Bayesian Updating of Infrastructure Deterioration Models

YUN LU AND SAMER MADANAT

Deterioration forecasting plays an important role in the infrastructure management process. The precision of facility condition forecasting directly influences the quality of maintenance and rehabilitation decision making. One way to improve the precision of forecasting is by successive updating of deterioration model parameters. A Bayesian approach that uses inspection data for updating facility deterioration models is presented. As an empirical study with bridge deck data indicates, the use of this methodology significantly reduces the uncertainty inherent in condition forecasts.

The process of infrastructure management consists of three activities: data collection and inspection, deterioration modeling and forecasting, and maintenance and rehabilitation (M&R) decision making (*1*). After facility condition data are collected, deterioration models are developed by using these data to forecast future facility performance; both data and models are later used to support M&R decision making. One fact that should be emphasized is that the relationship between condition data and deterioration models has traditionally been a static one; that is, once a deterioration model is developed, subsequently collected data are not used for model updating.

In contrast to this approach this paper presents a method that exploits the condition data collected during facility inspections to improve the precision of deterioration models. In this method data are used not only for deterioration modeling but also for model updating.

The advantage of this method is that one can start to develop a model even with limited data. Later the model can be updated as additional data become available. Therefore this approach increases the precision of forecasting and is expected to decrease facility life cycle costs.

The updating method used in this paper is the Bayesian approach, which will be introduced in the next section. PONTIS (*2*), the California Department of Transportation–FHWA network optimization system for bridge improvement and maintenance, used the same methodology to update the transition probabilities representing bridge deterioration. The only difference between the work presented in this paper and the updating procedure used in PONTIS is that the present research deals with a continuous deterioration model, whereas PONTIS is based on a discrete model that uses Markov transition probabilities as model parameters.

BAYESIAN APPROACH

Bayesian analysis is defined as the approach to statistics that formally seeks to use prior information. In statistics the Bayesian approach is widely used to estimate an unknown parameter Θ. Bayesian analysis is performed by combining the prior information and the sample information (x) into what is called the posterior distribution of Θ given x, from which all decisions and inferences are made (*3*).

Let $\pi(\Theta)$ denote the prior distribution of Θ and let $L(x, \Theta)$ denote the likelihood function, then Bayes's theorem can be expressed as follows:

$$\pi(\Theta|x) = \frac{L(x, \Theta)\pi(\Theta)}{\int_{-\infty}^{\infty} L(x, \Theta)\pi(\Theta)d(\Theta)} \quad (1)$$

where

$L(x, \Theta) = f(x|\Theta)$ = likelihood of experimental outcome x, that is, conditional probability of obtaining a particular experimental outcome assuming that the parameter is Θ;

$\pi(\Theta)$ = prior probability of Θ, that is, before availability of experimental information;

$\pi(\Theta|x)$ = posterior probability of Θ, that is, probability that has been revised in the light of experimental outcome x.

It is observed from Equation 1 that both the prior distribution and the likelihood function contribute to the posterior distribution of Θ. The prior information enters the posterior probability density function (pdf) via the prior pdf, whereas all of the sample information enters via the likelihood function. In this manner judgmental and observational data are combined properly and systematically (*4*).

Likelihood Function

For observed data, x, the function $L(x, \Theta) = f(x|\Theta)$ considered as a function of Θ is called the *likelihood function*.

Given a set of observed values $x_1, x_2, \ldots, x_n$, which represent a random sample from a population of X with underlying density $f_x(X)$, the probability of observing this particular set of values, assuming that the parameter of the distribution is Θ, is

$$f(x|\Theta) = L(x, \Theta) = \prod_{i=1}^{n} f_x(x_i|\Theta) \quad (2)$$

From Equation 2 it can be observed that the likelihood function $L(x, \Theta)$ is the product of the density function of X evaluated at $x_1, x_2, \ldots, x_n$.

School of Civil Engineering, Purdue University, West Lafayette, Ind. 47907.

Prior Information

Generally the prior information comes from past experience, the nature of the problem, or previous work. A prior distribution for which $\pi(\Theta)$ can be easily calculated is the so-called conjugate prior. For example the class of normal priors is a conjugate family for the class of normal (sample) densities. That is if X has a normal density and Θ has a normal prior, then the posterior density of Θ given x is also normal.

A conjugate prior greatly simplifies the application of Bayes's theorem for determination of a posterior distribution. Conjugate distributions provide a convenient model that may be realistic in many situations.

Posterior Distribution

The posterior distribution is the combination of the prior information and the likelihood function. Just as the prior distribution reflects beliefs about Θ prior to experimentation, so $\pi(\Theta|x)$ reflects the updated beliefs about Θ after (posterior to) observing the sample x. In other words the posterior distribution combines the prior beliefs about Θ with the information about Θ contained in the sample x to give a composite picture of the final beliefs about Θ.

The two important quantities of the posterior distribution are the mean and the variance. The mean value of Θ that is used as the Bayesian estimator of the parameter is

$$E^{\pi(\Theta|x)}(\Theta) = \int_{-\infty}^{\infty} \Theta\pi(\Theta|x)d\Theta \qquad (3)$$

and the variance is given by

$$V^{\pi(\Theta|x)}(\Theta) = \int_{-\infty}^{\infty} \Theta^2\pi(\Theta|x)d\Theta - [E^{\pi(\Theta|x)}(\Theta)]^2 \qquad (4)$$

In the case of a conjugate distribution, once the mean and the variance of the posterior distribution are calculated, one can directly write its probability density function.

The Bayesian approach has many advantages in the area of engineering planning and design. It systematically combines uncertainties associated with randomness and those arising from error of estimation and prediction. It provides a formal procedure for systematic updating of information and increases the prediction precision.

APPLICATION OF UPDATING METHODOLOGY

In this section the Bayesian approach is applied to the problem of updating facility deterioration models. A logistic model representing the fraction of bridge deck area delaminated is used as an example.

The logistic model has an S shape; its attractive mathematical property is that it has a bounded function that lies between 0 and 1; it is therefore well-suited for representing the progression of the damaged fraction of a bridge deck. Figure 1 shows an application of the logistic model that represents the percentage of a bridge deck area damaged (P_t) from year 0 to year 40 during which no maintenance or rehabilitation is performed. It can be observed from this graph that between years 0 and 10 there is very little damage since the bridge is new; after year 10 the rate of deterioration increases rapidly and P_t reaches 0.94 at year 25. After that the deterioration tends to slow again. This trend is consistent with observations of bridge deck deterioration over time (*5,6*).

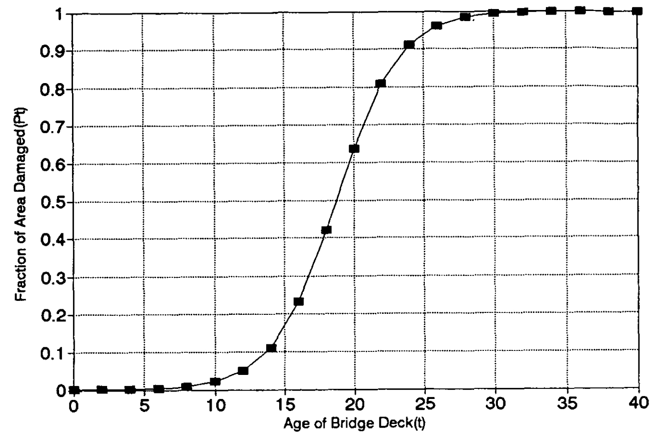

FIGURE 1 Fraction of bridge deck area damaged.

The mathematical function of the logistic model is of the form

$$P_t = \frac{1}{1 + e^{a+bt+\epsilon}} \qquad (5)$$

where

- P_t = fraction of area of bridge deck damaged;
- t = age of bridge deck;
- a, b = parameters specific to each bridge deck type; and
- ϵ = a random error term that captures the uncertainty associated with the deterioration process; it is usually assumed to be normally distributed, with mean μ_ϵ equal to 0 and variance σ_ϵ^2 (the variance σ_ϵ^2 is bridge deck type specific).

In Equation 5 there are two parameters, a and b, that determine the rate of deterioration. Generally if sufficient observations of bridge decks of a given type consisting of t and P_t are available, these two parameters can be estimated statistically.

In reality no deterioration model is perfectly accurate because of the limited sample size, the inherent randomness of the process, observation errors, and so on. The Bayesian approach can be used to update the deterioration model parameters to increase the precision of forecasting. For mathematical simplicity this paper considers updating parameter b only while treating parameter a as constant. Extending the method to update both parameters simultaneously is conceptually straightforward, but somewhat mathematically cumbersome.

Derivation of Prior Distribution and Likelihood Function

To use Bayes's theorem one needs to find out the prior distribution of b and the likelihood function $L(P_t, b)$.

The prior distribution of b is easy to determine. If regression is used to estimate the parameters of Equation 5, then the parameter b usually can be assumed to be normally distributed with mean μ_b and variance σ_b^2 (7). Therefore the prior density function of b can be written in the form

$$\pi(b) = \frac{e^{-\frac{1}{2}\frac{(b-\mu_b)^2}{\sigma_b^2}}}{\sqrt{2\pi}\,\sigma_b} \qquad (6)$$

To obtain the likelihood function the density function of P_t must first be generated.

In general if δ is normally distributed with mean a and variance σ^2 and δ is log τ, then τ will be log-normally distributed (8). In Equation 5 let $x = e^{a+bt+\epsilon}$ and $\Theta = a + bt + \epsilon$, then $x = e^{\Theta}$. Since $\Theta = a + bt + \epsilon$ is normally distributed and $\Theta = \log x$, x is therefore log-normally distributed according to the above rule. Its density function is of the form

$$f(x) = \frac{e^{\frac{-(\log x - a - bt)^2}{2\sigma_\epsilon^2}}}{\sqrt{2\pi}\,x\sigma_\epsilon} \qquad (7)$$

Generally, if the pdf of x is known as $f(x)$ and $y = h(x)$, one can obtain the pdf of y by the following relation:

$$f(y) = \left|\frac{\partial x}{\partial y}\right| f(x) \qquad (8)$$

Therefore the density function of P_t is given by

$$f(P_t) = \left|\frac{\partial x}{\partial P_t}\right| f(x) = \frac{e^{\frac{[\log(\frac{1}{P_t}-1)-a-bt]^2}{-2\sigma_\epsilon^2}}}{\sqrt{2\pi}\,\sigma_\epsilon P_t(1-P_t)} \qquad (9)$$

The likelihood function is the product of the pdf of P_t over all observations:

$$L(P_t, b) = \prod_{i=1}^{k} f(P_{t,i}|b) = \frac{e^{\sum_{i=1}^{k}\frac{[\log(\frac{1}{P_{t,i}}-1)-a-bt]^2}{-2\sigma_\epsilon^2}}}{\prod_{i=1}^{k}\sqrt{2\pi}\,\sigma_\epsilon P_{t,i}(1-P_{t,i})} \qquad (10)$$

Derivation of Posterior Distribution

According to Equation 1 the posterior distribution of b is

$$\pi(b|P_t) = \frac{L(P_t, b)\pi(b)}{\int_{-\infty}^{\infty} L(P_t, b)\pi(b)d(b)} \qquad (11)$$

Substituting $L(P_t, b)$ and $\pi(b)$ with Equations 10 and 6, respectively, one obtains

$$\pi(b|P_t) \propto \frac{e^{\sum_{i=1}^{k}\frac{\left[\log(\frac{1}{P_{t,i}}-1)-a-bt\right]^2}{-2\sigma_\epsilon^2}}}{\prod_{i=1}^{k}\sqrt{2\pi}\,\sigma_\epsilon P_{t,i}(1-P_{t,i})} \cdot \frac{e^{\frac{(b-\mu_b)^2}{-2\sigma_b^2}}}{\sqrt{2\pi}\,\sigma_b} \qquad (12)$$

It is tedious to use this equation. Fortunately the prior pdf is a conjugate normal distribution. Therefore the posterior distribution is also a normal distribution, with the mean and the variance shown in Equations 13 and 14, respectively (the derivation of these two equations is omitted for simplicity; it can be shown that Equation 12 simplifies to a normal distribution).

$$\mu' = \frac{\sigma_b^2 t\left\{E\left[\log\left(\frac{1}{P_{t,i}}-1\right)\right]-a\right\}+\frac{\mu_b\sigma_\epsilon^2}{k}}{\frac{\sigma_\epsilon^2}{k}+t^2\sigma_b^2} \qquad (13)$$

$$\sigma'^2 = \frac{\sigma_\epsilon^2\sigma_b^2}{\sigma_\epsilon^2 + kt^2\sigma_b^2} \qquad (14)$$

Therefore the pdf of the posterior distribution is of the form

$$\pi(b|P_t) = \frac{e^{\frac{(b-\mu')^2}{-2\sigma'^2}}}{\sqrt{2\pi}\,\sigma'} \qquad (15)$$

PARAMETRIC ANALYSIS

In this section a parametric analysis is performed to evaluate the effect of performing Bayesian updating on the forecasting precision of a logistic model of bridge deck deterioration. The deterioration model studied in this section has the following form (9):

$$P_t = \frac{1}{1 + e^{a+bt+\epsilon}} = \frac{1}{1 + e^{8.72-0.44t+\epsilon}} \qquad (16)$$

Bayes's theorem was used to update the parameter b to increase the model's prediction precision. Hence the parameter b is treated as a random variable instead of a constant. Model estimation results indicated that the prior distribution of b is normal, with mean μ_b equal to -0.44 and standard deviation σ_b equal to 0.2 (9). By calculating the posterior mean and variance of b through Equations 13 and 14, the parameter b can be updated repeatedly, thus reducing the standard deviation of b. Therefore the standard deviation of P_t is also expected to decrease. Monte Carlo simulation was used in the study to compute the standard deviation of P_t after each model update. The use of Monte Carlo simulation was necessitated by the form of the deterioration model. As Equation 16 shows the relationship between the parameter b and P_t is strongly nonlinear, which makes the derivation of the variance of P_t as a function of σ_b analytically rather difficult.

Two cases are compared: one is without updating and the other is with updating. In the first case parameter b is normally distributed with mean μ_b equal to -0.44 and standard deviation σ_b equal to 0.2. Since no updating is performed in this case the mean and standard deviation of b will remain constant from year 2 to year 20, which is the horizon used in the analysis. In the second case Bayesian updating is performed every 2 years after each inspection cycle. The number of observations k collected in every inspection period in the parametric study is assumed to be 10. Each observation consists of a pair (t, P_t) pertaining to a particular bridge deck among the population of bridges. In this case the parameter b is still normally distributed, but the mean and standard deviation are updated every 2 years according to Equations 13 and 14, respectively. Figure 2 depicts the standard deviation of b under these two scenarios. It can be observed from Figure 2 that the standard deviation of b decreases significantly as a result of updating, especially during the first update.

To show that the prediction accuracy is increased by updating the parameter b, the standard deviation of P_t needs to be compared under these two cases. Monte Carlo simulation was used to calculate the standard deviation of P_t after each update. The results are summarized in Figure 3, which shows that the standard deviation of P_t becomes substantially smaller when the parameter b is updated. The shapes of the two curves in Figure 3 are instructive. The upper curve, corresponding to the nonupdating case, shows the standard deviation of the forecast of P_t increasing up to year 20, after which it decreases until it becomes 0 at year 40 (not shown in the figure). This behavior stems from the fact that the P_t function is bounded from below by 0 and from above by 1, which forces the standard deviation to 0 at these two extremes. The lower curve, which corresponds to the updating case, shows a similar behavior, except that the maximum forecast standard error occurs earlier because of the contribution of Bayesian updating to reducing the standard deviation of b.

A study of the change in infrastructure life cycle costs with the level of uncertainty in condition forecasting can be found in another study (*10*). Figure 4 is adapted from that study. In Figure 4 the x-axis shows the standard error of conditional forecasting measured in PCI (Pavement Condition Index) units, a measure of pavement performance. The y-axis shows the expected life cycle cost, which is the sum of agency costs and user costs over the planning horizon, associated with the optimal policies. The M&R decision model used to compute these minimum life cycle costs is a Markov decision process–based optimization model. Under this decision model the inspection frequency is predetermined once per time period. Figure 4 shows that increasing the uncertainty in forecasting of the condition leads to a substantial increase in the value of the minimum expected cost. This result demonstrates the economic benefits of improving the precision of the deterioration models used in pavement management. Although the application in the present study (bridge decks) is different from the one used in that study, a similar result should hold because of the similarities in the cost structures of the two problems.

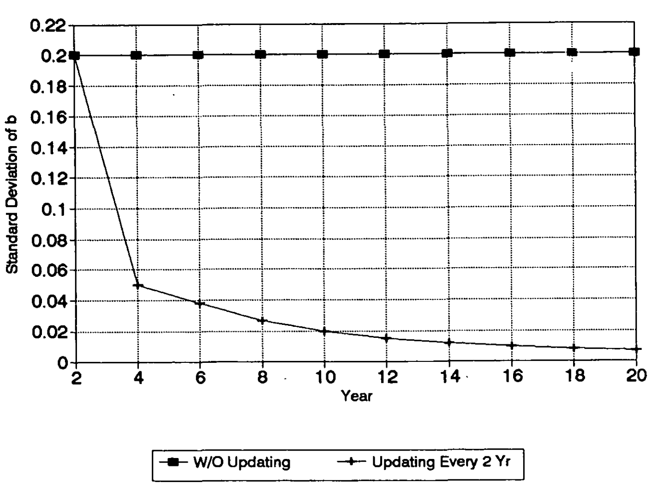

FIGURE 2 Standard deviation of b for updating and nonupdating cases.

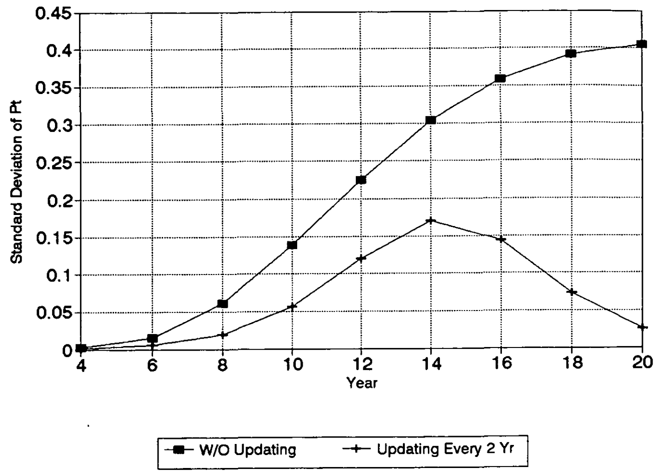

FIGURE 3 Standard deviation of P_t for updating and nonupdating cases.

CONCLUSION

In this paper a Bayesian methodology for updating deterioration models in infrastructure management was presented. This method has the following advantages:

1. It significantly decreases the uncertainty inherent in the forecasting of facility condition, thus decreasing the expected facility life cycle cost, and
2. It allows the development of a deterioration model even with limited data; the model is repeatedly updated and improved as inspection data are incorporated.

A deterioration model representing the damaged fraction of a bridge deck area was chosen as an application example of this methodology. As the empirical study indicated, the use of the methodology presented in this paper for updating the deterioration model parameter reduces the uncertainty associated with forecasting bridge deck condition significantly. Therefore the facility life cycle cost can be expected to decrease.

Although the sample problem dealt with a deterioration model for bridge decks, this methodology is also applicable to highway pavements as long as an appropriate deterioration model can be developed.

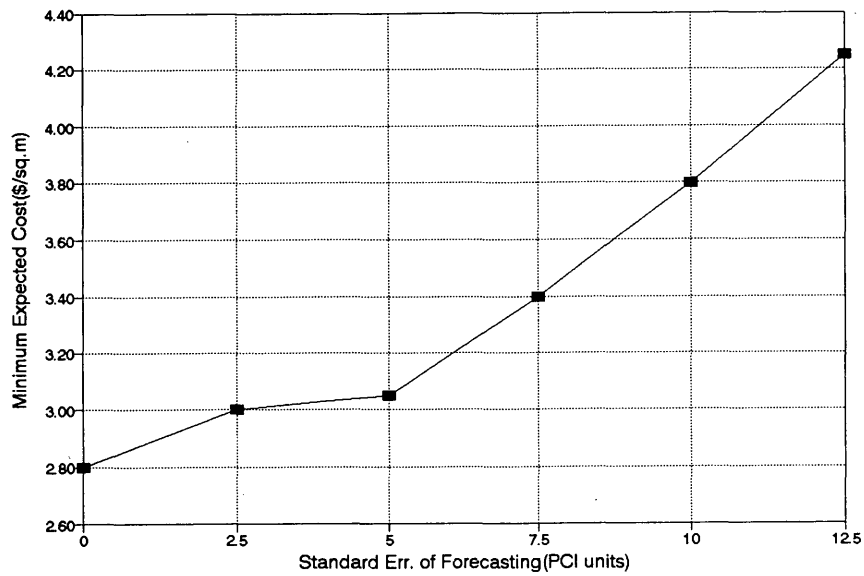

FIGURE 4 Minimum expected life cycle cost versus standard error of forecast.

To make this methodology operational and simplify the implementation, the following considerations should be taken into account:

1. The likelihood function plays an important role in Bayes's theorem. Therefore the choice of functional form for the deterioration model is a critical issue in the implementation of this methodology. To simplify the application of Bayesian updating, deterioration models whose forecasts follow a commonly known distribution such as a normal or a log-normal distribution are recommended.
2. Although many choices for the prior distribution are often available, a conjugate prior distribution is recommended because it facilitates the implementation of this methodology.

In this paper only parameter b was updated. Parameter a was treated as a constant for the sake of mathematical simplicity. The empirical study indicated that through updating parameter b the forecasting precision increased significantly. It is expected that the result will be even more significant if parameters a and b are updated simultaneously. The basic updating procedure will be the same as the one presented in this paper.

Generally deterioration models are classified into two categories: disaggregate and aggregate. PAVER (*11*) is an example of an aggregate model. The deterioration model described in this paper is a disaggregate model that uses an individual damage measurement as a measure of performance. Since an aggregate deterioration model represents a combination of different types of damage, updating of such model may be more difficult. The appropriate way to update such a model may be to update each damage model first and then to combine the results properly if individual damage models exist. On the other hand if the deterioration model used consists of a single condition index it will be necessary to perform Bayesian updating on that aggregate model directly. The application of Bayes' theorem for this case needs further study.

ACKNOWLEDGMENT

The authors are grateful to Minghui Chen of Purdue University for clarification of some statistical issues in Bayesian methodology.

REFERENCES

1. Ben-Akiva, M., F. Humplick, S. Madanat, and R. Ramaswamy. Infrastructure Management Under Uncertainty: Latent Performance Approach. *Journal of Transportation Engineering*, ASCE, Vol. 119, No. 1, 1993.
2. *PONTIS, Network Optimization System for Bridge Improvements & Maintenance.* Interim report. Cambridge Systematics, Inc., 1991.
3. Berger, J. O. *Statistical Decision Theory & Bayesian Analysis.* Springer-Verlag Inc., New York, 1985.
4. Benjamin, J. R., and C. A. Cornell. *Probability, Statistics & Decision for Civil Engineers.* McGraw-Hill Book Company, Inc., New York, 1970.
5. Cady, P. D., and R. E. Weyers. Deterioration Rates of Concrete Bridge Decks. *Journal of Transportation Engineering*, ASCE, Vol. 110, No. 1, 1984.
6. Markow, M., S. Madanat, and D. Gurenich. Optimal Rehabilitation Times for Concrete Bridge Decks. In *Transportation Research Record 1392*, TRB, National Research Council, Washington, D.C., 1993, pp. 79–89.
7. Neter, J., W. Wasserman, and M. H. Kotner. *Applied Linear Statistical Models.* Richard D. Irwin, Inc., 1990.
8. Wang, Z. X. *Probabilities.* People's Publishing Inc., Beijing, People's Republic of China, 1970.
9. Lu, Y. *A Methodology for Updating Deterioration Models in Infrastructure Management.* M. S. thesis. Department of Civil Engineering, Purdue University, West Lafayette, Ind., 1993.
10. Madanat, S. Incorporating Inspection Decisions in Pavement Management. *Transportation Research*, Vol. 27B, 1993, pp. 425–438.
11. Shahin, M., and S. Kohn. *Pavement Management for Roads and Parking Lots.* Technical Report M-294. Construction Engineering Research Laboratory, U.S. Army Corps of Engineers, 1980.

Publication of this paper sponsored by Committee on Structures Maintenance and Management.