:

$$S_{ki} = \sum_{j=1}^{n} (s_{kj} \cdot Z_{ji});\ k = 1, 2, \ldots 5 \tag{5}$$

where

S_{ki} = factor score for factor k and observation i;
s_{kj} = scoring coefficient for factor k and variable j;
Z_{ji} = value of variable j for observation i; and
n = number of variables, 15 in this case.

The matrix of scoring coefficients can be roughly approximated by dividing factor loadings by factor eigenvalues. The factor scores estimated by the scoring coefficients are also uncorrelated for all practical purposes *(8)*. These scores can be used to characterize and quantify the state of visual pavement deterioration instead of the original 15 variables. For example, the factor score for the general cracking factor (Factor 1) is

$$S_{1i} = 0.151 \cdot Z_{Ai} + 0.061 \cdot Z_{Bi} - 0.185 \cdot Z_{Ci} + \ldots + 0.006 \cdot Z_{Oi} \tag{6}$$

where subscripts $A, B, C, \ldots O$ refer to variables (distress manifestations) defined in Tables 1 to 4, and subscript i identifies an observation. Thus, for example, Z_{Ai} is a value of raveling and coarse aggregate loss for observation i, and Z_{Bi} is a value for flushing for the same observation.

The concept of characterizing, quantifying, and ultimately predicting visual pavement distresses in terms of factor scores is examined first for Factor 4. Factor 4 was named fatigue cracking and its two variable loadings are unique: longitudinal wheel track alligator cracking and centerline alligator cracking have salient, or statistically significant, factor loadings only on Factor 4. Furthermore, both variables measure the same phenomenon of cracking and are measured on the same scale. Thus, this type of fatigue cracking can be predicted independently of virtually any other distress manifestation. This prediction, of course, has been a common practice in the past; the present results provide further statistical justification for the practice.

Factor 1, named thermal cracking, is affected by two variables that do not provide unique loadings (raveling and coarse aggregate loss and pavement edge single and multiple cracking) as well as by heterogeneous variables—variables that are measured on different scales and that generally measure different deterioration phenomena such as rutting and cracking. Overall, Factor 1 loadings indicate that it is not possible to examine and predict the associated variables in isolation. For example, raveling and wheel track rutting both contribute to longitudinal wheel track cracking, transverse single and multiple cracking, and to other distresses. The prediction of any one of these distresses (rutting, for example) is questionable without taking into account the influence of the other distresses. Similarly, considering Factor 3 and its salient loading, it appears unattainable to predict roughness (in terms of distortion, rippling, and shoving) without taking into account flushing and longitudinal meander and midlane cracking.

The prediction of factor scores overcomes this limitation because the scores encompass the contributions of all relevant variables and provide uncorrelated fundamental measures of pavement visual distresses. Also, if the factor scores are known (or predicted), the scoring coefficients of Table 4 can be used to calculate values of the individual pavement distress variables.

CONCLUSIONS AND RECOMMENDATIONS

1. The evaluation of visible pavement distresses benefits from the application of factor analysis techniques. Factor analysis techniques, which can identify fundamental and uncorrelated categories of pavement distress manifestations, are useful in explaining and investigating relationships between highly correlated pavement distresses. As with any statistical procedure, factor analysis is a valuable tool only if used correctly. In addition, because of the skills required for factor interpretation, the successful user must have an in-depth technical knowledge of the problem domain.

2. Principal component analysis identified five basic independent factors that together explained about 60 percent of the variance previously explained by 15 observed variables. These five factors were named thermal cracking, pavement edge cracking, surface instability, fatigue cracking, and random cracking.

3. The results of factor analysis helped to quantify statistically associations between various distresses (such as raveling and cracking or distortion and flushing) in terms of more basic concepts. However, because of the interaction between the pavement damage attributes (e.g., traffic loads and temperature changes) and pavement structures, it was not possible to isolate the effects of the individual damage attributes (expressed as visible pavement surface distress manifestations) and assign them only to specific factors.

4. Due to the complex interdependency of individual pavement distresses, predictions of pavement distresses should be done in terms of the fundamental factors (i.e., factor scores) rather than in terms of the traditional distresses. The prediction of pavement distresses in terms of five factor scores will

provide sufficiently detailed data for the selection and timing of pavement rehabilitation treatments and for life-cycle economic analysis. For planning purposes, the separate prediction of pavement distresses in terms of the 15 pavement distress manifestation variables appears to be both impractical (because too many prediction models would be required) and logically incorrect (because distresses are interdependent).

5. All 15 distress manifestation variables currently used to characterize visible pavement deterioration contribute significantly to one or more of the fundamental factors and should be retained. Also, the individual distress variables are required for the selection of specific maintenance treatments and for establishing the existing values of the five factors.

ACKNOWLEDGMENTS

The authors express their appreciation to members of the MTC regional geotechnical sections for providing the pavement condition data that made this report possible. Appreciation is also extended to W. A. Phang, Manager of Pavement and Roadway Research, MTC, and to G. W. Bennett, Statistical Consulting Service, University of Waterloo, for many valuable comments and suggestions.

REFERENCES

1. W. R. Hudson, F. N. Finn, R. D. Pedigo, and F. L. Roberts. *Relating Pavement Distress to Serviceability and Performance.* Report FHWA-RD-80-098. FHWA, U.S. Department of Transportation, July 1980.
2. J. J. Hajek, W. A. Phang, G. A. Wrong, and G. M. Scott. *Pavement Condition Index (PCI) for Flexible Pavements.* Report PAV-86-02. Ontario Ministry of Transportation and Communications, Downsview, Ontario, Canada, July 1986.
3. J. J. Hajek. *Life-Cycle Pavement Performance Modelling Using a Knowledge-Based Expert System Technology.* Unpublished report. Department of Civil Engineering, University of Waterloo, Waterloo, Ontario, Canada.
4. B. G. Hutchinson. Principles of Subjective Rating Scale Construction. In *Highway Research Record 46,* HRB, National Research Council, Washington, D.C., 1964, pp. 60–70.
5. S. A. Mulaik. *The Foundations of Factor Analysis.* McGraw-Hill Book Company, New York, 1976.
6. H. H. Harman. *Modern Factor Analysis.* The University of Chicago Press, Chicago, Ill., 1968.
7. G. E. Mouchahoir. The Use of Multivariate Statistics in Transportation Planning. *Traffic Engineering,* Oct. 1971, pp. 44–48.
8. SAS Institute Inc. *SAS User's Guide: Statistics.* Version 5 Ed., SAS Institute, Cary, N.C., 1985.
9. R. L. Gorsuch. *Factor Analysis.* 2nd ed., Lawrence Erlbaum Associates, Inc., Hillsdale, N.J., 1983.
10. R. J. Harris. *A Primer of Multivariate Statistics.* Academic Press, Inc., New York, 1975.
11. R. L. McHattie and B. G. Connor. Description and Evaluation of Alaska's Pavement Rating Procedure. In *Transportation Research Record 938,* TRB, National Research Council, Washington, D.C., 1985, pp. 61–72.
12. J. J. Hajek and R. C. G. Haas. Predicting Low Temperature Cracking Frequency of Asphalt Concrete Pavements. In *Transportation Research Record 407,* TRB, National Research Council, Washington, D.C., 1972, pp. 39–54.
13. V. J. Marks and C. L. Huisman. *Reducing the Adverse Effects of Transverse Cracking.* Final Report, Iowa Department of Transportation, Ames, Jan. 1985.
14. L. L. Dickinson and J. W. Hall. Factor Analysis of Pedestrian Accidents. In *Transportation Research Record 605,* TRB, National Research Council, Washington, D.C., 1986, pp. 35–41.

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