Identifying Error-Generating Factors in Infrastructure Condition Evaluations

FRANNIE HUMPLICK

Infrastructure surface inspection and condition rating systems used today range from detailed automated inspections that use photographic and laser technologies to manual inspection that uses the human eye. The capability of these systems in measuring distressed areas varies because of several factors, including the principle of measurement, type of inspection strategy, manner of data reduction, and objectivity of data collection. The characteristics of the objects being measured and the surroundings in which they are inspected also affect the results. The types of errors affecting inspection results are presented, as is a set of hypotheses derived from theoretical expectations of the effect of the mentioned factors on the accuracy of inspection systems. These hypotheses are tested using data from state-of-the-art inspection systems. The conclusions are useful for designing, improving, and choosing systems and for adjusting inspection results for improved accuracy.

A variety of infrastructure inspection systems currently exists (*1–8*). These systems range from detailed automated inspections using photographic and laser technologies to manual inspection using the human eye. The capabilities of such systems in locating, recognizing, discriminating, and distinguishing among distresses, as well as scaling their size, extent, and severity, depend on a variety of factors. These include the principle of measurement, type of inspection strategy, and manner of data collection and reduction. Inspection results are also affected by the characteristics of the objects being measured, which create confounding measurement scenes for the inspection systems and hence limit their accuracy. Finally, the surroundings of the measured objects also affect inspection.

This paper discusses the types of error affecting the results of inspection and presents a set of hypotheses derived from theoretical expectations of the effect of characteristics on the accuracy of inspection systems. These hypotheses are then tested using data from a FHWA study entitled *Improved Methods and Equipment to Conduct Pavement Distress Surveys* (*4*). This data set will be referred to as the FHWA data set.

CLASSIFICATION OF ERRORS OF INSPECTION

A typical inspection process consists of the facility under inspection and the inspection system. Inspection errors originate from the inspected facility and inspection system, as well as the interface between them.

Center for Construction Research and Education, Massachusetts Institute of Technology, Cambridge, Mass. 02139. Current affiliation: Infrastructure and Urban Development Division, The World Bank, 1818 H Street N.W., Washington, D.C. 20433.

Intrinsic or Inherent Errors

Intrinsic errors are inherent in the inspected facility and the inspection system. They can be observed in laboratory or experimental conditions when all known influencing factors are controlled. In such situations the same object measured repeatedly by an inspection system can result in almost the same measured quantity or in highly varying measured quantities.

The first case is characteristic of most mechanical gauges, such as roughness measurements on highway pavements, where the required response from the inspection system is well known and can be determined. The second situation occurs when the response from the inspection system is not well behaved—that is, the results of measurement in the second situation are so varied that one cannot predict the underlying true value of the object without further knowledge about the distribution of the measured values or the causes of the discrepancies. The latter case is the most common in infrastructure condition evaluation, because the inspection systems used have multiple components, some of which are not well tested or designed. It has been observed that repeated measurements by the same system are highly variable, and the measured values by different systems of the same sections are even more variable (*4*).

The following forms of error can cause differences in observed results. These have been adapted from a work by Finkelstein (*9*) and are generalized to account for the inspection systems common in infrastructure condition evaluations:

1. Zero error occurs when the inspection system outputs a value even when there is no event present. For example, if a video inspection of an undistressed pavement results in a fixed or variable value of measured distress each time it is used, it is an indication of zero error. That is, the video inspection will always result in a value of distress even when it is not present. The analogy to a gauge-type measurement is that the gauge will have a misplaced zero position, reading a fixed value even before it has been applied in measurement.
2. Dynamic error results during the operation of an inspection system such as inspecting pavement surfaces with truck-mounted photographic equipment. The error in data recording with respect to location on the pavement is considered a dynamic error. The measured value of distress in this case is a function of the speed of the data acquisition (e.g., the shutter speed of a camera) and the speed of the operating vehicle.
3. Quantization or categorization errors result from inspection systems in which the measurement response changes

in discrete steps. For example, digitizing intensity values on a piece of film for future processing, by grouping observations into ranges of intensity values and assigning a value to each range, can result in a quantization error. This error type also appears in visual inspections in which distresses are categorized into ranges, resulting in a categorization error.

4. Viewing limitation errors occur when an inspection system can view only a fixed portion of the inspected facility or when detection capability is limited to a certain range. For example, a photographic camera has a fixed field of view and resolution that can be achieved in practice, limiting the size of the facility surface that can be viewed and the size of the objects detected. An analysis of the impacts of such error types can be found elsewhere (*10*).

Influence Errors

Influence errors arise during interaction between the inspection and inspected systems. They are caused by factors in the inspection environment that were not controlled in an experimental setting during the design of an inspection system. For example, a film of water on a pavement surface being inspected after a rainy period can change the reflectance property of the pavement surface. The system may not be designed to account for such a change.

Such errors can arise at the output interface and include the data reduction format used and whether individual, range estimates, or average values are reported. Alternatively, factors at the input interface such as the sampling strategy, characteristics of the inspected facility (the nature of the measured surface), and properties of the measured objects (pattern of occurrence or dimensions of the objects) may cause errors. Finally, uncontrolled influences due to departure from design or calibrated conditions can result in influence errors.

Uncontrolled influences affecting the results of inspection arise from the inspection environment. These include (a) mechanical vibrations during data collection affecting the relative position of measuring equipment; (b) electrical and thermal changes influencing the behavior of measuring equipment; (c) events occurring on the inspected facility that were not planned for or are not part of the measurement, such as shadows, oil spots, and debris, which can confound automated systems; and (d) fatigue due to long hours of operation of equipment or humans, inducing measurement errors.

Intrinsic and influence errors have a systematic component and a random component. Systematic errors are errors that can be predicted from past knowledge or use of an inspection system. For example, if it is known that a human inspector tends to add a fixed amount to all measurements, the measurements can be corrected by this amount, which can be obtained from past observations. Random measurement errors, on the other hand, are those that cannot be predicted on an individual measurement basis but that can be statistically estimated from multiple measurements. These errors are due to short-term variations of factor influencing measurement. For example, if an inspection system is repeatedly measuring the same object using the same procedures, the scatter of the measured value, caused by temporal and local variations in influences, is a realization of random measurement errors.

Formulation of Measurement Problem

The difference between the result of measurement and the true value of the measured quantity is an error of measurement. This error can be defined as ε_t where

$$\varepsilon_t = d - d^* \tag{1}$$

and d and d^* represent the measured and true values, respectively.

The true value is the quantity that would be obtained from a perfect measurement. Because all measurements are subject to error, the true value is essentially unknown (latent). For calibration purposes, an approximation obtained from a measurement deemed close enough to the true value for all practical purposes is often used. For example, inspection systems of moderate accuracy are calibrated against systems of high accuracy, by using the measured quantities by the high accuracy system as the "true values" [see Jeyapalan, Cable, and Welper (*11*)].

Let us denote the values of the error-generating factors that an inspection system encounters during measurement by θ. Measuring an object with a true value d^* with these settings of the error-generating factors results in

$$d = f(d^*, \theta) \tag{2}$$

If we denote the values of the error-generating factors for which an inspection system has been designed and calibrated for by θ^r, then the influence error can be defined as the difference between the resulting measurements from the actual and design settings. This can be expressed as follows:

$$\varepsilon_{\text{inf}} = f(d^*, \theta) - f(d^*, \theta^r) \tag{3}$$

where ε_{inf} measures the departure from design conditions.

From the definition of total error in equation 1, and denoting the error due to the measuring principle as ε_m, we have

$$\varepsilon_t = \varepsilon_{\text{inf}} + \varepsilon_m$$

These errors are additive, because they are assumed independently of each other; that is, removing the effect of an influence error (such as the resolution limitation of a camera) does not result in a change in the error due to the measurement principle, which remains essentially the same (indirect). Hence, the effect of the influence errors and the errors due to the measurement principle are additive.

A measurement principle can be direct or indirect. In direct measurement the inspection system receives as input properties of the measured object (e.g, the length and width of a crack) and gives as output a measure of these properties (e.g., length in feet). Visual inspection by human beings is a form of direct measurement. In indirect measurement the inspection system receives as input the properties of the measured object, which it senses depending on the measurement principle employed, and converts the information into signals that represent a proxy for these properties. These signals are then processed. The processing transforms the sensed proxy into measures or properties of the original object through some mapping function, which then outputs the measured value.

A rule system relates the numerical value obtained as output to the value of the properties of the object input. Transferring intensity values on a piece of film to measures of distress is an example of such mapping.

This process leads to two components of errors due to measuring principle: data acquisition and data processing. In general, a measurement process can be represented as

$$d^* \rightarrow \delta \rightarrow d$$

where δ is a proxy of the actual object realized as an intermediate step. Thus, the data acquisition error can be expressed as

$$\delta = h_1(d^*) + \bar{\varepsilon}_{da} + \bar{\varepsilon}_{\inf} \quad (4)$$

and the data processing error can be expressed as

$$d = h_2(\delta) + \varepsilon_{dp} \quad (5)$$

where h_1 (.) is a transformation mapping the true distress value into a proxy (the measured value of distress) and $h_2(.)$ maps the measured proxy of distress into the inspection output.

Substituting Equation 4 into Equation 5 gives

$$d = h_2[h_1(d^*) + \bar{\varepsilon}_{da} + \bar{\varepsilon}_{\inf}] + \varepsilon_{dp}$$

Assuming without loss of generality that the mapping from d^* to δ is linear,

$$h_1(d^*) = \alpha + \beta d^* = x \qquad \text{then}$$

$$h_2(\delta) = \left(\frac{\delta - \alpha}{\beta}\right) \qquad \text{and}$$

$$d = \frac{h_1(d^*) + \bar{\varepsilon}_{da} + \bar{\varepsilon}_{\inf} - \alpha}{\beta} + \varepsilon_{dp}$$

$$= \frac{\alpha + \beta d^* - \alpha + \bar{\varepsilon}_{da} + \bar{\varepsilon}_{\inf}}{\beta} + \varepsilon_{dp}$$

$$= d^* + \frac{\bar{\varepsilon}_{da}}{\beta} + \varepsilon_{dp} + \frac{\bar{\varepsilon}_{\inf}}{\beta}$$

Defining the following,

$$\varepsilon_{da} = \frac{\bar{\varepsilon}_{da}}{\beta}$$

$$\varepsilon_{\inf} = \frac{\bar{\varepsilon}_{\inf}}{\beta}$$

we obtain

$$\varepsilon_t = d - d^* = \varepsilon_{da} + \varepsilon_{dp} + \varepsilon_{\inf} \quad (6)$$

where $\varepsilon_m = \varepsilon_{da} + \varepsilon_{dp}$, and ε_{da}, ε_{dp}, and ε_m are data acquisition, data processing, and measurement principle errors, respectively.

Not all the errors in Equations 1 through 6 can be determined determined empirically. Particularly, the error due to data acquisition in Equation 5 cannot be ascertained because it is related to an intermediate value whose relationship to the original input is not well studied. For photographic inspection of pavement surfaces, the relationships between the intensity values on a piece of film and the true values of distress on a surface are not well studied. Procedures to identify and classify these intensity values in terms of distresses occurring on a surface are still in an experimental stage (*6*). Therefore it is not possible to determine the data acquisition error for such systems from analytical deductions.

A generalized measurement error model specified from the error structures derived thus far is expressed as

$$d_{ijk} = f(d^*_{ik}, \theta_i, \theta_j, \theta_k) \quad (7)$$

where

d_{ijk} = measured distress on Section i of Distress Type k by Inspection System j;
$f(.)$ = function representing the relationship between the measured distress, the true value of distress, and factors affecting the measurement;
d^*_{ik} = unobserved true value of distress of Type k on Section i; and
θ_i, θ_j, and θ_k = vectors representing error-generating factors from the inspection environment (section), inspection system (technology), and the measured objects (distresses), respectively.

Without loss of generality, we can express the function $f(.)$ in Equation 7 in a linear form with respect to the true distress, as is commonly done for calibration purposes.

$$d_{ijk} = \alpha_{jk} + \beta_{jk} d^*_{ik} + \varepsilon_{ijk} \quad (8)$$

where α_{jk}, β_{jk}, and ε_{ijk} are the systematic additive, systematic multiplicative, and additive random error of Inspection System j while measuring distress of Type k in Section i.

HYPOTHESIS DEVELOPMENT AND TESTING

The function in Equation 8 was estimated for a variety of inspection systems, distress types, and pavement sections using the FHWA data set (*4*). Details of the estimation can be found in unpublished data by Ben-Akiva and Humplick. Empirical results from the estimation are used in this section to investigate the effects of error-generating factors on measurements results.

The FHWA data set included measurements by seven inspection systems measuring distresses in experimental units consisting of three pavement types (flexible, composite, and rigid) and three condition levels (good, moderate, and poor). The inspection systems are a manual mapping method, detailed visual surveys using manual recording, automated data logging, the GERPHO device (Photo1), the PASCO Roadrecon survey vehicle (Photo2), the ARAN survey vehicle (Video), and the Laser RST device (Laser). These systems will be referred to as "Mapping," "Manual," "Logging," "Photo1," "Photo2," "Video," and "Laser." For a detailed description of these systems and the type of measurements they performed, see the work by Hudson et al. (*4*).

Hypotheses About Inspection System Characteristics

Two hypotheses were developed to test for the impact of inspection system characteristics on measurement accuracy.

Effects of Inspection Strategy and Data Reduction Format

The manner in which data is reduced and the percentage area of pavement inspected are expected to affect measurement results. An inspection system that either views less than 100 percent of the pavement surface or reports average values or ranges of distress values for a given section is expected to be less accurate than one that views 100 percent of the section and measures individual distress elements. The inspection systems used in the FHWA study can be grouped as

- Total area observed, individual measures of distress made, or both (Manual, Photo1, and Photo2); and
- Sample area observed and range or average values reported (Mapping, Logging, Video, and Laser).

The distinction is made because the inspection systems falling into the second group require some kind of estimate to obtain the total value of distress on a section. The types of errors in the inspection results of the second group that are not present in those of the first group may be due to extrapolating from a small sample size, or averaging by eye. To test whether there is a difference between the inspection results of the systems in the two groups, the following hypothesis is set up

$$H_0 : \beta_{\text{Man}} = \beta_{\text{Photo1}} = \beta_{\text{Photo2}} \quad \text{and}$$

$$\beta_{\text{Map}} = \beta_{\text{Log}} = \beta_{\text{Video}} = \beta_{\text{Laser}} \tag{9}$$

Effect of Data Collection Process

The type of data collection process employed, whether objective or subjective, is expected to affect the accuracy of the results of measurement. In particular, inspection systems making objective measurements are expected to have smaller random biases than those based on subjective rankings, unless they suffer from interpretation problems; then the random biases would be large. For example, objective measures of alligator cracking obtained by a photographic technique should have less variation than those obtained from an eye estimate by a human inspector. However, one expects more classification errors to affect the photographic technique, because there is no visual verification of the types of distress present. Subjective evaluations are expected to measure the extent of distress less accurately, but they suffer less from interpretation and classification errors. Inspection systems using both subjective and objective measures should, therefore, have more accurate results.

From the descriptions of the inspection systems in the FHWA study, they can be grouped as follows:

- Objective: Mapping, Photo2;
- Subjective: Manual, Logging; and
- Objective and subjective: Photo1, Video, Laser.

The Photo1, Video, and Laser used both subjective and objective measures to estimate the level of distress on a section. The following hypotheses are set up:

$$H_0 : \beta_{\text{Man}} = \beta_{\text{Log}}$$

$$\beta_{\text{Map}} = \beta_{\text{Photo2}}$$

$$\beta_{\text{Photo1}} = \beta_{\text{Video}} = \beta_{\text{Laser}} \tag{10}$$

The hypotheses in Equations 9 and 10 can only be tested for inspection systems in which all other factors affecting measurement accuracy are similar or insignificant.

Table 1 shows the organization of the FHWA data according to the factors mentioned. From this table we can test for the effect of the data reduction format and inspection strategy by comparing Manual to Logging and Photo1 to Video. We can test the data collection process by comparing Photo1 and Photo2 and Mapping to Logging.

A paired Tukey test was selected to perform multiple comparisons of the estimated bias parameters. The Tukey test is constructed as follows: assume the multiplicative biases $\hat{\beta}_j$ for the inspection systems $j = 1, \ldots, J$ are distributed with a mean $\bar{\beta}$ and variance σ^2_β. The range of the $\hat{\beta}_j$s is

$$R = \max_j \hat{\beta}_j - \min_j \hat{\beta}_j \tag{11}$$

Let s^2_β have an estimator of σ^2_β having ν degrees of freedom, and assume s^2_β and $\hat{\beta}_j$ are independent. Then $Q_{J,\nu} = R/S$, where J is the number of inspection systems being compared and S, the standard error of the βs, is Student t distributed. The confidence interval for the $(\hat{\beta}_j - \hat{\beta}_l)$, where $j \neq l$, taking into account that all possible comparisons can be made, is given by Larsen and Marx (*12*) and Box et al. (*13*).

TABLE 1 INSPECTION SYSTEM CHARACTERISTICS IN FHWA DATA

Inspection system (j)	Inspection System Characteristics: Measurement Principle	Data Reduction & Inspection Strategy	Data Collection Process
Mapping	-	+	O
Manual	-	-	S
Logging	-	+	S
Photo1	+	-	C
Photo2	+	-	O
Video	+	+	C
Laser	+	+	C

NOTE:
Measurement principle: - = direct; + = indirect
Data reduction & Inspection strategy: - = individual/total; + = average/range/sample
Data collection process: O = objective; S = subjective; C = combined

$$(\hat{\beta}_j - \hat{\beta} l) \pm \frac{q_{J,v,\alpha/2}}{\sqrt{2}} s \sqrt{\frac{1}{n_j} + \frac{1}{n_l}} \quad (12)$$

where

$\frac{q_{J,v,\alpha/2}}{\sqrt{2}}$ = a tabulated upper significant value of the Studentized range for J variables and v degrees of freedom.

One would reject the hypothesis that the parameters are equal if zero is not contained within this interval. The Tukey test was performed using the estimated values of β_j in Table 2. The results of the Tukey test are shown in Table 3.

The hypothesis on the equality of the multiplicative biases was rejected for all pairs of inspection systems except Mapping and Manual, Mapping and Logging, Mapping and Video, Mapping and Photo1, Manual and Video, and Logging and Photo1. The effect of the data reduction format and inspection strategy captured by the difference ($\beta_{Man} - \beta_{Log}$) and ($\beta_{Photo1} - \beta_{Video}$) was found significant, as the hypothesis that these differences are zero was rejected. Similarly, the effect of the data collection process, represented by the difference ($\beta_{Photo1} - \beta_{Photo2}$), was found significant. However, it was found insignificant for ($\beta_{Map} - \beta_{Log}$). This discrepancy may be because the Photo1 and Photo2 inspection systems employ photographic imaging techniques with the same measurement principle, so the effect of the data collection process is more pronounced than when Mapping and Logging are compared. They are both direct measurement technologies that use human inspectors, but they differ extremely in the manner in which data is actually collected: Mapping uses a sampling strategy and measures each individual distress on a sample unit to get an estimate of distress on the section, whereas Logging observes the entire section but gives a range estimate of distresses on the section.

Pairwise differences were computed for the random bias parameters using the results in Table 2. The Tukey interval was estimated, and the results are shown in Table 4. This test resulted in rejecting 6 out of 15 pairwise differences. The pairs

TABLE 2 ESTIMATED BIASES FOR DIFFERENT INSPECTION SYSTEMS (ALLIGATOR CRACKING ON FLEXIBLE PAVEMENTS)

Inspection system (j)	Estimated Parameters (standard errors of the estimates) $\hat{\alpha}_j$ (Sqft)	$\hat{\beta}_j$	$S.D.(\epsilon_{ij}) = \sqrt{\psi_j}$ (Sqft)	Coeff. of det. R^2
1. Mapping	-73.0 (474.2)	0.83 (0.17)	396.9	0.94
2. Manual	37.5 (363.7)	0.49 (0.10)	262.9	0.94
3. Logging	570.0 (845.1)	1.29 (0.54)	646.2	0.61
4. Photo1	-154.5 (527.0)	1.09 (0.21)	444.5	0.95
5. Photo2	-501.3 (551.6)	1.85 (0.23)	551.3	0.99
6. Video	134.0 (474.2)	0.44 (0.17)	472.3	0.77
7. Laser[a]	---	---	---	---

a The Laser inspection system did not report measures for alligator cracking as denoted by --- in the table.

TABLE 3 TUKEY TEST FOR EQUALITY OF MULTIPLICATIVE BIASES (ALLIGATOR CRACKING ON FLEXIBLE PAVEMENTS)

Hypothesis Tested	Results of Test
$\beta_{map} = \beta_{man}$	0.34 Accept
$\beta_{map} = \beta_{log}$	-0.46 Accept
$\beta_{map} = \beta_{photo1}$	-0.26 Accept
$\beta_{map} = \beta_{photo2}$	-1.02 Reject
$\beta_{map} = \beta_{video}$	0.39 Accept
$\beta_{man} = \beta_{log}$	-0.80 Reject
$\beta_{man} = \beta_{photo1}$	-0.60 Reject
$\beta_{man} = \beta_{photo2}$	-1.36 Reject
$\beta_{man} = \beta_{video}$	0.05 Accept
$\beta_{log} = \beta_{photo1}$	0.20 Accept
$\beta_{log} = \beta_{photo2}$	-0.56 Reject (not significant)
$\beta_{log} = \beta_{video}$	0.85 Reject
$\beta_{photo1} = \beta_{photo2}$	-0.76 Reject
$\beta_{photo1} = \beta_{video}$	0.65 Reject
$\beta_{photo2} = \beta_{video}$	1.41 Reject
95% Tukey Interval	±0.57
99% Tukey Interval	±0.71

TABLE 4 TUKEY TEST FOR PARAMETER EQUALITY— RANDOM BIAS (ALLIGATOR CRACKING ON FLEXIBLE PAVEMENTS)

Hypothesis Tested	Results of Test
$\sigma^2_{map} = \sigma^2_{man}$	99.5 Accept
$\sigma^2_{map} = \sigma^2_{log}$	-480.1 Reject
$\sigma^2_{map} = \sigma^2_{photo1}$	76.3 Accept
$\sigma^2_{map} = \sigma^2_{photo2}$	423.3 Reject
$\sigma^2_{map} = \sigma^2_{video}$	-200.8 Accept
$\sigma^2_{man} = \sigma^2_{log}$	-380.6 Reject
$\sigma^2_{man} = \sigma^2_{photo1}$	175.8 Accept
$\sigma^2_{man} = \sigma^2_{photo2}$	522.8 Reject
$\sigma^2_{man} = \sigma^2_{video}$	-101.3 Accept
$\sigma^2_{log} = \sigma^2_{photo1}$	556.4 Reject
$\sigma^2_{log} = \sigma^2_{photo2}$	903.4 Reject
$\sigma^2_{log} = \sigma^2_{video}$	279.3 Accept
$\sigma^2_{photo1} = \sigma^2_{photo2}$	347.0 Accept
$\sigma^2_{photo1} = \sigma^2_{video}$	-277.1 Accept
$\sigma^2_{photo2} = \sigma^2_{video}$	624.1 Reject
95% Tukey Interval	±422.25
99% Tukey Interval	±508.11

σ^2_j is the variance of measurement by inspection system j.

of inspection systems whose random biases are statistically different are

- Mapping and Logging,
- Mapping and Photo2,
- Manual and Logging,
- Manual and Photo2,
- Logging and Photo1,
- Logging and Photo2,
- Logging and Video, and
- Photo2 and Video.

This indicates that the Logging and Photo2 inspection systems have random biases that are significantly different in nature from those of the other inspection systems. For the Photo2 system this may be because alligator cracks and other crack types were jointly reported. For Logging, the difference may be due to the averaging of range estimates of distress that are reported instead of the individual values of distress on a section. However, more pairs of random biases were found statistically equal (9 out of 15 hypotheses were accepted) than were pairs of multiplicative biases (3 out of 15 hypotheses on equality of pairs of biases were accepted). A possible explanation for this difference is that the impact of factors affecting measurement results may more seriously affect multiplicative biases than random biases. This is a useful finding, because one can correct the measured results for systematic biases using the results of a calibration and hence only worry about minimizing random error.

The results of these tests indicate that the inspection systems used in the FHWA data had varying capabilities in measuring alligator cracking (represented by the multiplicative bias β_j). The most distinct inspection systems were Logging and Photo2; their parameters were statistically different from those of the other systems. The main differences between these inspection systems and the others is the limitations of the Logging device, which cannot measure individual distresses and reports ranges of distress instead, and of the Photo2 system, which jointly measures alligator and other areal distresses such as block cracking and patched cracks.

Hypotheses About Distress Characteristics

The distress characteristic that could be tested using the FHWA data set is the dimension of distress, mainly whether linear, area, or volumetric. Table 5 compares the estimated multiplicative biases β_j for distresses with these dimensions on flexible pavements. measures of volumetric distresses showed a range in values (0.34 to 2.09) larger than the ranges of linear (0.43 to 1.48) and areal (0.44 to 1.85) distresses. In general there was an increase in the range of parameter estimates as the number of distress dimensions increased, with the lowest range being for the case of linear distresses. This indicates that the inspection systems are measuring areal and volumetric distresses in a dissimilar way as compared to linear distresses. These differences are mainly due to the additional complexity of the measurement scene when volumetric distresses are involved.

A Tukey test on the equality of the estimated multiplicative bias parameters when measuring different distress types was performed. This test involves the comparison of 7 × 7 = 49 pairs of parameters for each distress dimension, which leads to 49 × 3 = 147 pairs. Only the differences between the same inspection systems for each distress dimension are of interest. These are presented in Table 6, in which they are compared to the Tukey intervals at 95 and 99 percent confidence.

The hypothesis on the equality of parameters was accepted for all the differences between multiplicative biases for linear versus areal distresses and rejected for all differences for linear versus volumetric distresses. This indicates that there is an effect of distress dimension on the results of measurement.

The effect of distress dimension is statistically significant especially for the inspection systems employing optical techniques (Photo1, Photo2, and Video). This result is expected, because the complexity of measurement due to distress dimension is supposed to affect optical techniques more than techniques (such as inspection by humans) that do not depend

TABLE 5 HYPOTHESES ABOUT DISTRESS CHARACTERISTICS (VARIOUS DISTRESSES ON FLEXIBLE PAVEMENTS)

	Estimated Multiplicative Biases (standard errors of the estimates)		
Inspection System (j)	Longitudinal and Transverse Cracking (linear)	Alligator Cracking (areal)	Potholes and Patches (volumetric)
1. Mapping	0.95 (0.26)	0.83 (0.17)	--
2. Manual	--	0.49 (0.10)	0.61 (0.36)
3. Logging	1.48 (0.65)	1.29 (0.54)	0.37 (0.15)
4. Photo1	0.43 (0.40)	1.09 (0.21)	2.09 (0.20)
5. Photo2	--	1.85 (0.23)	1.59 (0.12)
6. Video	1.17 (0.64)	0.44 (0.17)	0.34 (0.16)
7. Laser	0.98 (0.45)	--	--
Range in estimates of β_j(Max − Min)	(0.43 - 1.48) 1.05	(0.44 - 1.85) 1.41	(0.34 - 2.09) 1.65

-- denotes no parameter estimates for the given technology.

TABLE 6 TUKEY TEST FOR EFFECTS OF DISTRESS CHARACTERISTICS

	Results of Hypothesis Test		
Hypothesis Tested	Linear Vs. Areal	Linear Vs. Volumetric	Areal Vs. Volumetric
$\beta_{map} - \beta_{map}$	0.12 Accept	--	--
$\beta_{man} - \beta_{man}$	--	--	-0.12 Accept
$\beta_{log} - \beta_{log}$	0.19 Accept	1.11 Reject	0.92 Reject
$\beta_{photo1} - \beta_{photo1}$	-0.66 Accept	-1.66 Reject	1.00 Reject
$\beta_{photo2} - \beta_{photo2}$	--	--	0.26 Accept
$\beta_{video} - \beta_{video}$	0.73 Accept	0.83 Reject	0.10 Accept
$\beta_{laser} - \beta_{laser}$	--	--	--
Tukey 95% confidence interval	±0.81	±0.54	±0.51
Tukey 99% confidence interval	±1.02	±0.66	±0.65
Ratio of number rejected	0/4	3/3	2/5

– denotes no parameter estimates for the given technology pair.

on a signal (such as intensity of light) coming from the measured object. The nonoptical inspection systems (Mapping, Manual, and Logging) showed no significant difference between the estimated multiplicative biases.

Hypotheses About Section Characteristics

Section characteristics are captured by three factors: the pavement type, the contrast between distresses and their background, and the pattern of distress occurrence. In the FHWA data there were three types of pavement: rigid, flexible, and composite. Rigid pavements can be categorized as having high contrast and a systematic pattern of distress occurrence. Flexible and composite pavement can be categorized as having moderate to low contrast and a haphazard pattern of distress. To test for the joint effect of the contrast and pattern of distress occurrence, one can use the pavement type as a proxy.

The following general hypothesis can be stated on the basis of these section characteristics:

> Inspection systems are equally efficient in detecting and measuring distresses (capability) but distresses differ in their "detectability." That is, if a distress is in a section with high contrast and a systematic pattern of distress occurrence, it is more easily detectable by a given inspection system than if it is in a section with low contrast and a haphazard pattern of distress.

Therefore, the contrast and pattern of distress occurrence characterize the detectability, and capability is represented by the estimated measurement biases α_j and β_j.

The hypothesis that can be tested is whether there is a difference in inspection system biases when measuring distresses from backgrounds with different contrast and pattern of distress occurrence. This hypothesis is tested for situations in which the distresses have the same dimension, to exclude the effects of interaction between distress and section characteristics. Linear distresses on flexible, rigid, and composite pavements were used.

The following unconstrained model system was specified:

$$d_{ij1} = \alpha_{j1} + \beta_{j1}d^*_{i1} + \varepsilon_{ij1}$$

$$d_{ij2} = \alpha_{j2} + \beta_{j2}d^*_{i2} + \varepsilon_{ij2}$$

$$d_{ij3} = \alpha_{j3} + \beta_{j3}d^*_{i3} + \varepsilon_{ij3} \qquad (13)$$

where

$\alpha_{j1},\alpha_{j2},\alpha_{j3},\beta_{j1},\beta_{j2},\beta_{j3}$ = additive and multiplicative errors for inspection system j when measuring distresses on pavement types 1 (flexible), 2 (rigid), and 3 (composite), respectively; and

$d^*_{i1},d^*_{i2},d^*_{i3}$ = true value of distress on section i for pavement types 1, 2, and 3, respectively.

The hypothesis that there is no difference in inspection system biases when measuring distresses from backgrounds with different contrast and pattern of distress can be stated as follows:

$$H_0: \alpha_{j1} = \alpha_{j2} = \alpha_{j3}$$

$$\beta_{j1} = \beta_{j2} = \beta_{j3}$$

$$[\text{var}(\varepsilon_{ij1}) = \psi_1] = [\text{var}(\varepsilon_{ij2}) = \psi_2] = [\text{var}(\varepsilon_{ij3}) = \psi_3] \qquad (14)$$

To test the hypothesis in Equation 14, the unconstrained model in Equation 13 is estimated to get the values of the parameters β^{UC}_j and ψ^{UC}_j. The true values of distress are extracted using latent variable estimation techniques described in the unpublished data by Ben-Akiva and Humplick, which will be denoted by $(d^*_i)^{UC}$ for each pavement type. Then the observed values of distress on the three pavement types are stacked and a constrained model is estimated, as shown.

$$d^C_{ij} = \alpha^C_j + \beta^C_j(d^*_i)^C + \varepsilon^C_{ij} \qquad (15)$$

where the superscript C denotes the results of estimation from the stacked data.

The constrained model in Equation 15 represents the hypothesis

$$\alpha_1 = \alpha_2 = \alpha_3 = \alpha^C$$

$$\beta_1 = \beta_2 = \beta_3 = \beta^C$$

$$\psi_1 = \psi_2 = \psi_3 = \psi^C$$

Similarly, the true values of distress $(d^*_i)^C$ can be extracted from estimated values of β^C_j and ψ^C_j. Because the d^*_is for the constrained and unconstrained models are estimated from the respective β_js and ψ_js, one can compare the extracted d^*_is to make inferences about the β_js and ψ_js. This is a preferred procedure because it does not require computation of the error sum of squares, which is tedious to calculate and is required for any test on the equality of the β_js and ψ_js.

The following regression is performed for all three pavement types:

$$\underset{(n_z \times 1)}{d^*_{uc}} = \gamma \underset{(n_z \times 1)}{d^*_c} + \underset{(n_z \times 1)}{\varepsilon} \qquad (16)$$

where

z = 1, 2, 3 pavement types,
n_z = $n_1 + n_2 + n_3$, and
n_1, n_2, n_3 = number of flexible, rigid, and composite pavement segments, respectively.

The hypothesis that $\gamma = 1$ is then tested. If $\gamma \neq 1$, then there is a difference in inspection system bias when measuring distresses from backgrounds with varying contrast and pattern of distresses.

The results of the regression in Equation 16 are summarized in Figure 1. The hypothesis $\gamma = 1$ was accepted for the case of composite pavements and rejected for flexible and rigid pavements. This indicates that there is an effect of contrast and pattern of distress occurrence on the results of measurement.

Flexible Pavements

$$d_{UC} = \underset{(0.10)}{0.61}\, d_c \quad R-Square = 0.79$$

Rigid Pavements

$$d_{UC} = \underset{(0.11)}{0.91}\, d_c \quad R-Square = 0.89$$

Composite Pavements

$$d_{UC} = \underset{(0.57)}{1.18}\, d_c \quad R-Square = 0.46$$

Hypotheses

$$H_0: \gamma_f = \gamma_r = \gamma_c = 1$$

f = flexible, r = rigid, c = composite

The statistic used for testing is:

$$t_{N-2} = \frac{\bar{\gamma} - \gamma_0}{S_{\bar{\gamma}}}$$

Results

Flexible Pavements
T-statistic = -3.90
Reject null hypothesis at 95% confidence interval

Rigid Pavements
T-statistic = -0.82
Reject null hypothesis at 95% confidence level

Composite Pavements
T-statistic = 0.32
Accept null hypothesis at 95% confidence interval

FIGURE 1 Results of hypothesis tests on effects of section characteristics.

CONCLUSIONS

A methodology for identifying the factors affecting the results of measurement was developed and tested using highway inspection data. The success of such a methodology depends on scientifically collected data, such as were generated by the experimental design presented by Hudson et al. (*4*). The methodology can be used to identify directions for future development of inspection technologies, to choose among existing inspection systems, and to correct inspection results for measurement errors.

ACKNOWLEDGMENTS

This work was sponsored by the Army Research Organization through a project in the Center for Construction Research and Education at MIT. The author thanks Moshe Ben-Akiva for his advice and useful comments, and Ronald Hudson, Gary Elkins, Waheed Uddin, and Karen Reilley for providing the data.

REFERENCES

1. R. F. Carmichael III, D. S. Halbach, and L. O. Moser. *Pavement Condition Survey Manual for Rhode Island Pavement Evaluation and Management System*. Report IR-212. ARE Inc., 1985.
2. M. I. Darter, J. M. Becker, M. B. Snyder, and R. E. Smith. *NCHRP Report 277: Portland Cement Concrete Pavement Evaluation System, COPES*. TRB, National Research Council, Washington, D.C., 1985.
3. M. DeWilder. *GERPHO and the APL: Roadway Evaluation Equipment Workshop*. Report FHWA-TS-85-210. FHWA, U.S. Department of Transportation, 1985.
4. W. R. Hudson, G. E. Elkins, W. Uddin, and K. T. Reilley. *Improved Methods and Equipment to Conduct Pavement Distress Surveys*. Report FHWA-TS-87-213. FHWA, U.S. Department of Transportation, 1987.
5. D. Mahler. *Final Design of Automated Pavement Crack Measurement Instrumentation From a Survey Vehicle*. Final Report DTFH-61-86-C-001. FHWA, U.S. Department of Transportation, 1985.
6. K. R. Maser. *Computational Techniques for Automating Visual Inspection*. Working Paper. Department of Civil Engineering, Massachusetts Institute of Technology, Cambridge, 1987.
7. R. E. Smith, M. I. Darter, and S. M. Herrin. *Highway Pavement Distress Identification Manual*. FHWA, U.S. Department of Transportation, 1986.
8. M. R. Wigan. Some Investigations of Image Processing Techniques Applied to Road Problems. Presented at 1st International Conference on Applications of Advanced Technologies in Transportation Engineering, San Diego, Calif., 1985.
9. L. Finkelstein. Theory and Philosophy of Measurement. *Handbook of Measurement Science*, Vol. 1, John Wiley & Sons, Inc., New York, N.Y., 1982.
10. S. McNeil and F. Humplick. Evaluation of Errors in Pavement Surface Distress Data Acquisition Using Optical Techniques. Presented at 1st International Conference on the Application of Advanced Technologies to Transportation Engineering, San Diego, Calif., 1989.
11. K. Jeyapalan, J. K. Cable, and R. Welper. *Automated Pavement Data Collection Equipment*. Demonstration Project 72. Iowa Department of Transportation, Des Moines, 1987.
12. R. J. Larsen and M. L. Marx. *An Introduction to Mathematical Statistics and Its Applications*. Prentice Hall, Inc., Englewood Cliffs, N.J., 1981.
13. G. E. P. Box, W. G. Hunter, and J. S. Hunter. *Statistics for Experimenters: An Introduction to Design, Data Analysis, and Model Building*. John Wiley & Sons, Inc., New York, N.Y., 1978.

DISCUSSION

WAHEED UDDIN
Department of Civil Engineering, University of Mississippi, University, Miss. 38677.

The author has done an excellent job of presenting and applying a hypothesis-testing methodology to identify sources of error in the subjective manual, semiautomatic, and high-speed noncontact-type inspection systems for monitoring and evaluating pavement condition. The results discussed in this paper have important implications for selecting equipment and collecting data to evaluate pavement condition for infrastructure maintenance, preservation, and development. The following comments and discussion are related to the FHWA distress data base, on which the author relied to formulate the measurement error analysis problem and hypothesis testing.

The writer was one of the principal team members of the comprehensive FHWA study of pavement condition evaluation equipment (*1–3*) for which pavement test section selection and data collection were carried out with strict adherence to statistical experiment designs. The writer was primarily responsible for site selection and all field data collection.

This scientific study of equipment for and methods of monitoring pavement condition consisted of separate experiment designs for the following equipment categories: deflection, void detection, and distress survey.

Pavement nondestructive testing structural condition evaluation equipment included eight deflection devices (*1*):

- Slow-moving wheel load with manual data recording (dial gauge) and manual processing, requiring stops at the test locations (Benkelman beam).
- Continuously moving equipment with automated data recording by seismic geophone and automated data processing (Curviameter).
- Harmonic dynamic load equipment with automated deflection sensing by seismic geophone and automated processing, requiring stops at the tests locations (Dynaflect and Road Rater).
- Impact dynamic load falling weight deflectometer (FWD) equipment with automated data recording by seismometers or geophones and automated data processing, requiring stops at the test locations (three models of FWD and one replicate FWD unit).

The measurement of voids under concrete pavement for evaluation of structural integrity and assessment of concrete pavement restoration needs required very special equipment (*2*). The following devices were investigated for evaluating their capability of void detection and measurement of void size:

- Proof rolling and visual inspection,
- Deflection survey,
- Ground-penetrating radar equipment,
- Infrared thermography, and
- Transient dynamic response method.

Unfortunately, all of these methods required intensive manual data interpretation and special operator skills.

Seven varieties of equipment and methods were investigated for their suitability and reliability in distress survey and condition evaluation (*3*). Table 1 describes and groups the inspection system characteristics of these methods. Because the distress survey equipment is the subject of the paper, the measurement and processing principles of different distress data elements are summarized for the readers. The main differences among these methods are also highlighted.

- Mapping: Detailed direct manual measurements of all distress types including rutting by walking on selected inspection units within the pavement test section; procedure based on the AASHO Road Test distress mapping procedure; manual data processing. This is the method coded as Mapping.
- Manual visual surveys (PAVER/COPES): Detailed direct manual severity rating and extent measurement by walking using specific sampling and visual inspection guidelines for all distress types including rutting on selected inspection units within the pavement sections; manual data processing. This is the method coded as Manual.
- Semiautomated data logger: Measurements similar to detailed manual visual surveys by walking survey and entering data directly on a hand-held data logger (portable PC); automated data processing. This is the method coded as Logging.
- GERPHO: High-speed automatic imaging of pavement surface on continuous 35-mm photo film, at night only; manual distress data interpretation on full section length; no rutting data; automatic data processing. This is the method coded as Photo1.
- PASCO-Roadrecon survey vehicle: Multifunction high-speed automatic data collection; automatic data processing. Imaging of pavement surface on continuous 35-mm photo films at night only; manual distress data interpretation on full section length; automatic rutting data processing from digitized photo records of transverse profiles; longitudinal profile measurement by laser sensors. This is the method coded as Photo2.
- ARAN video condition inventory survey vehicle: Multifunction high-speed automatic data collection; automatic data processing; video imaging of perspective view and pavement surface; no interpretation of distress data from video. Windshield visual manual distress data collection, using integrated data logger on full section length; automatic rutting data processing from transverse profiles measured by ultrasonic sensors; longitudinal roughness measurement by accelerometer. This is the method coded as Video.
- Laser RST survey vehicle: Multifunction high-speed automatic data collection; automatic data processing; laser survey of pavement surface for measuring longitudinal and transverse profiles and texture data processing (only some transverse cracking data were produced from laser survey and no other distress data were interpreted from laser survey). Windshield visual manual survey for alligator, longitudinal, and edge cracking data and other distress data collection, using integrated data logger on full section length; automatic rutting data processing from transverse profiles measured by laser sensors. This is the method coded as Laser.

It is obvious from these comments that the alligator cracking, longitudinal cracking, and edge cracking data from both Video and Laser devices are essentially collected in the windshield-type visual survey mode using on-board integrated data loggers. These and other distress data, excluding rutting data, are visual, manual, subjective measurements reported by these high-speed multifunction devices. Therefore, these data are not expected to be of the same quality as the distress data processed from the Photo1 and Photo2 equipment, Logging, Manual, and Mapping methods.

The author is encouraged to examine rutting data for hypothesis testing. As described, rutting data were collected by objective measurements by Photo2, Video, and Laser. Direct manual objective measurement was used in Mapping, and subjective data collection procedures were used for rutting survey in Manual and Logging methods.

Pavement management system (PMS) development and implementation is a top priority area on federal-aid highway systems throughout the United States. Pavement condition data monitoring and evaluation, particularly distress data and rutting data, are integral components of the PMS process. Multifunction and high-speed equipment providing objective measurements of pavement condition data are attractive and cost-effective alternatives during the PMS equipment selection process (*4*). However, speed and productivity should not be the only selection criteria; quality and accuracy of pavement condition evaluation and prediction are important as well. Note that the maintenance need assessment and the maintenance work program and budgets depend on the quality of pavement condition data. The author is commended for bringing the subject of data quality and sources of error in distress survey methods and equipment to the attention of pavement community.

REFERENCES

1. W. R. Hudson, G. E. Elkins, W. Uddin, and K. T. Reilley, *Evaluation of Pavement Deflection Equipment*. Final Report FHWA-TS-87-208. ARE, Inc.; FHWA, U.S. Department of Transportation, March 1987.
2. W. Uddin, W. R. Hudson, and G. E. Elkins. *Evaluation of Equipment for Measuring Voids Under Pavements*. Final Report FHWA-TS-87-229. ARE, Inc.; FHWA, U.S. Department of Transportation, Sept. 1987.
3. W. R. Hudson, G. E. Elkins, W. Uddin, and K. T. Reilley, *Improved Methods and Equipment to Conduct Pavement Distress Surveys*. Final Report FHWA-TS-87-213. ARE, Inc.; FHWA, U.S. Department of Transportation, April 1987.
4. W. R. Hudson and W. Uddin. Future Pavement Evaluation Technologies: Prospects and Opportunities. *Proc., 2nd North American Conference on Managing Pavements*, Vol. 3, Toronto, Ontario, Canada, 1987.

AUTHOR'S CLOSURE

The discussant has provided a valuable description of the inspection technologies that is useful for interpreting the significance of the results of this paper. As mentioned in the conclusion, scientifically collected data, gathered by the methods used in the background papers presented by the discussant, are necessary input to the success of the methodology developed in this paper.

As suggested by the discussant, bias parameters were estimated for rutting data. The results are presented in Table 7. As can be seen from these results, all systems perform very well with respect to the standard error of measurement (which is practically zero for all systems). However, the inspection system with the lowest additive bias is Photo2, which underestimates rutting by only 0.01 in.

The impact of lack of objectivity is seen by an underestimation by Manual of 0.11 in., and the impact of system limitations (resolution) is exemplified by a 0.32 in. overestimation by Video. The systems with the least multiplicative bias, however, are Mapping, which underestimates rutting by a factor of only 0.02, and Logging, which overestimates rutting by a factor of 0.02. The seriousness of over- or underestimation depends on the use to which the data are put. A methodology for choosing among inspection technologies on the basis of their accuracy of measurement and whether the data are used to predict performance or make maintenance decisions can be found elsewhere (*1*).

The Photo2 technology has the highest multiplicative bias for rutting measurements, overestimating them by a factor of 1.23. However, this is not a problem, because the inspection results can be corrected for using the results of the calibration in Table 7. The results in Table 7 indicate that the benefits of direct measurement are undermined by the spatial variation of measurements that cannot be captured by sampling strategy employed by Mapping. However, these effects are additive in nature and hence can be factored out using the results of calibration. On the other hand, the advantages of automation (such as when using Photo2) can be achieved only if the results of measurement are corrected for error of inspection. The advantage of the rutting data is that all the technologies at the moment have insignificant random errors and, because one can correct for systematic errors, there should be no advantage other than cost and speed of data collection.

The author suggests the use of spatial models to estimate the impact of spatial effects on measurement errors. Such work is ongoing; preliminary results have been published by Koutsopoulos and Mishalani (*2*).

TABLE 7 ESTIMATED BIASES FOR DIFFERENT INSPECTION SYSTEMS (RUTTING ON FLEXIBLE PAVEMENTS)

	Estimated Parameters (standard errors of the estimates)			
Inspection system (j)	$\bar{\alpha}_j$ (Sqft)	$\bar{\beta}_j$	$S.D.(\epsilon_{ij}) = \sqrt{\psi_j}$ (Sqft)	Coeff. Of det. R^2
1. Mapping	-0.20	0.85 (0.36)	0.00	0.82
2. Manual	-0.11	0.81 (0.30)	0.00	0.86
3. Logging	0.07	1.14 (0.18)	0.00	0.85
4. Photo1	--	--	--	--
5. Photo2	-0.01	2.11 (0.21)	0.00	0.88
6. Video	0.32	0.31 (0.09)	0.00	0.76
7. Laser	-0.06	0.78 (0.27)	0.00	0.86

-- The Photo1 inspection system did not report measures for rutting as denoted by -- in the table. Additionally, the standard errors of the additive biases were not calculated as this is a very time consuming activity.

Since an unbiased system has a multiplicative bias of one, the degree of over or underestimation is calculated as $(\beta_j - 1)^2$, which is 0.02 for Mapping, 0.04 for Manual, 0.02 for Logging, -- for Photo1, 1.23 for Photo2, 0.48 for Video, and 0.05 for Laser respectively.

REFERENCES

1. M. Ben-Akiva, F. Humplick, S. Madanat, and R. Ramaswamy. Infrastructure Management Under Uncertainty: The Latent Performance Approach. *Journal of Transportation Engineering*, ASCE, 1992 (in press).
2. H. N. Koutsopoulos and R. G. Mishalani. A Method for Automated Analysis and Spatial Representation of Pavement Distress Data. Presented at 71st Annual Meeting of the Transportation Research Board, Washington, D.C., Jan. 1992.

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