Infrastructure Management under Uncertainty: Latent Performance Approach

By Moshe Ben-Akiva,[1] Frannie Humplick,[2] Samer Madanat,[3] and Rohit Ramaswamy[4]

Abstract: This paper presents a new framework for the analysis of infrastructure performance and the planning of inspection, maintenance, and rehabilitation (M&R) activities. The main facets of this framework are: explicitly modeling the errors in measurement of facility performance indicators; treating the key variable in the process, facility performance, as a latent variable that manifests itself through measured condition indicators; and explicitly analyzing and including in the decision process the uncertainties in infrastructure condition measurement and performance forecasting. Our approach recognizes the errors inherent in the measurement of the condition indicators; these errors are quantified through the use of measurement error models. A model system relating the latent facility performance to explanatory variables and to observed indicators is developed. Our M&R activity planning model accounts for the presence of uncertainty in the output of the inspection of facility condition and the forecasting of facility performance. An application example is presented.

Introduction

Infrastructure management is the process by which agencies monitor and maintain built systems of facilities, with the objective of providing the best possible service to the users, within the constraints of available resources. More specifically, the management process refers to the set of decisions made by an infrastructure agency concerning the allocation of funds among a system of facilities and over time. The decisions that are made by infrastructure agencies are maintenance and rehabilitation (M&R) decisions, the focus of which is the type of M&R activity to perform on each facility in each year of a specified planning horizon.

These decisions require several data items, such as the available budget, the cost and effectiveness of different activities, and the current and projected levels of usage; however, the most important data item is the performance (or condition) of each facility in the system.

There are two forms of information on infrastructure performance: information on current (measured) performance, which is obtained through facility inspection; and information on future (forecasted) performance, which is obtained using performance forecasting models. Both forms of performance information are characterized by substantial uncertainty, which has important implications on the management decisions and the life-cycle costs of infrastructure facilities. The treatment of uncertainty is the major motivation for the research reported in this paper.

The relationships between the measurement and forecasting of facility

---

[1]Prof., Dept. of Civ. Engrg., Massachusetts Inst. of Tech., Cambridge, MA 02139.
[2]Infrastructure Systems Analyst, Infrastructure and Urban Development Dept., The World Bank, Washington, D.C.
[3]Asst. Prof., School of Civ. Engrg., Purdue Univ., West Lafayette, IN 47907.
[4]Member of Tech. Staff, AT&T Bell Labs., Holmdel, NJ 07733.

Note. Discussion open until June 1, 1993. To extend the closing date one month, a written request must be filed with the ASCE Manager of Journals. The manuscript for this paper was submitted for review and possible publication on October 23, 1990. This paper is part of the *Journal of Transportation Engineering*, Vol. 119, No. 1, January/February, 1993. ©ASCE, ISSN 0733-947X/93/0001-0043/$1.00 + $.15 per page. Paper No. 761.

condition and M&R decision making are shown in Fig. 1. The facility condition data collected using different inspection technologies is used in two ways. First, it is one of the items used in the estimation of infrastructure performance models. Second, as mentioned previously, it is one of the major inputs in the selection of maintenance and rehabilitation (M&R) strategies in a given time period. Infrastructure performance models are used in planning present and future M&R activities. In the decision-making block, M&R strategies are selected for the present and future time periods.

In this paper, we present our development of a methodological framework to support the infrastructure management process. Fig. 2 presents this methodological framework. In this diagram, rectangles represent observed quantities, ellipses represent latent variables, and diamonds represent decisions. At the upper level of the diagram, the true values of condition (or performance) indicators of an infrastructure facility are estimated from the indicator measurements obtained through different measurement technologies. The relationship between the two is explained by a "measurement error model," which is a function of technological, environmental, facility, and damage-specific factors. The true values of condition indicators are an input to the second level, where the performance of a facility is estimated through a "performance model." The other inputs to the model system in Fig. 2 are exogenous variables such as usage, environment, past maintenance and rehabilitation activities, facility characteristics, and possible performance ratings. This model system provides unbiased estimates of the present

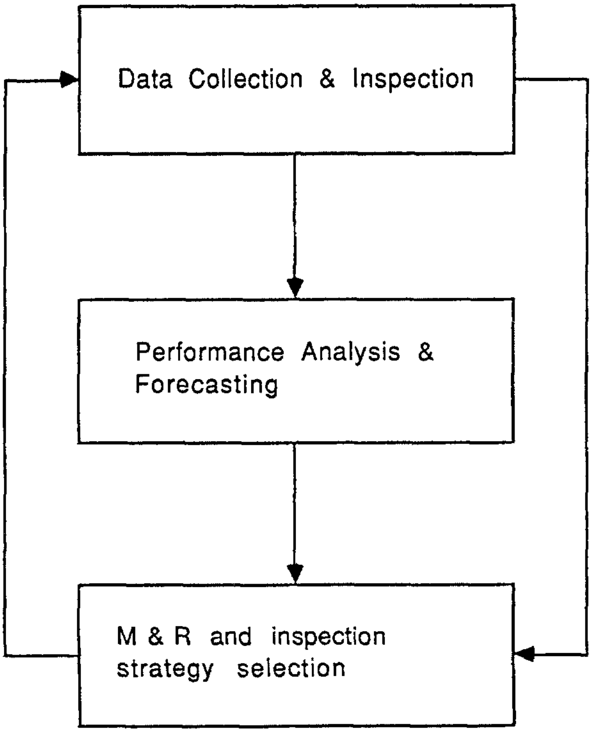

FIG. 1. Infrastructure Management Process

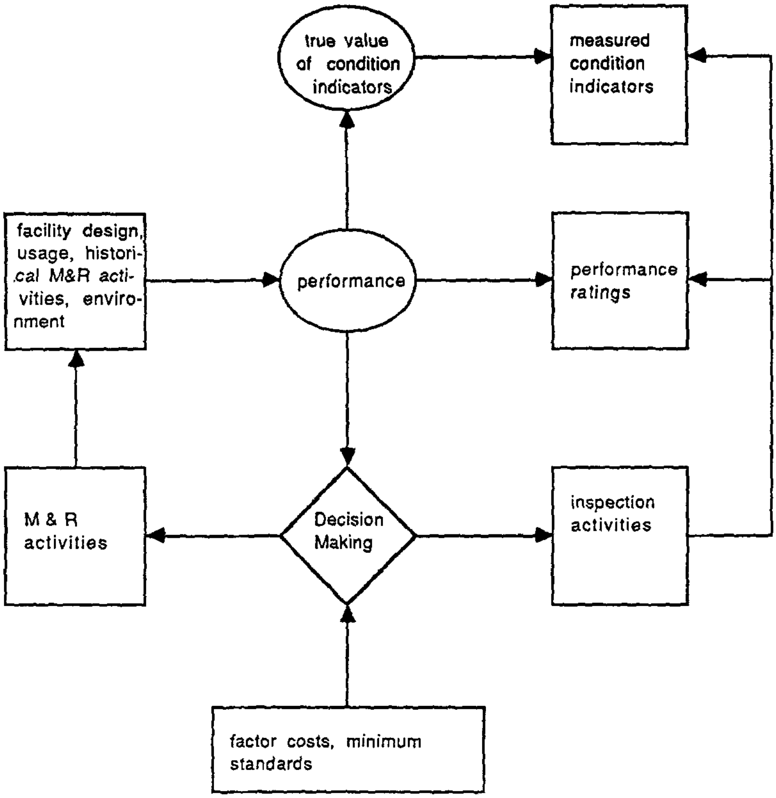

FIG. 2. Latent Performance Framework

performance of an infrastructure facility, and can be used to forecast future performance. Both present and forecasted performance are inputs to the third level, where maintenance and rehabilitation decisions are jointly made. Other inputs to the decision-making algorithm are the measurement error models estimated in the first level, the performance models estimated in the second level, the costs of different activities, and minimum performance standards. The outputs of the algorithm are optimal maintenance and rehabilitation policies that feed back into the higher levels of the model.

In the following sections, our methodological approach is presented in detail. The next section overviews the decision-making methodology that selects maintenance and rehabilitation activities. Two sections following present, respectively, the two major inputs to this methodology: the measurement error models (which describe the uncertainty in measured performance indicators); and the facility performance models (which combine these indicators into a performance index and describe the uncertainty in forecasted performance). A case study using this methodological approach is presented next, followed by conclusions and suggestions on further research. The specific type of infrastructure facilities we emphasize in the following sections are highway pavements, due to the availability of data. The methodologies presented are, however, applicable to any type of infrastructure.

METHODOLOGY FOR DECISION MAKING

Several research efforts to bring about systematic methodologies to address M&R decisions (using operations research techniques) have been made in the last decade. However, most of these have ignored the inherent uncertainty in measuring and forecasting facility performance, which have made them of little use when applied in the field. Exceptions to this trend (Carnahan et al. 1987; Feighan et al. 1988) adopted a stochastic optimization approach, based on the Markov decision process (MDP), to account for the uncertainty in predicted performance.

In an MDP, the facility condition is represented by a set of discrete states, $i = 1, \ldots, n$, and the condition prediction is represented by a set of Markovian transition probabilities, denoted by:

$$p(x_t = j | x_{t-1} = i, a_{t-1}); \qquad i, j = 1, \ldots, n \quad \ldots\ldots\ldots\ldots\ldots\ldots (1)$$

where x_t = facility condition state at the beginning of time period t; a_{t-1} = M&R activity performed during $t - 1$; i, j = indices of elements in the set of discrete condition states; and p = a known probability mass function.

This approach, however, assumes that there is no uncertainty in the measured facility performance. Since infrastructure management requires a combination of inspection and prediction of facility performance, and since the uncertainty in inspection output is not negligible (as will be demonstrated later), there is a need to develop a methodology that takes into account both sources of uncertainty.

Our methodology recognizes the uncertainty in both facility performance prediction and measurement, through the use of the latent Markov decision process (LMDP). This methodology was developed in the field of manufacturing (Smallwood and Sondik 1973). The LMDP is an extension of the traditional Markov decision process (MDP) methodology, but differs from it in one major aspect: It does not assume that the measurement of facility performance is necessarily error free. Instead, it recognizes that the decision maker observes outputs from measuring the performance of a facility, which are probabilistically related to the true performance.

Mathematically, this relation is expressed by:

$$q(\hat{x}_t = k | x_t = j); \qquad t = 0, 1, \ldots, T \quad \ldots\ldots\ldots\ldots\ldots\ldots (2)$$

where $\hat{x}_t$ = measured condition state of the facility at start of t; x_t = true condition state of the facility at start of t; j, k = indices of elements in the set of discrete condition states; and q = a known probability mass function.

To select M&R policies in the presence of uncertainty, in a meaningful manner, the decision maker cannot base decisions solely on the measured condition of the facility. Instead, all the information available to the decision maker about the facility can be relevant to future decisions. This information (the history of measured condition states and M&R activities) is referred to as "the state of the information." Denoting this state by I_t, we have:

$$I_t = \{I_0, a_0, \hat{x}_1, a_1, \ldots, \hat{x}_{t-1}, a_{t-1}, \hat{x}_t\}; \qquad t = 1, 2, \ldots, T$$

$$I_0 = \{\hat{x}_{-t'}, a_{-t'}, \ldots, \hat{x}_{-1}, a_{-1}, \hat{x}_0\} \quad \ldots\ldots\ldots\ldots\ldots\ldots (3)$$

where t' = number of years between first inspection of facility and the start of the planning horizon.

It can be shown that if the deterioration of the facility is represented by a finite state Markov chain, and the measurement uncertainty is represented by a set of discrete probabilities, then the state of the information itself will evolve in a Markovian fashion (Madanat 1992). This evolution is expressed mathematically by

$$P(I_{t+1} = K|I_t, a_t) \quad \text{(4)}$$

where K = a possible state of the information; and P = a known probability mass function. The details of the mathematical derivation of P are given in (Madanat 1992).

In this manner, the M&R decision problem is transformed from one where the state (condition of the facility) is latent to one where the state (information about the facility) is observed. Once this transformation has been performed, the same solution methods used to find optimal M&R policies for a Markov decision process (such as dynamic programming in the case of nonrandomized policies) can be adapted to the latent Markov decision process.

The similarity between the LMDP and the classical MDP can best be explained in terms of the underlying decision trees. Fig. 3 depicts a classical MDP tree. At the beginning of time period t, the true state x_t is observed.

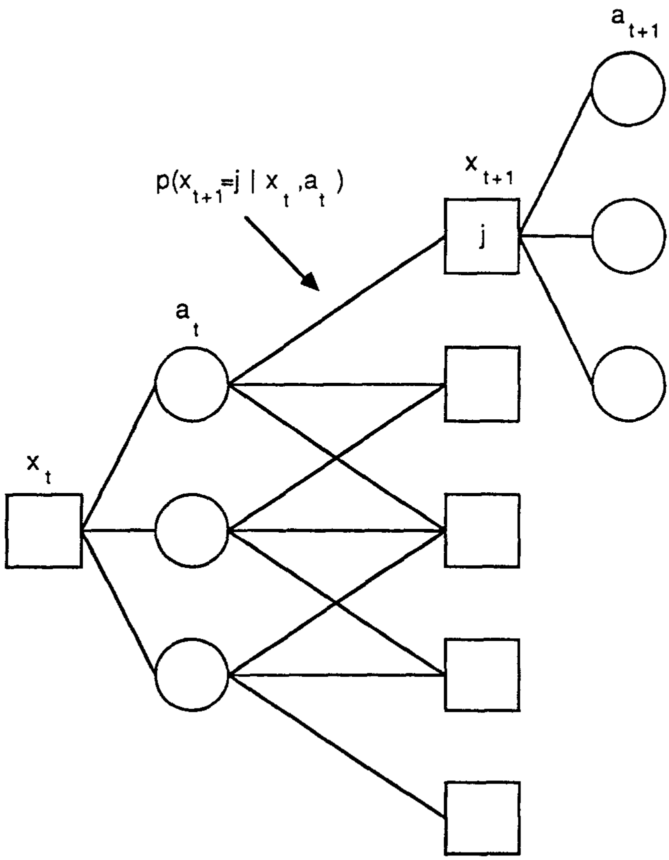

FIG. 3. Decision Tree for Markov Decision Process

Based on this knowledge, the decision maker selects an activity a_t. Given the condition state x_t and the selected activity a_t, the facility moves to one of the states $x_{t+1} = j$ with probability $p(x_{t+1} = j|x_t, a_t)$. The same process is then repeated in time period $t + 1$, and so on.

In Fig. 4, part of a LMDP tree is shown. We start in period t, when the decision maker has available the state of the information I_t. Based on this information, an activity a_t is selected. Given the information state I_t and a_t, the system moves to one of the states $I_{t+1} = K$, with probability $P(I_{t+1} = K|I_t, a_t)$. The same process is then repeated in time period $t + 1$, and so on.

Before the capabilities of this decision-making model are demonstrated, the derivation of the two major inputs must be described. These inputs are:

- The measurement probabilities of (2),
- The transition probabilities of (1).

These two inputs are derived using the measurement error models and the facility performance models developed in the next two sections, respectively.

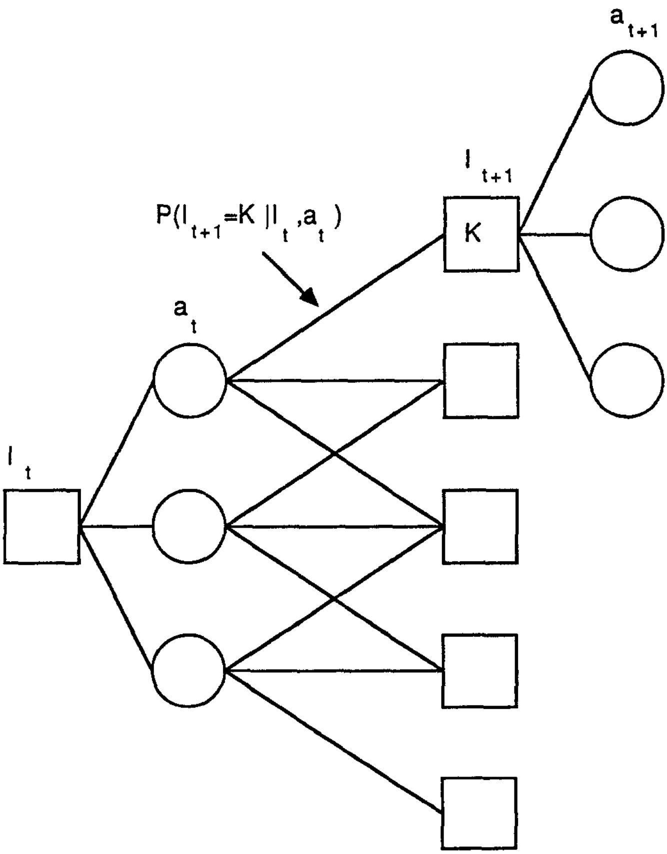

FIG. 4. Decision Tree for LMDP with Annual Inspections

Measurement Error Models

This section develops a model for establishing the relationship between the measured condition indicator and the true condition, which are inputs to establishing the measurement probabilities in (2). The model is referred to as a measurement error model, which is a mathematical expression explaining the difference between the true value of an indicator at a given location and its measured value. This is done in terms of systematic biases and a random error. A linear specification of such a model is:

$$\hat{d}_{lmk} = \alpha_{mk} + \beta_{mk} d_{lk} + \varepsilon_{lmk} \qquad (5)$$

where $\hat{d}_{lmk}$ = measured value of a condition indicator of type k, on section l by measurement process m; d_{lk} = true value of condition indicator k on section l; α_{mk}, β_{mk} = additive and multiplicative systematic biases of measurement process m when measuring indicator type k; and ε_{lmk} = random error of measurement for indicator type k, on section l by measurement process m, with zero mean, and a standard deviation denoted by Ψ_m.

For example, for a given highway pavement segment, the measured indicators $\hat{d}_{lmk}$ can be the degree of cracking, rutting, potholes, or roughness. The measurement processes m, which are typically used to measure such indicators, include detailed mapping, manual inspection, photographic imaging, video imaging, and laser measurement.

A combination of factors can lead to systematic measurement biases and random measurement errors. These factors include:

1. Technological factors (e.g., resolution and field of view).
2. Distress factors (e.g., dimensions of distress).
3. Section or location factors (e.g., pavement type).

Separate measurement equations can be specified for each indicator k, which reduces (5) to $k = 1, \ldots, K$ separate measurement equations of the form:

$$\hat{d}_{lm} = \alpha_m + \beta_m d_l + \varepsilon_{lm} \qquad (6)$$

where the parameters α_m, β_m, and Ψ_m characterize the relationship between the true value and measured value for each measurement process m. An unbiased measurement process is one that has $\alpha_m = 0$, $\beta_m = 1$, and a precise measurement process is one that has small Ψ_m.

Once the systematic biases α_m, β_m have been statistically estimated using a sample of measurements by different processes, they can be corrected before using the measured results for decision making. However, the standard deviation of the random error Ψ_m, which measures the uncertainty of the measurement, needs to be accounted for in the decision-making processes as outlined previously. The estimation procedure to obtain the systematic and random errors of measurement from the model specification in (6) are discussed next.

There are two major estimation approaches for obtaining the parameters in (6). These are regression against an average and factor analysis normalized to an average. The regression approach is currently used by a variety of agencies [see Hudson et al. (1987)]. This approach assumes that there is an unbiased measurement process, whose results can be used to calibrate other measurements, and obtain their biases and random errors. This approach results in biased estimates of the β_m parameters in (6), as shown in Humplick (1992).

In our research, the factor analytic approach was used to obtain unbiased estimates of the parameters in (6). This approach assumes that the true value is latent. The measurement biases and random error of measurement are obtained using the covariances between multiple measurements by different technologies. These covariances are used to obtain estimates of the random error of the measurement Ψ_m directly, and estimates of the multiplicative bias β_m through normalization against a reference measurement. In this manner, the estimation of the random error of measurement by the factor analytic approach is independent of the reference measurement. In addition, the factor analysis approach uses more information to estimate the biases and random error and, hence, results in a better fit to the data.

A reference measurement can be developed by averaging measurements from processes with radically different measurement principles (such as a video imaging technology and a laser nondestructive measurement technology). The major assumption is that the measurement errors from radically different measurement principles will cancel out, resulting in an unbiased average with low variance. Such reference measurements are developed and tested in Humplick (1992), and the estimation procedures using these averages are also reported.

The factor analytic approach was applied to a data set from a Federal Highway Administration (FHWA) (Hudson et al. 1987). The study included measurement of surface distresses on highway pavements using a variety of existing and newly developed technologies and methods. The technologies included three direct measurement techniques (mapping, manual, logging) involving visual inspection by humans; and four indirect measurement techniques involving optical imaging (photo1, photo2, video) and a laser nondestructive measurement technique.

An example of estimated models for measurements of the area of alligator cracking on pavement sections in the study is shown in Table 1. This table compares the estimated results obtained by the factor analytic approach to those by the traditional regression approach. From Table 1, we can see that the estimated parameters β_m by the regression approach are generally smaller than those by the factor analytic approach. This is an indication of the underestimation of the biases and random error of measurement as a result of calibrating against a reference measurement. In addition, the goodness of fit (R^2) obtained from the factor analytic specification is better than that obtained using traditional regression techniques.

The estimated additive and multiplicative biases α_m, β_m in Table 1, from the factor analytic specification, can be used to correct the measured values of distress for future use. The uncertainty of measurement Ψ_m cannot be corrected but can be included in the decision-making processes to select future inspection frequencies or future M&R actions. The manner in which these random errors are included in the decision-making model is discussed herein.

The estimated parameters in Table 1 can also be used to select among available technologies, by comparing accuracies and precisions of a variety of technologies. For example, one would select photo1 over the other measurement processes based on lower additive and multiplicative bias. However, assuming one could correct for systematic biases given that they are known, one would select manual over the other measurement processes on the basis of lower random error of measurement.

TABLE 1. Comparison of Alternate Estimation Approaches Case of Alligator Cracking on Flexible Pavements

Technology (m) (1)	ESTIMATED PARAMETERS (STANDARD ERRORS IN PARENTHESES)							
	α_m (m²)		β_m,		Standard Deviation $(\varepsilon_{lm}) = \sqrt{\Psi_m}$ (m²)		Coefficient of Determination (R^2)	
	Regression (2)	Factor analysis (3)	Regression (4)	Factor analysis (5)	Regression (6)	Factor analysis (7)	Regression (8)	Factor analysis (9)
Mapping	−4.35 (17.6)	−6.78 (44.05)	0.81 (9.08)	0.83 (0.17)	42.6	36.9	0.92	0.94
Manual	4.90 (10.72)	3.48 (33.79)	0.48 (0.05)	0.49 (0.10)	66.3	24.4	0.91	0.94
Logging	40.25 (59.28)	52.95 (78.5)	1.41 (0.29)	1.29 (0.54)	143.6	60.0	0.75	0.61
Photo1	−11.44 (21.78)	−14.35 (49.0)	1.06 (0.11)	1.09 (0.21)	52.6	41.3	0.93	0.95
Photo2	−43.67 (23.28)	−10.8 (51.2)	1.82 (0.11)	1.85 (0.23)	56.5	51.2	0.97	0.99
Video	14.31 (18.16)	12.4 (44.1)	0.42 (0.09)	0.44 (0.17)	75.5	43.9	0.74	0.77
Laser	—	—	—	—	—	—	—	—

In our research, measurement error models were estimated for seven different technologies, five types of condition indicators, and two pavement types. A more complete description of our work is given in Humplick (1989, 1992). The next section discusses how the measured indicators discussed in this section are used to develop infrastructure performance models, which are the second input to the decision-making process.

Facility Performance Model

The scope of this section is the estimation of an infrastructure deterioration model. A deterioration model relates the performance of an infrastructure facility to a set of causal variables, such as traffic, age, maintenance history, etc. Such a model is required as an input to selecting maintenance and rehabilitation activities for infrastructure facilities.

The main problem in estimating such a model is that "performance" is not directly observable. What can be observed are indicators of performance, such as roughness, cracking, rutting, and skid resistance, as was discussed in the previous section of this paper.

Numerous studies have attempted to devise "performance indexes," which combine different indicators into a single quantity. This includes the pavement condition index (PCI), developed by Shahin and Kohn (1981), and the present serviceability index (PSI), developed in AASHO ("The AASHO" 1962). These indexes were based on the subjective judgment of pavement experts, lack rigorous justification, have poor explanatory power, and use a predetermined set of indicators, which precludes incorporation of new condition indicators. Furthermore, when these indexes were used together with common causal variables, to estimate a deterioration model, the resulting fit to data was poor.

Our approach, on the other hand, does not rely on subjective judgment for devising a "performance index." Indeed, it does not require the predetermination of such an index. Instead, it treats performance as a latent variable. This variable (S) is linked to explanatory (or causal) variables ($\mathbf{X}$) and maintenance actions ($\mathbf{A}$) through a deterioration model. Moreover, it is linked to a set of condition indicators ($\mathbf{D}$) through a measurement model. These models form a system of equations estimated simultaneously, thereby producing a much better fit to data than traditional deterioration models. Fig. 5 shows a schematic representation of the process. In this figure, rectangles represent observed quantities, while the ellipse represents the unobserved latent performance.

This approach was applied to a data set consisting of 3,837 1,609-km (1-mi) pavement sections from Nevada. This set contained information on several condition indicators (cracking, rut depth, roughness) for each section, as well as a set of causal variables (average daily traffic, percentage of truck traffic, age, maintenance by several activities, and several environmental variables).

In the process of estimating our model system, we uncovered a major problem that had not been resolved in existing deterioration models, which are estimated with data from in-service pavements. Such pavements are subjected to maintenance performed by highway agencies in response to their level of traffic, percentage of trucks, etc. As a result, pavements with the highest level of usage will receive higher levels of maintenance. In other words, two conflicting mechanisms are acting on these pavements:

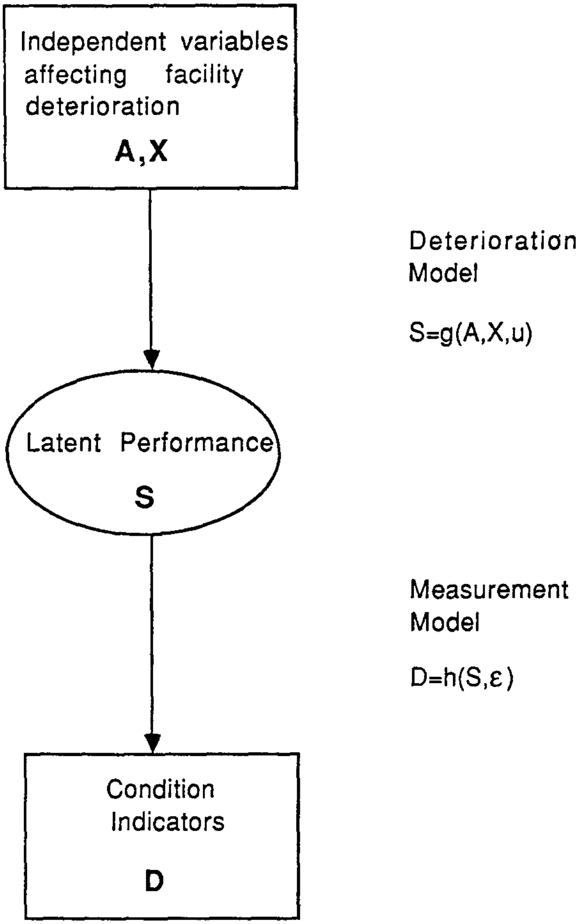

FIG. 5. Integrated Model of Latent Facility Performance and Condition Measurement

1. A deterioration mechanism, where condition decreases as traffic and age increase.
2. A maintenance mechanism, where a condition increases as traffic increases.

An attempt to estimate a deterioration model from in-service pavements without taking the second mechanism into account, as is usually the case with state-of-the-art deterioration models, will result in biased, counterintuitive parameter estimates. The correct specification for such a situation is a simultaneous equation model, including two separate relationships—one for each of the two mechanisms described.

Several examples of an estimated deterioration equation of a latent variable model system are given in Ramaswamy (1989), Ramaswamy and Ben-Akiva (1990), and Ben-Akiva and Ramaswamy (1989). The results show that all the parameter estimates have intuitively correct signs. This is a result of the simultaneous equation specification described. The value of R^2, which is a measure of fit to data, is also reported and is higher than is usually the case in existing deterioration models. This is a result of using the latent variable approach. A complete description of our work in this area is given in the aforementioned references.

Application of Methodology

In this section, a case study in the area of highway pavements is presented to illustrate the application of the methodology of this paper.

The first step in applying the derivations of the previous sections, is to select an appropriate discretization of the performance scale. We relied on previous work by other authors (Carnahan et al. 1987, Feighan et al. 1988) to select a discretization scheme. As a result, it was decided to select eight performance states, that is $i = 0, \ldots, 7$.

The next step consists of deriving the measurement probabilities and the costs of inspection associated with different technologies. The measurement probabilities were derived from the measurement error models estimated by Humplick (1989), which were illustrated previously. These measurement error models relate the measured values of the indicators to their true values, whereas the measurement probabilities relate the measured performance to the true performance. Hence, in order to use the measurement error models, the relation between performance and the performance indicators is required. This relation is provided by the performance index equations estimated by Ramaswamy (1989). This equation, discussed earlier, relates the latent performance index to roughness, cracking, rut depth, potholes, and patching. The resulting relation between the true performance S and the measured performance with any technology m, $\hat{S}_m$, is given in Madanat (1991).

Using that relation, we can compute the discrete measurement probabilities. These probabilities have the form:

$$q(\hat{x} = k|x = j, m); \qquad 0 \leq j, k \leq 7 \quad \text{(7)}$$

where $\hat{x}$ = measured performance state; x = true performance state; and m = measurement technology.

The variable costs of inspection, associated with different technologies were given in the original report from which the data, used in the estimation of the measurement error models, was taken (Hudson et al. 1987).

The measurement probabilities and the costs of inspection associated with the measurement technology used in this case study are given in Table 2.

The transition probabilities, in this case study, were derived from a deterioration model taken from Ramaswamy (1989), where the dependent variable was the latent performance index. This model was of the form shown in an earlier section. Due to the presence of age (time since last reconstruction) as an explanatory variable in the deterioration model, the resulting transition probabilities were age dependent, that is, nonhomogeneous. The latent performance deterioration model included only one M&R activity (routine maintenance) among the explanatory variables. The transition probabilities for that activity, i.e., for routine maintenance, were solved for using age-dependent condition state probabilities (the probabilities of the condition being in different states as a function of age), calculated from the forecasts obtained with the deterioration model, and assuming that the facility was in the highest possible condition at age zero. The solution

TABLE 2. Measurement Probabilities and Inspection Costs

True state	Measured State							
	7	6	5	4	3	2	1	0
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
7	0.70	0.20	0.10	0.00	0.00	0.00	0.00	0.00
6	0.30	0.40	0.20	0.10	0.00	0.00	0.00	0.00
5	0.10	0.20	0.40	0.20	0.10	0.00	0.00	0.00
4	0.00	0.10	0.20	0.40	0.20	0.10	0.00	0.00
3	0.00	0.00	0.10	0.20	0.40	0.20	0.10	0.00
2	0.00	0.00	0.00	0.10	0.20	0.40	0.20	0.10
1	0.00	0.00	0.00	0.00	0.10	0.20	0.40	0.30
0	0.00	0.00	0.00	0.00	0.00	0.10	0.20	0.70

Note: Inspection costs = $0.012/m².

uses basic relations between state probabilities and transition probabilities, from Markov chain theory (Madanat 1991).

The costs per unit area associated with routine maintenance were included in the study by Ramaswamy and were taken directly from that study.

A second activity (reconstruction) was added for the purposes of this case study. It was assumed that reconstruction brings the performance of the facility back to state $i = 7$, which is the best possible state, and brings the age back to zero. This assumption has been made elsewhere (Carnahan et al. 1987). The costs per unit area associated with reconstruction were taken from the study by Carnahan et al. (1987), where they were assumed to be independent of the performance state before reconstruction. This assumption allowed us to use the same unit costs without concern for the fact that the performance index used by Carnahan et al. (the PCI) was different than ours (the latent performance index). We also needed to assume that the highest performance states on the PCI scale and the latent performance measure scale coincide, which is a reasonable assumption.

It was assumed that, at the beginning of the planning horizon, we started with a new facility, that is: age (τ) = 0, and $i = 7$. Note that the following notation was used: τ for age and t for time. The time-dependent transition probabilities and the costs associated with the two M&R activities (routine maintenance and reconstruction) are given in Tables 3 and 4. In Table 3, only rows corresponding to nonzero state probabilities are shown. For example, since at age 0, the facility can only be in state 7, only the row corresponding to state 7 is shown for $\tau = 0$.

The final item required for the decision-making algorithm was the minimum performance standard, to guarantee that the user costs did not exceed a certain threshold. It is reasonable to assume that a PSI (present serviceability index) of 2.0 (on a scale of 0–5.0), should be used as the minimum performance standard. In order to establish a threshold in terms of the latent performance measure used in this study, it was necessary to be able to translate this minimum standard into a latent performance standard. The PSI standard was easily "translatable," because it was based on the same set of performance indicators as the latent performance measure. It was therefore possible to directly correlate the two measures of performance and to obtain a minimum standard in terms of the latent performance. The corresponding minimum acceptable condition state for the facility was $x_{min} = 3$.

TABLE 3. Transition Probabilities and Costs of Routine Maintenance (for Pavement Ages 0, 1, 2 Years)

$x(\tau)$ (1)	$x(\tau+1)$ 7.00 (2)	6.00 (3)	5.00 (4)	4.00 (5)	3.00 (6)	2.00 (7)	1.00 (8)	0.00 (9)
				(*a*) $\tau = 0$				
7.00	0.92	0.08	0.00	0.00	0.00	0.00	0.00	0.00
				(*b*) $\tau = 1$				
7.00	0.88	0.12	0.00	0.00	0.00	0.00	0.00	0.00
6.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
				(*c*) $\tau = 2$				
7.00	0.75	0.25	0.00	0.00	0.00	0.00	0.00	0.00
6.00	0.00	0.94	0.06	0.00	0.00	0.00	0.00	0.00
				(*d*) Costs of routine maintenance ($/m²)				
$ca(x(\tau))$	0.05	0.18	0.37	0.78	0.99	1.67	2.39	8.25

Note: τ = pavement age at beginning of year

TABLE 4. Transition Probabilities and Costs of Reconstruction

$x(\tau)$ (1)	$x(1)$ 7.00 (2)	6.00 (3)	$ca[x(\tau)]$ ($/m³) (4)
7.00	0.92	0.08	31.06
6.00	0.92	0.08	31.06
5.00	0.92	0.08	31.06
4.00	0.92	0.08	31.06
3.00	0.92	0.08	31.06
2.00	0.92	0.08	31.06
1.00	0.92	0.08	31.06
0.00	0.92	0.08	31.06

Note: τ = pavement age at beginning of year ($\tau = 0, 1, \ldots, 19$)

The planning horizon of this case study was set to 20 years. The discount rate was assumed to be 5%, which corresponds to a discount amount factor of $1/(1 + 0.05)^1 = 0.9524$.

The value of the discounted total minimum expected cost, as computed by the latent Markov decision process algorithm, is: \$15.7/m² (\$13.1/sq yd). More extensive empirical work, to investigate the effect of the magnitude of the standard deviations of the measurement error model and the deterioration model on the minimum expected cost, is given in Ben-Akiva et al. (1991).

Conclusion and Further Research

This paper presented a new methodological approach for infrastructure management. This approach explicitly recognizes the uncertainty in both facility condition measurement and prediction, and incorporates this information in a decision-making model based on the latent Markov decision process.

The paper also presented the estimation of the two major inputs to the decision-making model:

- The measurement-error models, which describe the uncertainty in performance indicator measurement.
- The facility performance model, which allows for the combination of these separate indicators into a performance index and describe the uncertainty in performance forecasting.

Several refinements on the methodologies presented in this paper are possible. Some of these are:

1. Linking the latent performance variable to different measures of user costs. This will allow us to include user costs in the cost minimization algorithm of the decision-making model, instead of relying on minimum performance standards, as is the case now. This can be achieved by estimating a system of user cost equations simultaneously with the performance model system described previously. Such an effort is currently in progress at M.I.T.
2. Extending the joint decision-making algorithm to handle system-level considerations, such as a budget constraint. The problem illustrated throughout this paper was a facility-level problem; the system-level decision of "which facilities to allocate funds to in a given year" was not explicitly addressed. It is possible to achieve such an extension through the use of "randomized policies." Application of randomized policies to traditional Markov decision processes in pavement management have been successfully implemented. Some issues pertaining to their implementation for the LMDP are discussed in Madanat (1991).

Acknowledgments

Support was provided by a grant from the Army Research Office to the Center for Construction Research and Education of the Civil Engineering Department at Massachusetts Institute of Technology. We are indebted to Sue McNeil, who was responsible for the early stages of this research while at Massachusetts Institute of Technology.

Appendix I. References

"The AASHO road test." (1962). *Special Report 73*, Nat. Academy of Sci., Nat. Res. Council, Washington, D.C.
Ben-Akiva, M., Humplick, F., Madanat, S., and Ramaswamy, R. (1991). "Latent performance approach to infrastructure management." *Transp. Res. Record* 1311.
Ben-Akiva, M., and Ramaswamy, R. (1989). "Estimation of latent pavement performance from damage measurements." *Selected Proc. 5th World Conf. on Transportation Research*, WCTR Society.
Carnahan, J. V., Davis, W. J., Shahin, M. Y., Keane, P. L., and Wu, M. I. (1987). "Optimal maintenance decisions for pavement management." *J. Transp. Engrg.*, ASCE, 113(5), 554–572.
Feighan, K. J., Shahin, M. Y., Sinha, K. C., and White, T. D. (1988). "An application of dynamic programming and other mathematical techniques to pavement management systems." *Transp. Res. Record*, 1200.
Hudson, W. R., Elkins, G. E., Uddin, W., and Reilley, K. T. (1987). "Improved

methods and equipment to conduct pavement distress surveys," *Final Report, # FHWA-TS-87-213*, Federal Highway Admin., Washington, D.C.
Humplick, F. (1989). "Theory and methods of analyzing infrastructure inspection output: Application to highway pavement surface condition evaluation," Ph.D. Dissertation, Massachusetts Inst. of Tech., Cambridge, Mass.
Humplick, F. (1992). "Highway pavement distress evaluation: Modeling measurement error." *Transp. Res. B*, 26B(2), 135–154.
Madanat, S. (1991). "Optimizing sequential decisions under measurement and forecasting uncertainty: Application to infrastructure inspection, maintenance and rehabilitation," PhD Dissertation, Massachusetts Inst. of Tech., Cambridge, Mass.
Madanat, S. (1992). "Optimal infrastructure management decisions under uncertainty." *Transp. Res. C*, (to be published).
Ramaswamy, R. (1989). "Estimation of latent pavement performance from damage measurements," PhD Dissertation, Massachusetts Inst. of Tech., Cambridge, Mass.
Ramaswamy, R., and Ben-Akiva, M. (1990). "Estimation of highway pavement deterioration from in-service pavement data." *Transp. Res. Record*, 1272, 96–106.
Shahin, M. Y., and Kohn, S. D. (1981). "Pavement maintenance for roads and parking lots." *Tech. Report M-294*, Construction Engrg. Res. Lab., U.S. Army Corps of Engineers, Champaign, IL.
Smallwood, R. and Sondik, E. (1973). "The optimal control of partially observable Markov processes over a finite horizon." *Operations Res.*, 21, 1071–1088.

APPENDIX II. NOTATION

The following symbols are used in this paper:

$\mathbf{A}$ = vector of maintenance activities;
$\mathbf{D}$ = vector of distress indicators;
$\hat{d}_{lm}$ = measured value of indicator by technology m;
d_1 = true value of indicator;
I_t = state of information at beginning of t;
$P(I_{t+1} = K|I_t, a_t)$ = probability of facility being in information state K at beginning of $t + 1$ given information state at beginning of t and activity performed during t;
$p(x_t = j|x_{t-1} = i, a_{t-1})$ = probability of facility being in condition state j at beginning of t given it was in state i at beginning of $t - 1$ and activity a was performed during $t - 1$;
$q(\hat{x}_t = k|x_t = j)$ = probability of measuring condition state k given that true condition state of facility is j;
S = latent facility performance;
$\mathbf{X}$ = vector of explanatory variables (traffic, environment);
$\hat{x}_t$ = measured condition state of facility at beginning of time period t;
α_m, β_m = additive and multiplicative systematic biases of technology m;
a_{t-1} = M&R activity performed during $t - 1$;
ε_{lm} = random error of measurement by technology m with mean = 0, and standard deviation = Ψ_m; and
x_t = facility condition state at beginning of time period t.