Transportation Research Part B 43 (2009) 311–324

Contents lists available at ScienceDirect

Transportation Research Part B

journal homepage: www.elsevier.com/locate/trb

Optimal infrastructure condition sampling over space and time for maintenance decision-making under uncertainty

Rabi G. Mishalani [a,*], Liying Gong [b]

[a] The Ohio State University, Civil and Environmental Engineering and Geodetic Science, 2070 Neil Avenue, Room 470, Colombus, OH 43210, United States
[b] Capital One, 15000 Capital One Drive, Richmond, VA 23238, United States

ARTICLE INFO

Article history:
Received 3 June 2007
Received in revised form 10 July 2008
Accepted 10 July 2008

Keywords:
Infrastructure management
Inspection, maintenance, and rehabilitation decision-making
Optimization under uncertainty
Condition assessment
Spatial sampling

ABSTRACT

Infrastructure management is the process through which inspection, maintenance, and rehabilitation (IM&R) decisions are made to minimize the total life-cycle cost. Measurement, forecasting, and spatial sampling are three main sources of errors introducing uncertainty into the process. The first two uncertainties are captured in the infrastructure management literature. However, the third one has not been recognized and quantified. This paper presents a methodology where the spatial sampling uncertainty in question is captured and the sample size is incorporated as a decision variable in an optimization framework. An illustrative realistic example is presented to demonstrate an application of the developed framework. The results indicate that by not addressing the sampling uncertainty and decisions, the optimum IM&R decisions would not be achieved, and consequently, marked unnecessary overspending could take place.

1. Introduction and motivation

Infrastructure management, consisting of integrated mathematical tools, is utilized to support infrastructure agencies in making decisions on inspection, maintenance, and rehabilitation (IM&R) activities with the objective of minimizing the total life-cycle cost. After initial construction, infrastructure facilities deteriorate naturally over time mainly due to usage, environmental effects, and aging. Given the complex nature of deterioration, inspections are regularly made to assess the facility condition and to forecast its deterioration. Based on inspection outcomes, maintenance and rehabilitation (M&R) activities are performed to counter deterioration. The corresponding economic effects include user, inspection, and M&R costs, which computed over time amount to the total life-cycle cost. More accurate assessment of facility condition through inspection actions, have the potential to lead to more effective M&R decisions. Consequently, the expected combined user costs and M&R costs are reduced over the planning horizon. However, more accurate information requires more resources such as increased inspection frequency, advanced inspection technologies and data processing methods, larger sample sizes, or less correlated observations. Therefore, the optimum combination of inspection decisions and M&R decisions should be determined based on an economic evaluation that captures the long-term costs and benefits.

In the inspection data collection and utilization process, several sources of errors introduce uncertainty. In this study a field is defined as a section of infrastructure where condition behaves homogeneously over space and time (Mishalani and Koutsopoulos, 1995, 2002) and represents a facility of length L for which decisions are made. Henceforth the terms field, facility, and section are used interchangeably. A condition variable x – examples in the case of roadways include the extent of some pavement distress such as cracking or rutting, or the value of some aggregate condition variable, such as the Pavement Condition Index (PCI) (Shahin and Kohn, 1981) – is distributed along a field with mean μ_x, variance σ_x^2, and spatial correlation

* Corresponding author.
E-mail address: mishalani@osu.edu (R.G. Mishalani).

0191-2615/$ - see front matter
doi:10.1016/j.trb.2008.07.003

ρ_x. The values of condition variable x at n sample locations $\{x_1, \ldots, x_i, \ldots x_j, \ldots, x_n\}$ vary around μ_x due to inherent spatial processes discussed by Mishalani and Koutsopoulos (1995, 2002) where it is argued and empirically shown that positive spatial correlation is exhibited. A set of measured condition values $\{\hat{x}_{1,r}, \ldots, \hat{x}_{i,r}, \cdots \hat{x}_{j,r}, \ldots, \hat{x}_{n,r}\}$ are observed values of $\{x_1, \ldots, x_i, \ldots x_j, \ldots, x_n\}$ using inspection technology r. These measured values reflect an additional degree of uncertainty due to measurement errors associated with measurement technology limitations and environmental effects (Humplick, 1992). The sample mean of measured condition $\bar{\hat{x}}_{n,r}$ is calculated from these observations, which is an estimate of the current condition of that section, and, in turn, is one primary input to forecasting future condition. Hence, the difference between the condition estimate $\bar{\hat{x}}_{n,r}$ and the true condition μ_x reflects both measurement and spatial sampling uncertainty, where the latter is due to the spatial variation in the condition variables x and positive correlation among them.

Therefore, two types of errors are involved in this process: measurement error and spatial sampling error. Together with forecasting error, which is introduced in predicting future condition, the combination of these sources of uncertainty could affect the resulting M&R decisions through their influence on the quality of condition estimation and forecasting.

This study focuses on the prevalent discrete-state-based approach to decision-making commonly adopted in practice. To reduce the number of possible continuous condition states from infinity to a manageable finite number for the purpose of decision-making, it is necessary to discretize the continuous state space. Therefore, the effect of discretization on uncertainty must also be captured. Such a treatment is discussed in Section 3.2 of this paper.

Humplick (1992) developed a measurement error model whereby the difference between the true value of a distress variable at a given location and its measured value is explained in terms of systematic biases and a random error. Madanat (1991, 1993), Ben-Akiva et al. (1993), Madanat and Ben-Akiva (1994), and Ellis et al. (1995) extended the traditional infrastructure management framework known as the Markov Decision Process (MDP), which captures forecasting uncertainty, to the Latent Markov Decision Process (LMDP) by incorporating the measurement error model into the decision-making framework. The LMDP framework therefore captures both forecasting and measurement uncertainty. Nevertheless, the uncertainty associated with inspection spatial sampling has not been recognized or quantified. Not capturing this uncertainty can lead to sub-optimal IM&R decisions. Hence, even though the cost associated with inspection activities might be small, the impact of the sampling uncertainty on M&R decisions could still be appreciable.

This paper focuses on presenting a methodology to capture inspection sampling uncertainty within the LMDP decision-making framework and to optimize inspection spatial sampling decisions in the field of infrastructure management. Through developing a sampling error model and incorporating it into the LMDP framework, the resulting new framework, which is henceforth referred to as the Latent Markov Decision Process in the presence of Spatial Sampling (LMDPSS), is able to comprehensively capture the uncertainties associated with infrastructure condition assessment and forecasting and to determine optimum IM&R decision policies.

In the following section a brief overview of the LMDP framework is provided. Section 3 presents the derivation and formulation of sampling error, cost components, and objective function for IM&R decision-making in the presence of spatial sampling. In the fourth section a realistic illustrative example is presented to demonstrate an application of the LMDPSS framework, followed by a comparison between the decision policies arrived at by the LMDP and LMDPSS in order to assess the value of the LMDPSS framework. (A comprehensive evaluation approach is discussed in Mishalani and Gong (2007, 2008) in detail.) The last section summarizes the paper and includes a discussion of possible future research. Some mathematical details and implementation related issues are presented in the Appendices.

2. Review of LMDP framework

The LMDP framework is an extension of the traditional MDP framework, but differs in several aspects including the recognition that condition measurements are not error-free. Measurement technologies of different accuracies are assumed to be available in the LMDP framework. Madanat and Ben-Akiva (1994) also included the inspection frequency as part of the decision process where inspection costs are reflected in the life-cycle cost minimization.

In the LMDP framework the facility condition at any time is represented by one of several finite discrete states (which, as mentioned in the introduction, is consistent with the state-of-practice commonly employed at many infrastructure agencies) and the deterioration and renewal process is represented by discrete transition probabilities of the following form:

$$p(x_{t+1}^d = j | x_t^d = i, a_t); \quad 1 \leqslant i, j \leqslant w; \quad t = 0, 1, \ldots, T-1 \tag{1}$$

where x_t^d is the discrete true condition state of the facility at the beginning of period t, w is the total number of condition states, i and j are condition state indices, a_t is the M&R action applied at the beginning of period t, and T is the planning (time) horizon. This conditional probability indicates the likelihood that the facility condition is in state j at time $t+1$ given that its condition is in state i at time t. The transition probability captures both the deterioration process and the impact of the performed maintenance activity and, as a result, reflects forecasting uncertainty.

With the measurement uncertainty introduced, what the decision-maker measures at the beginning of time t results in an observed condition state, which is probabilistically related to the true condition state of the facility. This relationship is mathematically stated as follows:

$$q(\hat{x}_t^d = k | x_t^d = i, r_t) = f(\sigma_r^2); \quad 1 \leqslant i, k \leqslant w; \quad t = 0, 1, \ldots, T-1 \tag{2}$$

where $\hat{x}_t^d$ is the discrete condition state based on measured condition at time t, $q(\cdot|\cdot)$ is the conditional probability that the observed discrete condition state $\hat{x}_t^d$ is k given that the true discrete condition state x_t^d is i, r_t is the measurement technology index of the inspection made at beginning of time period t, $f(\cdot)$ is a known function, and σ_r^2 is the variance of the measured condition using inspection technology r after it has been corrected for the systematic bias captured by the measurement error model developed by Humplick (1992).

Madanat and Ben-Akiva (1994) considered the no-inspection case as a totally imprecise technology with zero cost, since the effect of not inspecting the facility at the beginning of period t is to leave the previous information unchanged. That is, the measurement probability for the no-inspection case is given by the following:

$$q(\hat{x}_t^d = k | x_t^d = i, r_t = \text{no-inspection}) = \frac{1}{w} \quad \forall i, k \tag{3}$$

Eq. (3) reflects the situation where all observed condition states are equally likely no matter what the true state is. This situation represents an inspection technology which provides no information to the decision-maker and, therefore, does not change any previous information.

To address the limitation of the traditional MDP framework whereby the state at a given time only reflects the condition at that time, and to avoid making M&R decisions only on the basis of the observed condition of the facility when measurement uncertainty is present, in the LMDP framework a new state is defined to account for all the information available to the decision maker at a given time and relevant to future decisions (Madanat, 1991, 1993; Madanat and Ben-Akiva, 1994). The information available to the decision maker at the beginning of time t includes the entire history of measured states up to time t and the decisions made up to time $t-1$. This information represents the augmented state at time t, denoted by I_t as follows:

$$I_t = \{I_0, a_0, r_1, \hat{x}_1^d, \ldots, \hat{x}_{t-1}^d, a_{t-1}, r_t, \hat{x}_t^d\}; \quad t = 1, 2, \ldots, T \tag{4}$$

where $I_o = \{r_{-\tau}, \hat{x}_{-\tau}^d, a_{-\tau}, \ldots, \hat{x}_{-1}^d, a_{-1}, r_0, \hat{x}_0^d\}$, and τ is the number of time periods (a time period is typically one or two years in the case of roadways) between the first inspection of the facility and the beginning of the planning horizon.

Under this representation the evolution of the state of information is modeled in a manner where the current condition depends on all the past condition states, which reflects an important improvement over the MDP framework. In the LMDP framework, the probabilistic evolution of the state of information is denoted by the following:

$$P(I_t | I_0, a_0, r_1, \hat{x}_1^d, \ldots, \hat{x}_{t-1}^d, a_{t-1}, r_t) = P(I_t | I_{t-1}, a_{t-1}, r_t); \quad t = 1, 2, \ldots, T \tag{5}$$

And, the probability mass function of the true condition state at t conditional on I_t, is denoted by the following:

$$p_t(x_t^d | I_t) \quad \forall x_t^d, \ \forall I_t, \ \forall t \tag{6}$$

The probabilities of the true condition given the augmented state can be calculated using Bayes' law where the measurement probabilities and the transition probabilities are known. In vector form, this probability mass function is denoted by the following:

$$P_t | I_t \quad \forall I_t, \ \forall t \tag{7}$$

where $P_t|I_t$ is a w-dimensional vector, which is referred to as the information vector. With the introduction of the augmented state, the same solution method of dynamic programming that is used to find optimal M&R policies by minimizing expected costs in the MDP framework is adopted in the LMDP framework.

3. LMDPSS framework

3.1. Spatial sampling error

As was discussed in Section 1, there are sampling errors in addition to measurement errors contributing to the uncertainty of inspection results. In this section the mathematical representation of sampling error assuming a constant distance between adjacent observations is introduced and, subsequently, the representation where this assumption is relaxed is presented.

3.1.1. Estimate of facility condition

The facility condition μ_x is estimated by the sample mean of condition observations within a field, and represents one primary input to IM&R decision-making. This sample mean is given by the following:

$$\bar{\hat{x}}_{n,r} = \frac{1}{n} \sum_{l=1}^{n} \hat{x}_{lr} \tag{8}$$

where $\bar{\hat{x}}_{n,r}$ is the sample mean of condition observations across the facility, $\hat{x}_{lr}$ is the measured condition at location l ($l = 1, 2, \ldots, n$) using measurement technology r, and n is the sample size. In this study, it is assumed that once inspection is taking place at a facility, only one inspection technology is adopted. This is a reasonable assumption given the following.

First, several inspection technologies in effect consist of a system of integrated observational techniques (including different sensors mounted on a van in the case of roadway infrastructure inspection) whereby the characteristics of several distress types (the values of which could comprise inputs to calculating aggregate condition variables such as the PCI) could be measured simultaneously by a single inspection technology of this integrated kind. Second, an extra cost would be associated with simultaneously adopting multiple alternative technologies (of the integrated type or otherwise) for the same condition variable or set of condition variables.

It is also assumed that condition measurements are corrected for the systematic biases modeled by Humplick (1992) before being used further in decision-making. This correction is possible under certain conditions, namely when the measurement error model is known, and therefore this assumption is reasonable. That is, the following holds:

$$\hat{x}_{lr} = x_l + \varepsilon_{lr} \tag{9}$$

where x_l is the true condition at location l, and ε_{lr} is the random measurement error of mean zero at location l using measurement technology r after systematic biases have been corrected for.

3.1.2. Statistics of condition estimate

Based on statistical theory, it is clear that the sample mean calculated in Eq. (8) is an unbiased estimate of the mean of the condition variable over a field. The uncertainty associated with the spatial sampling is captured by the variance of the sample mean, which is denoted by $\text{Var}(\bar{\hat{x}}_{n,r})$ and estimated as follows:

$$\text{Var}(\bar{\hat{x}}_{n,r}) = \text{Cov}\left(\frac{\hat{x}_{1r} + \hat{x}_{2r} + \cdots + \hat{x}_{nr}}{n}, \frac{\hat{x}_{1r} + \hat{x}_{2r} + \cdots + \hat{x}_{nr}}{n}\right) = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\text{Cov}(\hat{x}_{ir}, \hat{x}_{jr}) \tag{10}$$

In order to arrive at a specific solution to Eq. (10), the following assumptions are made:

(i) $\text{Cov}(x_i, \varepsilon_{ir}) = 0 \ \forall i, r$ (11)

(ii) $\text{Cov}(\varepsilon_{ir}, \varepsilon_{jr}) = 0 \quad \text{if } i \neq j \ \forall r, \quad \text{and}$ (12)

(iii) $\{x_i\}$ is a stationary process, which requires that the following holds (Brockwell and Davis, 2002):

$$\text{Cov}(x_i, x_j) = \gamma(j - i) = \gamma(s), \tag{13a}$$
$$\gamma(s) = \gamma(-s), \tag{13b}$$
$$\gamma(0) = \sigma_x^2, \quad \text{and} \tag{13c}$$
$$\gamma(s) = \rho(s) \cdot \sigma_x^2 \tag{13d}$$

where $\gamma(s)$ is the covariance between two condition variables within the field that are separated from one another by a distance of s, and $\rho(s)$ is the correlation between condition variables at two locations within the field that are separated from one another by a distance s.

Eq. (11) reflects the independence between the condition variable and the random measurement error at the same location and Eq. (12) reflects the independence among the measurement errors associated with measured condition at different locations. Eqs. (13a) and (13b) reflect the situation where the covariance between condition variables at two locations is only a function of the absolute distance between them. Eq. (13c) reflects the fact that the condition variable x at different locations has constant variance σ_x^2, and as a result the covariance in Eq. (13d) is equal to the product of the correlation coefficient and the constant variance.

Based on Eqs. (9), (11), and (12), a solution to Eq. (10) can be derived as follows:

$$\begin{aligned}\text{Var}(\bar{\hat{x}}_{n,r}) &= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\text{Cov}(x_i + \varepsilon_{ir}, x_j + \varepsilon_{jr}) = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left[\text{Cov}(x_i, x_j) + \text{Cov}(x_i, \varepsilon_{ir}) + \text{Cov}(x_j, \varepsilon_{jr}) + \text{Cov}(\varepsilon_{ir}, \varepsilon_{jr})\right] \\ &= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left[\text{Cov}(x_i, x_j) + \text{Cov}(\varepsilon_{ir}, \varepsilon_{jr})\right] = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n}\text{Cov}(x_i, x_j)}{n^2} + \frac{\sigma_r^2}{n}\end{aligned} \tag{14}$$

In order to specify the covariance among condition variables at different locations in Eq. (14), $\text{Cov}(x_i,x_j)$, the stationary assumption represented by Eqs. (13a)–(13d) is applied with the additional initial assumption of a constant distance between adjacent observations. The case of a variable distance between adjacent observations is discussed subsequently.

If the distance between locations i and $i + 1$ is constant for all i, the optimum value for this constant distance denoted by h is $L/(n - 1)$. In other words, the largest possible constant distance between adjacent observations results in the most precise estimation of the facility condition through the sample mean. The mathematical proof is presented in Appendix A. With this additional assumption over the stationary assumption, the specific solution to Eq. (14) is the following:

$$\text{Var}(\bar{\hat{x}}_{n,r}) = \frac{\sigma_r^2}{n} + \frac{\sigma_x^2}{n} + \frac{2\sigma_x^2\sum_{i=1}^{n-1}(n - i) \cdot \rho(i \cdot h)}{n^2} \tag{15}$$

When the distances between adjacent observations are not assumed to be constant, they are denoted by h_i, $i = 1, \ldots, n-1$, and have to satisfy the constraint $\sum_{i=1}^{n-1} h_i \leqslant L$. As a result, the solution to Eq. (14) is the following:

$$\mathrm{Var}(\bar{x}_{n,r}) = \frac{\sigma_r^2}{n} + \frac{\sigma_x^2}{n} + \frac{2\sigma_x^2 \sum_{i=1}^{n-1} \sum_{j=1}^{i} \rho(\sum_{s=1}^{n-i} h_{j+s-1})}{n^2} \tag{16}$$

Eqs. (15) and (16) combine both the measurement and sampling uncertainties through the variance of the corrected condition measurement σ_r^2, variance of the condition variable σ_x^2, spatial correlation among observations $\rho(s)$, and sample size n. Clearly, the larger the sample size is, the smaller the contributions of the variances of the condition variable and its measurement are to the overall variance. And, the smaller the correlation among observations is, which is captured by the last additive element of Eqs. (15) and (16), the more accurate the estimate of facility condition is. The only difference between Eqs. (15) and (16) is in the last additive element of each, and both elements reduce to zero if the condition variables at different locations are independent.

In the general case of variable spacing between adjacent observations, it can be mathematically shown that for at least small sample sizes n, a constant spacing is optimal when "edge effects" – where the absence of information on either end of the field skews the results in the vicinity of the edges (Cressie, 1993) – are excluded (Gong, 2006). Furthermore, with the additional assumption of constant spacing between adjacent observations, the number of variables associated with spatial sampling affecting the accuracy of facility condition estimation is reduced markedly from $n + 1$ (n location variables plus 1 sample size variable) to 1 (the sample size variable). And, as discussed earlier and proven in Appendix A, the largest possible constant distance between adjacent observations, $h = L/(n-1)$, minimizes the uncertainty resulting from spatial sampling. Therefore, this largest possible constant distance is assumed in the remainder part of this paper.

Since the estimate of mean condition within a field is the basis for IM&R decision-making, the nature of the sample mean statistic is further developed by introducing distributional assumptions. The condition variable x is assumed to follow a Normal distribution with mean of μ_x and a variance of σ_x^2. The corrected condition measurement is assumed to follow a Normal distribution with mean zero and a variance of σ_r^2. It follows that the distribution of the sample mean $\bar{x}_{n,r}$ is Normal with mean of μ_x and a variance as given by Eq. (15) (Casella and Berge, 2002).

3.2. Measurement and sampling probability

To quantify the effect of the spatial sampling uncertainty on facility condition assessment, the measurement probability represented by Eq. (2) is revised and now is referred to as the measurement and sampling probability. It is mathematically represented as follows:

$$q(\bar{x}^d_{n,r,t} = k | \mu^d_{x,t} = i, n_t, r_t) = f(\sigma_x^2, \sigma_r^2, \rho, n_t) \tag{17}$$

where $\bar{x}^d_{n,r,t}$ is the discrete condition state based on the continuous value of the sample mean across condition observations within the field at time t, and $\mu^d_{x,t}$ is the discrete true condition state based on the true mean of the condition variable for the given field at time t.

Comparing Eq. (17) with Eq. (2), it is clear that the measurement and sampling probability is not only a function of the factor related to the inspection technology, σ_r^2, as in Eq. (2), but also a function of factors corresponding to the properties of the observed facility, σ_x^2 and $\rho(s)$, and the sample size, n_t where the index t represents the time the sample is taken.

To capture the discretization of continuous condition into discrete states, the measurement and sampling probability of Eq. (17) is calculated as follows:

$$q(\bar{x}^d_{n,r,t} = k | \mu^d_{x,t} = i, n_t, r_t, \rho) = \frac{\int_{\mu_{x,t} \in i} \int_{\bar{x} \in k} f_{\bar{x}}(\bar{x}_{n,r,t} | \mu_{x,t}) \mathrm{d}\bar{x}_{n,r,t} \mathrm{d}\mu_{x,t}}{\int_{\mu_{x,t} \in i} f(\mu_{x,t}) \mathrm{d}\mu_{x,t}} \tag{18}$$

where $f(\mu_{x,t})$ is the probability density function (PDF) of the true mean of condition variable at time t assumed uniform over the condition interval defining the condition state $\mu^d_{x,t}$, and $f_{\bar{x}}(\bar{x}_{n,r,t} | \mu_{x,t})$ is the conditional PDF of the sample mean of measured condition variables given that the true mean is $\mu_{x,t}$, which is discussed at the end of Section 3.1.2.

3.3. Cost components

The costs of interest in infrastructure management and IM&R decision-making relate to M&R activities, inspection, users, and terminal condition of the facility in question. Each of these components is discussed in more detail next.

3.3.1. M&R cost

The cost of an M&R activity depends on its type and extent. The activity type can range from basic routine maintenance to major rehabilitation. The extent of a particular activity depends on the true condition state of the facility and has been addressed in earlier studies (Sharaf et al., 1985; Carnahan et al., 1987). Thus, the maintenance and rehabilitation cost is denoted by $R(a, \mu_{x,t})$.

3.3.2. Inspection cost

Inspection cost is incurred at the beginning of each period (which, as mentioned earlier is typically one or two years in the case of roadways) if inspection is scheduled to take place. Since there is a lack of detailed studies on this cost, a general representation is adopted as follows:

$$M(r,n) = f_r + u_r \cdot n^{1/c} \tag{19}$$

where $M(r,n)$ is the cost of inspection with measurement technology r and sample size n, f_r is the fixed cost when inspection technology r is applied, u_r is the unit observation cost using technology r, and c is the economies of scale factor with $c \geqslant 1$.

The fixed cost consists of many components, such as field data collection equipment cost, labor safety training and insurance cost, data processing equipment cost, and the logistics cost of transporting equipment and labor from the depot to the inspection sites. Unit costs occur whenever an observation is made (i.e., a sample is taken), which is determined by the inspection time and the labor hourly wage.

3.3.3. User cost

One of the components of user cost captures the impact of deterioration on the users as a function of the condition of the facility. The reduction in the user cost resulting from condition improvements is the benefit that the users derive from the application of M&R activities. Madanat (1991, 1993) and Madanat and Ben-Akiva (1994) did not quantify this cost because of the absence of a realistic user cost function. Instead, the minimum allowable condition state for the facility was introduced as a constraint while minimizing the agency cost (i.e., the cost associated with IM&R activities). Since then, Ben-Akiva and Gopinath (1995) developed a model that specifies the user cost as a function of underlying performance S and a set of independent variables X_1 including highway geometry, traffic composition, and traffic volume. This user cost model is adopted in the LMDPSS framework in order to capture user cost directly in the objective function rather than in the form of a constraint as done in earlier studies. The user cost is denoted by $U_1(S, X_1)$. It is important to point out that the condition mean μ_x of a facility is a manifestation of its performance S.

In addition, maintenance and rehabilitation results in further user cost in the form of delays at work zones. This cost has not been captured in the implementation of the LMDP framework. Recent research (Vadakpat et al., 2000) resulted in a model for predicting the additional user cost due to roadway work zones, which is a function of a set of variables X_2 including lane location, truck percentage, and traffic volume for which the delay is to be computed. This study introduces this model into the implementation of the LMDPSS framework. This additional user cost is denoted by $U_2(a, X_2)$. Similarly, inspection results in traffic disruption cost, which is denoted by $U_3(r, X_2)$. Finally, user cost is a function of network effects. However, since this study focuses on the facility level, such effects are not modeled.

3.3.4. Terminal cost

In this study terminal cost is defined as the hypothetical expense incurred at the end of the planning horizon for bringing the facility condition back to the best condition level in order to equalize the service life from that point onward. It is a function of the true state of the facility at time T. If the facility is in good condition, the facility can continue to supply high level of service from that point onward resulting in a relatively low terminal cost. However, if the condition is poor, the facility's potential to provide a reasonable level of service from that point onward is low resulting in a relatively high terminal cost. Thus, this cost is denoted by $\mathrm{TC}(\mu_{x,T})$.

3.4. Objective function and dynamic programming

The three main revisions to the LMDP framework required as a result of introducing spatial sampling are the following:

(i) the measured condition state is replaced by the condition sample mean,
(ii) the uncertainty associated with the sample mean is captured, and
(iii) the sample size becomes a decision variable.

These main revisions are mathematically incorporated into the LMDPSS framework, which are discussed in more detail next.

3.4.1. Information vector and its evolution

Since now the spatial sampling process is incorporated, the augmented state discussed in Section 2 is revised as follows:

$$I_t = \{I_0, a_0, n_1, r_1, \bar{\hat{x}}_1^d, \ldots, a_{t-1}, n_t, r_t, \bar{\hat{x}}_t^d\}; \quad t = 1, 2, \ldots, T \tag{20}$$

where $I_o = \{\bar{\hat{x}}_{-\tau}^d, a_{-\tau}, \ldots, a_{-1}, n_0, r_0, \bar{\hat{x}}_0^d\}$. Moreover, the probabilistic evolution of the state of information is denoted by the following:

$$P(I_t|I_0, a_0, n_1, r_1, \bar{\hat{x}}_1^d, \ldots, a_{t-1}, n_t, r_t) = P(I_t|I_{t-1}, a_{t-1}, n_t, r_t), \quad t = 1, 2, \ldots, T \tag{21}$$

The probability mass function of the true condition state at time t conditional on I_t (consisting of the elements of the information vector $P_t|I_t$) is calculated using Bayes' law, the measurement and sampling probabilities, and the transition probabilities as follows:

$$p_t(\mu^d_{x,t}=j|I_t)=\frac{p_t(\mu^d_{x,t}=j,I_t)}{p_t(I_t)}=\frac{p_t(\mu^d_{x,t}=j,I_{t-1},a_{t-1},n_t,r_t,\bar{\hat{x}}^d_t)}{p_t(I_{t-1},a_{t-1},n_t,r_t,\bar{\hat{x}}^d_t)}=\frac{q(\bar{\hat{x}}^d_t|\mu^d_{x,t}=j,n_t,r_t)p_t(\mu^d_{x,t}=j|I_{t-1},a_{t-1})}{\sum_{j'}q(\bar{\hat{x}}^d_t|\mu^d_{x,t}=j',n_t,r_t)p_t(\mu^d_{x,t}=j'|I_{t-1},a_{t-1})}$$
$$=\frac{q(\bar{\hat{x}}^d_t|\mu^d_{x,t}=j,n_t,r_t)\sum_i p(\mu^d_{x,t}=j|\mu^d_{x,t-1}=i,a_{t-1})p_{t-1}(\mu^d_{x,t-1}=i|I_{t-1})}{\sum_{j'}q(\bar{\hat{x}}^d_t|\mu^d_{x,t}=j',n_t,r_t)\sum_i p(\mu^d_{x,t}=j'|\mu^d_{x,t-1}=i,a_{t-1})p_{t-1}(\mu^d_{x,t-1}=i|I_{t-1})}\quad \forall j=1,\dots,w \tag{22}$$

where $p_t(\mu^d_{x,t}=j,I_t)$ is the joint probability of augmented state I_t and true condition of the facility at time t being in state j, $p_t(I_t)$ is the marginal probability of augmented state I_t, $q(\bar{\hat{x}}^d_t|\mu^d_{x,t}=j,n_t,r_t)$ is the measurement and sampling probability in the LMDPSS framework, and $p(\mu^d_{x,t}=j'|\mu^d_{x,t-1}=i,a_{t-1})$ is the transition probability in the LMDPSS framework. In Eq. (22), $p_t(\mu^d_{x,t}=j|I_{t-1},a_{t-1})$ captures the probability of true condition being in state j before an update with inspection results is available. If the information vector $P_0|I_0$ is assumed to be known, then the information vector $P_t|I_t$ can be calculated recursively using Eq. (22) for all t, starting from $t = 1$ to $t = T$.

3.4.2. Objective function

Based on the incorporation of spatial sampling and the various cost functions discussed above, the objective function of the LMDPSS framework is the following:

$$J_T(P_T|I_T)=TC(\mu_{x,T});$$
$$\begin{aligned}J_t(P_t|I_t)=\min_{(a_t,r_{t+1},n_{t+1})}\Bigg\{&\sum_{i=1}^{w}p_t(\mu^d_{x,t}=i|I_t)[R(a_t|\mu^d_{x,t}=i)+U_2(a_t,X_2)\\&+\alpha\sum_{j=1}^{w}p(\mu^d_{x,t+1}=j|\mu^d_{x,t}=i,a_t)\cdot U_1(\mu^d_{x,t+1}=j,X_1)+\alpha M(r_{t+1},n_{t+1})+\alpha U_3(r_{t+1},X_2)\\&\underbrace{+\alpha\sum_{j=1}^{w}p(\mu^d_{x,t+1}=j|\mu^d_{x,t}=i,a_t)\cdot\sum_{k=1}^{w}q(\bar{\hat{x}}^d_{t+1}=k|\mu^d_{x,t+1}=j,r_{t+1},n_{t+1})\cdot J_{t+1}(P_{t+1}|I_t,a_t,r_{t+1},n_{t+1},\bar{\hat{x}}^d_{t+1}=k)]}_{A}\Bigg\}\end{aligned}$$
$$t=1,\dots,T-1 \tag{23}$$

where $J_t(P_t|I_t)$ is the minimum expected cost-to-go at time t given that the condition state of the facility is represented by the information vector $P_t|I_t$, and α is the annual discount factor. The decision variables are M&R activity a_t, inspection technology r_{t+1}, and sample size n_{t+1} for each time period t. Also note that the terminal cost is assumed to take a non-zero value, which is a function of the facility's condition at time T.

In program (23), the element $\sum_{j=1}^{w}p(\mu^d_{x,t+1}=j|\mu^d_{x,t}=i,a_t)\cdot U_1(\mu^d_{x,t+1}=j,X_1)$ captures the expected user cost associated with all possible true condition states over time $t + 1$ given that the true condition at time t is in state i. The element labeled A in program (23) captures the minimum expected cost-to-go at time $t + 1$ across all possible observed condition states and transition outcomes given that the current true condition is in state i. The second summation term (over k) of this element captures the minimum expected cost-to-go across all possible observed condition states of the facility at time $t + 1$ given that the true condition at time $t + 1$ is in state j. The first summation term (over j) of this element incorporates the probability of the true condition of the facility at time $t + 1$ being in state j given that the true condition at time t is in state i. The other elements within the brackets in program (23) are self-explanatory. Summing all elements within the brackets provides the total cost from the current time onward given that the current true condition is in state i. The summation of the products of the above total cost with the probability of the true condition being in state i, which is given by $p_t(\mu^d_{x,t}=i|I_t)$ introduced earlier as expression (22), for $i = 1$ to $i = w$ results in the expected total cost associated with augmented state I_t at time t.

Based on all the expected total costs associated with each information vector, the optimal values of a_t, r_{t+1}, and n_{t+1} leading to the minimum expected cost-to-go at time t are determined for the corresponding augmented state at time t. This process continues recursively from $t = T$ to $t = 0$ for every information vector $P_t|I_t$ until $J_0(P_0|I_0)$ is obtained. By solving program (23) for a specified planning horizon T, the optimum total expected cost $J_0(P_0|I_0)$ and the optimum policy $\pi^*=\{\eta^*_0(P_0|I_0),\eta^*_1(P_1|I_1),\dots,\eta^*_{T-1}(P_{T-1}|I_{T-1})\}$ are determined. The expression $\eta^*_t(P_t|I_t)$ represents the optimum policy for time t, and specifies the optimum combination of M&R, inspection, and sampling activities for each possible information vector $P_t|I_t$ at time t.

3.4.3. Grouping of information vectors

The computational consideration involved in the computer implementation of the LMDPSS framework algorithm consists of a trade-off between computer running time and memory requirements on the one hand, and calculation accuracy on the other. Each information vector $P_t|I_t$ at time t leads to multiple information vectors at time $t + 1$. The number of information vectors at time $t + 1$ produced by $P_t|I_t$ is equal to the product of the number of maintenance activities N_a, the number of inspection technologies N_r, the number of considered sample sizes N_s, and the number of states w. The upper bound of the number of information vectors at each time is, therefore, given by $(N_a \times N_r \times N_s \times w)^t$, $t = 1, \ldots, T$. Consequently, the memory requirements of the algorithm are a limitation, especially when sampling is incorporated as a decision-making variable, whereby the dimensionality of the problem is further increased. Hence, combining similar information vectors into a representative vector becomes a possible solution to this difficulty. Some information vector grouping methods and the one developed in this study are discussed in Appendix B.

4. Numerical results and discussion

4.1. Example facility

In this analysis, a realistic 1 km long, 3.66 m wide homogeneous pavement section consisting of the right lane of a two-lane interstate highway is assumed. This field is divided into 100 subsections with equal area. Each subsection is considered a potential observation unit. Therefore, the maximum sample size is 100. The planning horizon T is set at 10 years and the interest rate at 5%, which corresponds to a discount amount factor α of 0.9524.

The condition variable is given by the PCI referred to in the introduction section, which takes a value ranging from 0 to 100. This range is equally divided into eight condition states (state 7 is the best and state 0 is the worst). The initial condition state is set at 7 with a probability of one. This discretization of the state space relies on previous studies (Carnahan et al., 1987; Feighan et al., 1988; Madanat, 1993; Madanat and Ben-Akiva, 1994). The maintenance actions considered are routine maintenance and rehabilitation. The corresponding transition probabilities are adopted from chapter 5 in Madanat (1991). Thus, 10 sets of transition probabilities in the presence of routine maintenance, one for each of the 10 years of the planning horizon, are specified. That is, the process is non-stationary in this case. It is also assumed that rehabilitation brings the condition of the facility back to the best possible state. This assumption has also been made in other studies (Madanat, 1993; Ben-Akiva et al., 1993; Madanat and Ben-Akiva, 1994).

The standard deviation of the condition variable as a function of the condition level is derived and discussed in more detail in Gong (2006). This relationship is mathematically expressed as follows:

$$\sigma_x = 192 - 1.90 \times \mu_{x,t} \tag{24}$$

In Eq. (24), the standard deviation of the condition variable σ_x and the condition level $\mu_{x,t}$ are in PCI.

There are 11 combinations of inspection activities: two types of inspection technologies and five levels of sample sizes (20%, 40%, 60%, 80% and 100%), in addition to the no-inspection option. The manual mapping method (henceforth referred to as technology 1) and the ARAN survey vehicle (henceforth referred to as technology 2) discussed by Humplick (1992) are considered as the two types of inspection technologies in this example. The standard deviations of inspection technologies (in PCI) are 7.99 and 17.95, respectively, which are computed as the product of the mean value (in PCI) with the coefficients of variation of these inspection technologies derived from data presented by Humplick (1992).

Mishalani and Koutsopoulos (1995, 2002) showed that the spatial correlations of pavement condition are positive and monotonically decreasing functions of the distance between observations. Therefore, the correlation function considered in this study reflects positively correlated observations and it is assumed to follow the exponential form as follows:

$$\rho(s) = \exp(\beta \cdot s) \qquad \forall \beta \leqslant 0, \;\; s \geqslant 0 \tag{25}$$

where β is the parameter of the exponential function, which is assumed to take a negative value in order to guarantee that the correlation coefficient is less than or equal to one. Such a function satisfies the necessary and sufficient condition of spatial correlation functions in one-dimensional space (Cressie, 1993), namely the covariance matrix should be positive-definite. Spatial correlation data of 11 fields (Mishalani, 1993) are used to estimate the parameter β of the exponential function associated with the PCI condition variable. The estimated value of the parameter β is found to be -0.026386 (Gong, 2006). Based on all of the above regarding inspection and field characteristics, the measurement and sampling probabilities are computed.

Four cost components are mentioned in Section 3.3. Specific cost functions for each are identified based on the literature. The numerical values are converted to the present value of the year 2003 through the inflation rate (InflationData.com, 2005). The results are shown in Table 1. More detailed discussions are presented by Gong (2006).

4.2. Comparison measures

In order to make meaningful comparisons between the LMDPSS and LMDP frameworks, in this section an adjusted LMDP framework (ALMDP) is introduced and measures useful for comparison purposes are subsequently defined.

4.2.1. Adjusted LMDP framework

In the traditional LMDP framework it is assumed that a unique observation is taken for estimating the condition of an infrastructure facility or field, which in turn is used in the decision-making process. It is important to note that the measurement probability in the LMDP framework, represented by Eq. (2), is only a function of the variance of the measurement random error since only a single observation is considered to represent the condition of a field. Based on basic statistical theory,

Table 1
Cost functions for the numerical example ($/m^2)

	Condition state							
	7	6	5	4	3	2	1	0
Costs that depend on condition state								
Routine maintenance cost	0.15	0.53	1.28	1.98	3.80	5.00	12.34	17.24
User cost	0.00	36.94	51.84	61.63	70.26	80.05	94.95	131.91
Terminal cost	0.15	5.20	14.06	30.91	46.52	64.87	64.87	64.87
Costs that do not depend on condition state								
Rehabilitation cost	64.87							
Additional user cost due to M&R	1.45 for routine maintenance and 10.13 for rehabilitation							
Additional user cost due to inspection	0.09 for inspection technology 1 and 0.0015 for inspection technology 2							
Fixed inspection cost	0.0012 for inspection technology 1 and 0.009 for inspection technology 2							
Unit inspection cost	0.00023 for inspection technology 1 and 0.000085 for inspection technology 2							

if multiple independent observations are averaged to represent the field condition, the variance of that mean condition would be σ_r^2/n.

In order to relax the assumption of a single observation in the original LMDP framework, the framework is revised for the purpose of this evaluation to reflect multiple observations by replacing σ_r^2 with σ_r^2/n. Therefore, the measurement probability in the adjusted LMDP (ALMDP) framework is calculated by the following:

$$q'(\hat{x}_t^d = k | x_t^d = i, r_t) = f(\sigma_r^2/n); \qquad 1 \leqslant i, k \leqslant w; \quad t = 0, 1, \ldots, T-1 \tag{26}$$

where $q'(\cdot|\cdot)$ is the measurement probability in the ALMDP framework. Notice that the variance of the mean condition of the ALMDP framework is the first additive element of the variance of the sample mean of the LMDPSS framework shown in Eq. (15).

All the comparisons conducted in this study are based on the ALMDP framework because it is straightforward for agencies to extend the LMDP framework to reflect a sample of condition observations within the limitations of that framework. Since the ALMDP framework is more advanced than the LMDP framework, the comparisons would not exaggerate the value of the LMDPSS framework.

4.2.2. Accuracy index

To compare the uncertainty associated with different combinations of inspection technologies and sample sizes, an "accuracy" index given measurement technology r and sample size n is defined to be the average value of the probabilities of estimating the condition state to be equal to the true condition state across all true states, which is expressed by the following equation:

$$AI_{n,r} = \frac{1}{w} \sum_{i=1}^{w} q(\bar{\hat{x}}_{n,r}^d = i | \mu_x^d = i) \tag{27}$$

where $AI_{n,r}$ is the accuracy index for the inspection activity with sample size n and inspection technology r.

Of course, this index represents the extent to which the estimated and true condition states are equal but does not capture the magnitude of the error between the estimated and true condition states when they are not equal. Therefore, this index provides an assessment of only one aspect of accuracy and is, therefore, not comprehensive in nature. Nevertheless, it gives an indication of an important aspect and, therefore, is worth examining. Naturally, the objective of the optimization is to minimize total life-cycle cost. However, the assessment of the accuracy of condition state estimates influences the optimization and, therefore, could prove useful in interpreting the numerical results as illustrated subsequently in Section 4.3.

4.2.3. Relative difference in expected total cost

Since the objective is to minimize the expected total cost, the value of the LMDPSS framework regarding meeting this objective is best captured by the relative difference between the expected total cost at optimality when applying the LMDPSS framework and the true cost associated with the optimal actions determined by applying the ALMDP framework. The latter is determined by simulating the optimal ALMDP-determined policy in the LMDPSS framework – see Mishalani and Gong (2008) for a detailed discussion of this evaluation approach.

The relative difference between the expected total costs associated with the decision policies determined by ALMDP and LMDPSS frameworks, denoted by ΔC, is mathematically expressed as a percentage by the following:

$$\Delta C = \frac{C_{\text{ALMDP}} - C_{\text{LMDPSS}}}{C_{\text{LMDPSS}}} \times 100\% \tag{28}$$

where C_{ALMDP} is the expected total cost associated with the ALMDP solution, and C_{LMDPSS} is the expected total cost associated with LMDPSS solution.

4.2.4. Relative difference in expected agency cost

The expected agency cost consists of the expected total IM&R cost, which could be used by infrastructure agencies in estimating their operating budget over the planning horizon. The LMDPSS framework may produce IM&R decision policies different from those produced by the ALMDP framework since it captures the uncertainty associated with the spatial sampling and incorporates the sample size as a decision variable. As a result, the two frameworks may lead to different expected agency costs. While the objective is to minimize expected total cost, differences in agency costs are indicative of the value of the LMDPSS framework in improving agency budget planning. Moreover, the comparison of the expected agency costs could also capture the extent of the differences in the IM&R decision policies determined by applying the two frameworks.

The relative difference between the expected agency costs associated with the decision policies determined by the ALMDP and LMDPSS frameworks, denoted by ΔAC, is mathematically expressed as a percentage by the following:

$$\Delta AC = \frac{AC_{\mathrm{ALMDP}} - AC_{\mathrm{LMDPSS}}}{AC_{\mathrm{LMDPSS}}} \times 100\% \tag{29}$$

where AC_{ALMDP} is the expected agency cost associated with the ALMDP solution, and AC_{LMDPSS} is the expected agency cost associated with LMDPSS solution.

4.3. Results and discussion

Fig. 1 shows the accuracy index for all combinations of inspection technologies and sample sizes under the ALMDP and LMDPSS frameworks. Note that the accuracy is increasing along with increasing sample size, which is consistent with statistical theory (for the case of independent observations). Also, note that for a given decision-making framework under the same sample size, inspection technology 1 is more accurate than technology 2, again as expected given that the variance of the measurements produced by technology 1 is smaller than the variance of those produced by technology 2.

Fig. 1 also allows for comparing the accuracy index values under the ALMDP framework with those under the LMDPSS framework. Not surprisingly, the accuracies under the ALMDP framework are higher than those under the LMDPSS framework reflecting the systematic overconfidence of a decision-maker operating under the ALMDP framework as evident from Eq. (15) whereby the actual variance of condition assessment captured in the LMDPSS framework consists of all three elements while the assumed variance under the ALMDP framework consists of the first element only as already discussed.

The optimum decision policies under the LMDPSS and ALMDP frameworks are determined over a 10-year planning horizon whereby the value of ΔC of Eq. (28) for this example is found to range between 2.64% and 4.59% depending on the predetermined sample size set under the ALMDP framework. While seemingly small, given the large costs involved, the actual magnitude of the saving resulting from applying the LMDPSS framework is substantial – see Mishalani and Gong (2007, 2008) for a more detailed analysis and discussion.

The optimum policy results for years 0, 1, and 2 are shown in Tables 2 and 3 as determined by each of the LMDPSS and ALMDP frameworks, respectively. Both frameworks produce the same optimum decision policies at time 0 for this specific example. However, the decision policies start to deviate from one another from year 1 onward. As already discussed, under the LMDPSS framework the agency reflects an accurate assessment of inspection uncertainty, while under the ALMDP framework the agency reflects systematic overconfidence in the accuracy of inspections resulting in different optimum policies produced by each of the frameworks.

For example, under the ALMDP framework in year 2, due to the additional data resulting from the predetermined 100% sample size, in comparison to the optimal 60% for year 2 determined by applying the LMDPSS framework (shown under the year 1 column in Table 2 as it is a decision made in year 1 to be applied in the following inspection period), the true

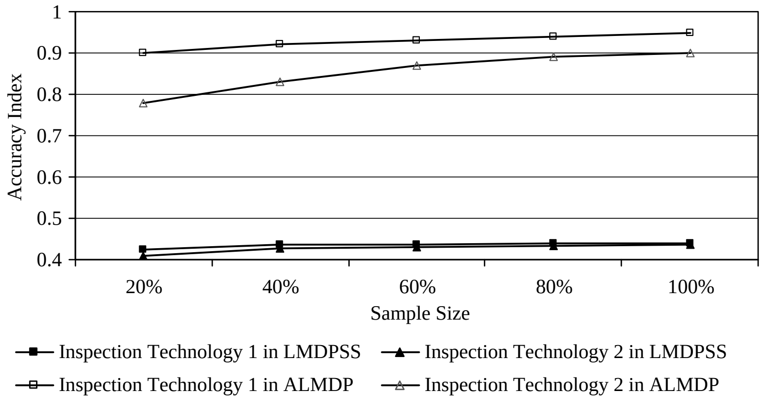

Fig. 1. Accuracy index values of inspection activities.

Table 2
Example of optimal decision policies under LMDPSS

Assessed condition state	Optimal decision policies for first three time periods		
	Year 0	Year 1	Year 2
7	Routine maintenance No-inspection	Routine maintenance Inspection tech 2 Sample size 60%	Routine maintenance Inspection tech 2 Sample size 40%
6		Routine maintenance Inspection tech 2 Sample size 60%	Routine maintenance Inspection tech 2 Sample size 100%
5–0		Routine maintenance Inspection tech 2 Sample size 60%	Rehabilitation No-inspection

Table 3
Example of optimal decision policies under LMDP

Assessed condition state	Optimal decision policies for first three time periods		
	Year 0	Year 1	Year 2
Predetermined sample size = 20%			
7	Routine maintenance No-inspection	Routine maintenance No-inspection	Routine maintenance Inspection tech 1
6		Routine maintenance No-inspection	Routine maintenance Inspection tech 1
5–0		Routine maintenance No-inspection	Routine maintenance Inspection tech 1
Predetermined sample size = 40%			
7	Routine maintenance No-inspection	Routine maintenance Inspection tech 2	Routine maintenance Inspection tech 1
6		Routine maintenance Inspection tech 2	Routine maintenance Inspection tech 1
5–0		Routine maintenance Inspection tech 2	Rehabilitation Inspection tech 2
Predetermined sample size = 60%			
7	Routine maintenance No-inspection	Routine maintenance Inspection tech 2	Routine maintenance Inspection tech 1
6		Routine maintenance Inspection tech 2	Routine maintenance Inspection tech 2
5–0		Routine maintenance Inspection tech 2	Rehabilitation Inspection tech 2
Predetermined sample size = 80%			
7	Routine maintenance No-inspection	Routine maintenance Inspection tech 1	Routine maintenance Inspection tech 1
6		Routine maintenance Inspection tech 1	Routine maintenance Inspection tech 1
5–0		Routine maintenance Inspection tech 1	Rehabilitation Inspection tech 2
Predetermined sample size = 100%			
7	Routine maintenance No-inspection	Routine maintenance Inspection tech 1	Routine maintenance Inspection tech 1
6		Routine maintenance Inspection tech 1	Rehabilitation Inspection tech 2
5–0		Routine maintenance Inspection tech 1	Rehabilitation Inspection tech 2

uncertainty is further underestimated with a consequent exaggeration of the value of the derived information leading in this case to a more aggressive set of M&R actions in year 2 (routine maintenance, rehabilitation, and rehabilitation for assessed condition states 7, 6, and 5–0, respectively) in comparison to the policy determined by applying the LMDPSS framework consisting of a less aggressive set of M&R actions in year 2 (routine maintenance, routine maintenance, and rehabilitation for assessed condition states 7, 6, and 5–0, respectively) where the evaluation of uncertainty is accurate.

Also, notice that for the cases where rehabilitation is the optimal M&R action under the ALMDP framework for year 2, inspection technology 2 is the optimal measurement technology for year 3 (shown under the year 2 column in Table 3 as it is a decision made in year 2 to be applied in the following inspection period), whereas under the LMDPSS framework for year 2, when rehabilitation is the optimal M&R action, no-inspection is optimal for year 3. Even though rehabilitation is the optimal M&R action in all of the above cases whereby the condition will improve to the best state possible (state 7), the action is applied at the beginning of year 2 and therefore by the beginning of year 3, when inspection could be applied next, some deterioration will have occurred. In this example the overconfidence in the value of information derived from the inspection actions under the ALMDP framework results in the situation where inspection (using technology 2) is considered optimal for year 3, whereas the accurate assessment of the value of information derived from the inspection actions under the LMDPSS framework results in no-inspection for year 3 being optimal. In general, unnecessary inspection coupled with the overconfidence in its value under the ALMDP framework could lead to sub-optimal policies reflecting either unnecessarily aggressive and costly M&R actions or delayed M&R actions with negative long-term consequences in the form of accelerated deterioration and higher user costs.

In addition to the comparison between the specific optimal policies over time, the relative differences in the expected agency cost ΔAC of Eq. (29) provide an aggregate measure of comparison between the ALMDP and LMDPSS frameworks. The expected agency cost associated with the ALMDP framework varies between \$267,000 and \$289,000 for the example facility (depending on the predetermined sample size), while that associated with the LMDPSS framework is \$242,000. The relative difference ΔAC ranges between 9.23% and 21.34%. The large values are further indication of the extent of the difference between the two policies over the planning horizon. Furthermore, the positive values of ΔAC indicate that, in the case of this example, the agency will unnecessarily overspend if the ALMDP framework is used. And, the large range of the values of ΔAC indicates that the “optimal” policy determined by applying the ALMDP framework is highly dependent on the predetermined sample size. These results confirm the value of incorporating sample size as a decision variable, one of the objectives of this research.

5. Summary and future research

The uncertainty associated with spatial sampling has not been recognized or quantified in the infrastructure IM&R decision-making literature. Moreover, the sample size is not incorporated as a decision variable. This paper presents a methodology where the uncertainty in question is captured and the sample size is incorporated as a decision variable in an optimization framework.

The results based on a specific example indicate that the systematic overconfidence of a decision-maker regarding the accuracy of the inspection activity under the ALMDP framework clearly produce suboptimal policies. Moreover, the expected agency costs are very different under the LMDPSS and ALMDP frameworks, indicating that the two frameworks produce very different IM&R policies. The difference in the agency costs also illustrates that by not capturing the sampling uncertainty and by not incorporating the sample size as a decision variable, marked unnecessary overspending could take place. More comprehensive comparisons based on a wide spectrum of example facilities and scenarios are developed by Mishalani and Gong (2007, 2008) to assess both the value of LMDPSS in capturing the uncertainty associated with spatial sampling and the value in incorporating the sample size as a decision variable.

A fruitful area worthy of research is the investigation of the impact of certain factors on the IM&R decision outcomes and the life-cycle cost. Factors including inspection cost, user cost, and the nature of the spatial correlation could be considered. The results would indicate which factors are critical in influencing the decision outcomes.

In addition, considering spatial sampling in making IM&R decisions while simultaneously considering multiple facilities, usually forming a network, is another important problem to address and doing so would clearly increase the value of addressing the condition sampling problem. Such decisions focus on selecting facilities for maintenance or rehabilitation under budget constraints. Smilowitz and Madanat (2000) utilized the LMDP framework to address this problem. However, spatial sampling is not considered. It is worthwhile to do so given that inspection resources need to be allocated across multiple facilities optimally as well. Only when simultaneously considering all the facilities under the jurisdiction of an infrastructure agency, budget constraints can be taken into account.

Moreover, as discussed in detail, the methodology developed in this study builds upon the LMDP framework. It would be worthwhile to consider the treatment of spatial sampling in more recent advances in infrastructure management employing adaptive control approaches – see for example Durango and Madanat (2002) and Guillaumot et al. (2003).

Finally, while the discrete condition state representation of the Markov Decision Process based decision-making frameworks (i.e., the MDP, LMDP, and LMDPSS frameworks) is consistent with condition assessment metrics adopted by numerous infrastructure agencies, approaches that are directly based on continuous condition variables are attractive given the continuous nature of most condition variables. Recent developments in continuous condition variable based decision-making have been taking place. Example methodologies include those developed by Ouyang and Madanat (2006), Durango-Cohen (2007), and Jido et al. (2008). Given their theoretical appeal, addressing the condition spatial sampling problem in the context of such decision-making frameworks is certainly worthwhile.

Acknowledgement

The presented research was funded by the US National Science Foundation under Grant 0093452.

Appendix A. Optimum distance between adjacent condition observations

Under the assumption that the distances between adjacent observations are equal to a constant denoted by h, $h = L/(n-1)$ is the optimum solution when minimizing expected total life-cycle cost. This result is mathematically proven by first recognizing that for a given sample size n the expected total life-cycle cost increases with increasing variance of the sample mean of condition and, subsequently, by showing that that variance is a minimum at the largest possible value of h.

Based on Eq. (15), the derivative of the variance of the sample mean with respect to h is given by the following:

$$\frac{\partial \mathrm{Var}(\bar{\hat{x}})}{\partial h} = \frac{2\sigma_x^2}{n^2} \times \sum_{i=1}^{i=n-1} (n-i)\frac{\partial \rho}{\partial h} \cdot i = \left[\frac{2\sigma_x^2}{n^2} \times \sum_{i=1}^{i=n-1} (n-i) \cdot i\right] \times \frac{\partial \rho}{\partial h} = \kappa \times \frac{\partial \rho}{\partial h} \tag{A.1}$$

where κ is a positive constant value independent of the variable h. Since Mishalani and Koutsopoulos (1995, 2002) showed that the spatial correlation function of pavement condition is positive and monotonically decreasing with respect to the distance between observations, the following holds:

$$\frac{\partial \rho}{\partial h} < 0 \tag{A.2}$$

Then, Eq. (A.1) implies the following:

$$\frac{\partial \mathrm{Var}(\bar{\hat{x}})}{\partial h} < 0 \tag{A.3}$$

Therefore, the larger the h, the smaller the variance of the sample mean and, hence, the smaller the expected total life-cycle cost. Clearly $h = L/(n-1)$ is the largest possible value h can take whereby two of the n observations are located at either end of the facility. That is, $h = L/(n-1)$ is the optimum solution.

Appendix B. Information vector grouping

Madanat (1991) discretized the probability space to reduce the number of information vectors. Discretization of the probability space consists of approximating the continuous [0, 1] space of each element of the $P_t|I_t$ vector – $p_t(x_t^d = i|I_t)$ in the case of the LMDP framework (introduced in Section 2 as expression (6)) or $p_t(\mu_{x,t}^d = i|I_t)$ in the case of the LMDPSS framework (introduced in Section 3.4.1 as expression (22)) – into a set of discrete points belonging to the [0, 1] interval, which in the case of Madanat (1991) are equally spaced. In general a $1/u$ discretization, where $u = 1, 2, \ldots$, reduces the interval to a set D containing $u + 1$ points. (Other numerical methods to this problem are available, but have not been considered in this study.) Each $P_t|I_t$ vector must also satisfy the condition that the sum of its elements must be equal to 1, which reduces the number of possible vectors even further. Madanat (1993) claimed that the nearest discretized vector to any information vector can be easily generated by rounding each element of the vector to the nearest point in the discrete set. This statement can be mathematically proven to be true. This approach is referred to as the "rounding" method. However, there are two important shortcomings associated with this "rounding" method: possible non-trivial dissimilarities between information vectors assigned to the same representative information vector, and a possible ambiguity in mapping information vectors to a representative information vector. These shortcomings are illustrated through two examples.

First, consider the following two information vectors as an example: $[0.8748 \quad 0.1251 \quad 0.0001]'$ and $[0.6250 \quad 0.3749 \quad 0.0001]'$. Based on a 1/4 discretization (i.e., $u = 4$), both information vectors map to $[0.75 \quad 0.25 \quad 0]'$. That is, they are classified into the same group represented by that discretized representative vector. Notice that the Manhattan and Euclidean distances between the original two information vectors is 0.4996 and 0.3533, respectively. Even though these values are large, the two vectors are grouped together resulting in Manhattan and Euclidean distances between the original information vectors and the representative information vector of 0.2498 and 0.1766, respectively, in both cases. Of particular concern is having the probability of being in the worst condition state change from 0.0001 to 0 resulting in a possible appreciable impact on decision-making because the large user cost associated with the worst condition state (the one with zero probability based on the discretized representative vector) will not play a role in determining the optimum policies after discretization in this example. Therefore, the rounding method might result in unnecessary loss of information.

Regarding the second shortcoming, consider the following information vector as an example: $[0.750 \quad 0.125 \quad 0.125]'$. Based on a 1/4 discretization (i.e., $u = 4$), it is not clear which of the following two vectors the information vector maps to: $[0.75 \quad 0.25 \quad 0]'$ or $[0.75 \quad 0 \quad 0.25]'$. The likelihood of the occurrence of the two shortcomings decreases as u increases. This property is taken advantage of subsequently in developing an alternative grouping approach.

In order to address the above shortcomings, hierarchical clustering is introduced to complement the "rounding" method. Hierarchical clustering methods are presented in detail in Johnson and Wichern (1988). The complete linkage method is found to be commonly used and most suited for the problem at hand. The basic algorithm of this method is the following (Johnson and Wichern, 1988):

(i) Start with N clusters where N is the total number of information vectors, and calculate the $N \times N$ symmetric matrix of distances D = $\{d_{ik}\}$ between all the information vectors.
(ii) Search for the nearest pair of clusters (U,V) based on the maximum distance (or farthest neighbor) between the members of each cluster U and V. The use of this metric is known as complete linkage.
(iii) Merge clusters U and V. Update the entries in the distances matrix by deleting the rows and columns corresponding to clusters U and V, and adding a row and column indicating the distances between the centroid of cluster (UV) and the centroids of the remaining clusters.
(iv) Repeat steps (ii) and (iii) a total of $N - 1$ times.

By applying the complete linkage method, the maximum distance among information vectors in each group is determined. Such a distance is then used in applying the grouping criterion whereby the final grouping consists of the smallest number of clusters such that the maximum distance among information vectors in each cluster does not exceed an upper bound. The larger the upper bound to this maximum distance is, the more information vectors are combined. As a result, the less memory and computation time are required at the expense of loss in accuracy.

The complete linkage method does not capture the anisotropy of information vectors, while the "rounding" method does. That is, the different elements of an information vector are associated with different cost values in the optimization process, which gives the rounding method an advantage when u is large in that each element in the original information vector will be close to the corresponding element of the representative information vector. Therefore, in order to take advantage of the desirable characteristics of both methods in this study, information vectors are grouped by the "rounding" method first using a large value for u, and then the resulting representative information vectors are further grouped using the complete linkage clustering method.

References

Ben-Akiva, M., Gopinath, D., 1995. Modeling infrastructure performance and user costs. Journal of Infrastructure Systems 1 (1), 33–43.
Ben-Akiva, M., Humplick, F., Madanat, S., Ramaswamy, R., 1993. Decision-making under uncertainty in infrastructure management: the latent performance approach. Journal of Transportation Engineering 119 (1), 43–58.
Brockwell, P.J., Davis, R.A., 2002. Introduction to Time Series and Forecasting. Springer, New York.
Carnahan, J.V., Davis, E.J., Shahin, M.Y., Keane, P.L., Wu, M.I., 1987. Optimal maintenance decisions for pavement management. Journal of Transportation Engineering 113 (5), 554–572.
Casella, G., Berge, R., 2002. Statistical Inference. Duxbury, Pacific Grove, CA.
Cressie, N., 1993. Statistics for Spatial Data. Wiley, New York.
Durango, P.L., Madanat, S.M., 2002. Optimal maintenance and repair policies in infrastructure management under uncertain facility deterioration rates: an adaptive control approach. Transportation Research Part A 36 (9), 763–778.
Durango-Cohen, P.L., 2007. A time series analysis framework for transportation infrastructure management. Transportation Research Part B 41 (5), 493–505.
Ellis, H., Jiang, M., Corotis, R.B., 1995. Inspection, maintenance, and repair with partial observability. Journal of Infrastructure Systems 1 (2), 92–99.
Feighan, K.J., Shahin, M.Y., Sinha, K.C., White, T.D., 1988. Application of dynamic programming and other mathematical techniques to pavement management systems. Transportation Research Record 1200, 90–98.
Gong, L., 2006. Optimal Spatial Sampling of Infrastructure Condition: a Life-cycle-based Approach under Uncertainty. Ph.D. Dissertation, Department of Civil and Environmental Engineering and Geodetic Science, The Ohio State University, Columbus, OH.
Guillaumot, V.M., Durango-Cohen, P.L., Madanat, S.M., 2003. Adaptive optimization of infrastructure maintenance and inspection decisions under performance model uncertainty. Journal of Infrastructure Systems 9 (4), 133–139.
Humplick, F., 1992. Highway pavement distress evaluation: modeling measurement error. Transportation Research Part B 26 (2), 135–154.
InflationData.com, 2005. URL: <http://inflationdata.com/Inflation/Inflation_Rate/InflationCalculator.asp>. Richmond, VA.
Jido, M., Otazawa, T., Kobayashi, K., 2008. Optimal repair and inspection rules under uncertainty. Journal of Infrastructure Systems 14 (2), 150–158.
Johnson, A.R., Wichern, D.W., 1988. Applied Multivariate Statistical Analysis. Prentice-Hall, Englewood Cliffs, NJ.
Madanat, S.M., 1991. Optimizing Sequential Decisions under Measurement and Forecasting Uncertainty: application to Infrastructure Inspection, Maintenance and Rehabilitation. Ph.D. Dissertation, Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, MA.
Madanat, S.M., 1993. Optimal infrastructure management decisions under uncertainty. Transportation Research Part C 1 (1), 77–88.
Madanat, S.M., Ben-Akiva, M., 1994. Optimal inspection and repair policies for infrastructure facilities. Transportation Science 28 (1), 55–62.
Mishalani, R.G., 1993. Extracting spatial information on infrastructure distress using a stochastic modeling approach: application to highway maintenance decision making. Ph.D. dissertation, Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, MA.
Mishalani, R.G., Gong, L., 2007. Spatial sampling of infrastructure condition: motivation and numerical analysis. In: Proceedings of the 11th World Conference on Transportation Research, Berkeley, CA.
Mishalani, R.G., Gong, L., 2008. Evaluating the impact of pavement condition sampling advances on life-cycle management. Transportation Research Record 2068.
Mishalani, R.G., Koutsopoulos, H.N., 1995. Uniform infrastructure fields: definition and identification. Journal of Infrastructure Systems 1 (1), 44–55.
Mishalani, R.G., Koutsopoulos, H.N., 2002. Modeling the spatial behavior of infrastructure condition. Transportation Research Part B 36 (2), 171–194.
Ouyang, Y., Madanat, S.M., 2006. An analytical solution for the finite-horizon pavement resurfacing planning problem. Transportation Research Part B 40 (9), 767–778.
Shahin, M., Kohn, S., 1981. Pavement management system for roads and parking lots. Technical Report M-294, Construction Engineering Research Laboratory (CERL), US Army Corps of Engineers, Champaign, IL.
Sharaf, E.A., Reichelt, E., Shahin, M.Y., Sinha, K.C., 1985. Development of a methodology to estimate pavement maintenance and repair costs for different range of pavement condition index. Transportation Research Record 1123, 30–39.
Smilowitz, K., Madanat, S., 2000. Optimal inspection and maintenance policies for infrastructure networks. Computer-Aided Civil and Infrastructure Engineering 15 (1), 5–13.
Vadakpat, G., Stoffels, S., Dixon, K., 2000. Road user cost models for network-level pavement management. Transportation Research Record 1699, 49–57.