Journal of Applied Statistics
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713428038

A Bayesian approach for improved pavement performance prediction

Eun Sug Park[a]; Roger E. Smith[ab]; Thomas J. Freeman[a]; Clifford H. Spiegelman[ac]
[a] Texas Transportation Institute, Texas A&M University System, College Station, TX, USA [b] Department of Civil Engineering, Texas A&M University, College Station, TX, USA [c] Department of Statistics, Texas A&M University, College Station, TX, USA

To cite this Article Park, Eun Sug , Smith, Roger E. , Freeman, Thomas J. and Spiegelman, Clifford H.(2008) 'A Bayesian approach for improved pavement performance prediction', Journal of Applied Statistics, 35: 11, 1219 — 1238
To link to this Article: DOI: 10.1080/02664760802318651
URL: http://dx.doi.org/10.1080/02664760802318651

Journal of Applied Statistics
Vol. 35, No. 11, November 2008, 1219–1238

Routledge
Taylor & Francis Group

A Bayesian approach for improved pavement performance prediction

Eun Sug Park[a]*, Roger E. Smith[a,b], Thomas J. Freeman[a] and Clifford H. Spiegelman[a,c]

[a]*Texas Transportation Institute, Texas A&M University System, College Station, TX, USA;* [b]*Department of Civil Engineering, Texas A&M University, College Station, TX, USA;* [c]*Department of Statistics, Texas A&M University, College Station, TX, USA*

We present a method for predicting future pavement distresses such as longitudinal cracking. These predicted distress values are used to plan road repairs. Large inherent variability in measured cracking and an extremely small number of observations are the nature of the pavement cracking data, which calls for a parametric Bayesian approach. We model theoretical pavement distress with a sigmoidal equation with coefficients based on prior engineering knowledge. We show that a Bayesian formulation akin to Kalman filtering gives sensible predictions and provides defendable uncertainty statements for predictions. The method is demonstrated on data collected by the Texas Transportation Institute at several sites in Texas. The predictions behave in a reasonable and statistically valid manner.

Keywords: pavement management information system; Bayesian adjustment; state-space models; Kalman filtering; Markov chain Monte Carlo

1. Introduction

Once a new roadway is built, various types of pavement distresses may develop over time. It is very important to monitor and predict the pavement distress for efficient maintenance of roadways. The US federal government has a transportation budget of nearly 60 billion dollars a year, and state budgets provide additional billions of dollars for transportation. Major portions of these budgets are spent on road and highway repair each year. Factors that state and federal highway departments consider before deciding on repair schedules for road and highway segments include: pavement type, road or highway condition and forecasted condition, ride score, average daily traffic, road or highway class, weather, and time since the last resurfacing. Among these, forecasted road or highway condition is an important statistical component.

Figure 1 shows an example of the predominant pavement distress found in Texas, USA, called longitudinal cracking. Longitudinal cracking consists of cracks or breaks that run approximately parallel to the pavement centerline. It is measured as the total length in linear feet per road segment. The current distress data collection method in Texas is to have the rating team drive along the

*Corresponding author. Email: e-park@tamu.edu

ISSN 0266-4763 print/ISSN 1360-0532 online

DOI: 10.1080/02664760802318651
http://www.informaworld.com

shoulder, or edge of the pavement, at <15 mph, and count the quantities of the different types of distresses such as failures, alligator cracking, longitudinal cracking, transverse cracking, raveling, and flushing in each section [8]. The sections are defined by Texas Department of Transportation's (TxDOT's) pavement management information system (PMIS) and are usually 0.5 mile in length. The lane with the most distresses is rated, which is usually the outside lane. One hundred percent of the sections are inspected every year.

There are usually large variations in observed cracking within a section of pavement over the years, not just due to measurement errors but also due to the inherent variability of cracking.

Figure 2 illustrates the longitudinal cracking levels measured at different times from one pavement section in Texas. The x-axis represents the pavement age in years, and the y-axis represents the level of longitudinal cracking in linear feet. These are converted to units of cracking per 100 feet of road segment. The circles represent the observations. As can be seen, there are not many observations, which is typical of pavement data. Pavement distress data for a section often consist of only a few observations (often <10) because the inspection is usually taken only once a year due to high cost, and since pavement rehabilitation eliminates all distress. The cost of annual inspection in Texas is $2.1 million. Because of the small number of observations, it is especially important to incorporate good engineering knowledge in pavement distress prediction. Thirty-six of 39 state agencies (including the District of Columbia) responding to a recent survey said that they used distress data [1]. Inspection costs of other agencies can also be found from the Federal Highway Administration's document on states data collection [1].

In the absence of measurement error, engineers frequently use the following sigmoidal equation to model pavement distress [5]:

$$L_t = \gamma + \alpha \mathrm{e}^{-(\chi\rho/(\mathrm{AGE}_t - \delta))^{\beta}}, \tag{1}$$

Figure 1. Longitudinal cracking.

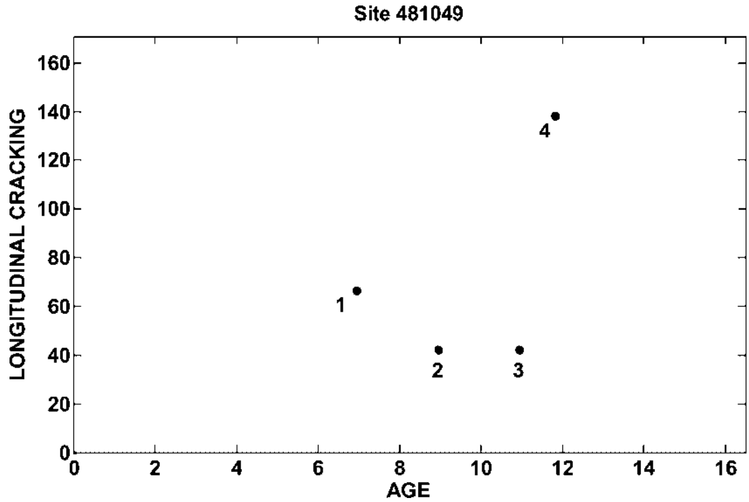

Figure 2. Longitudinal cracking data example.

where L_t is the level of distress at time AGE_t (i.e. for the tth observation), α the horizontal asymptote factor that controls the maximum range of percentage distress growth, δ the horizontal shift factor controlling the age at which the first sign of distress appears, γ the vertical shift factor giving the initial level of the distress, ρ the prolongation factor that controls how long the pavement will last before significant increases in distress occur, χ the modifying factor for ρ that accounts for a change in projection based on the observed performance, β the slope factor that controls how steeply L_t changes in the middle of the curve, and AGE_t is the pavement section age in years for the tth observation.

The values for γ and δ are usually determined by engineering knowledge of the distress definitions and the initial values. For example, a pavement will have most initial distresses set to zero, while for the international roughness index in inches/mile, the usual initial value is 65. For some distresses, such as rutting, distress begins immediately, but alligator cracking has a much longer time delay (horizontal shift factor). Depending on the distress being modeled and how it is modeled, shift factors that reflect the appropriate performance are utilized. As a matter of fact, it is often reasonable to assume γ and δ to be zeros (or other known values) in practice. If γ and δ are known to be 0, Equation (1) can be rewritten as

$$L_t = \alpha \mathrm{e}^{-(\chi\rho/\mathrm{AGE}_t)^{\beta}}. \tag{2}$$

In Section 2, we assume Equation (2) represents median distress level per road segment.

Figure 3 is a plot of a sigmoidal curve described by the above equation. Sets of values for the α, ρ, and β for predicting the level of individual distress types such as longitudinal cracking have been developed by the TxDOT for groups of similar pavements [6]. The sigmoidal curve having

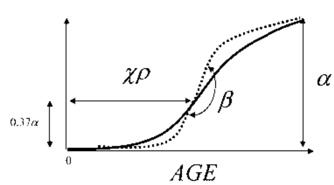

Figure 3. Theoretical distress (longitudinal cracking).

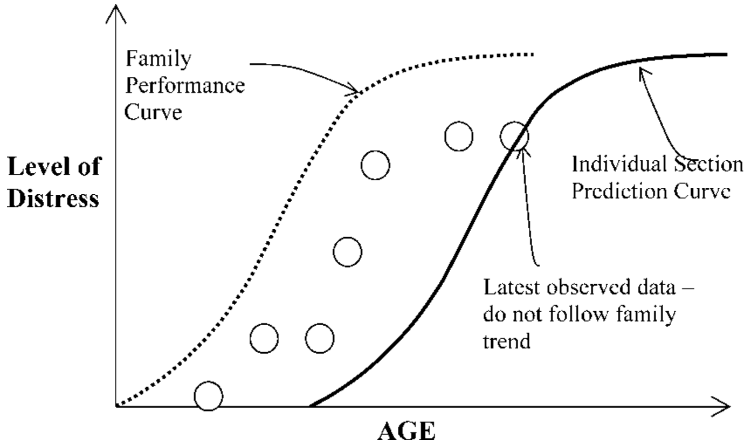

Figure 4. TxDOT method 1: individual pavement section prediction curve from family model.

those α, ρ, and β values and 1 for χ is called a family curve in the TxDOT PMIS. The family curve is assumed to give the average projected condition for distress for those similar pavement sections. However, individual pavement sections will perform quite differently from the average. In addition, there is some level of error in the observed quantities of distress measured in the field for each section of pavement. In the current study, α, β, ρ, γ, and δ are taken to be known constants, and χ is an unknown parameter that varies with a pavement section. Hereafter, χ will be referred to as the projection parameter. As mentioned earlier, all pavement sections are inspected each year (100% survey). The objective of this research is to develop a procedure for updating the expected pavement performance of individual sections as a new observation on distress (resulting from yearly inspection of the same portion of pavement sections) becomes available.

This process can then be used as a part of the TxDOT PMIS.

The approach currently used by TxDOT and many other state and local agencies (referred to as TxDOT method 1) is to obtain the prediction curve by shifting the family curve horizontally each time a new observation is entered, so that the prediction curve passes through the latest observation. This is illustrated in Figure 4. The other approach by TxDOT planned for the PMIS (referred to as TxDOT Method 2) is to 'bend' the family curve as a new observation comes in so that the prediction curve passes through the latest observation, as illustrated in Figure 5. Both these naive approaches seem to be problematic. Each prediction curve is based only on the last observation, and does not utilize any information in the past observations. Also, there is no consideration given to errors in the observed distress. As a result, if the observed condition varies to both the right and the left of the family curve over time, the future projected condition will vary considerably after each inspection.

Our goal here is to develop a practically improved, as well as statistically valid, method to predict future pavement conditions, based on Equation (2) with the unknown projection parameter χ and the observed distress data, while reducing the dramatic variability of prediction by the TxDOT methods. The value of the projection parameter χ can be assumed to be 1 initially. This is our prior belief about the projection parameter before we observe any data. Thus, the family curve will serve as a prediction curve when no data are available. After data are obtained, our initial belief on the projection parameter (χ) as well as the prediction curve can be modified to incorporate the data. As a matter of fact, with each new observation the prediction can be updated. A Bayesian framework is a natural fit to this problem. The rest of the article is organized as follows. In Section 2, the underlying physical model is extended to account for excessive variability in distress data. Section 3 contains the estimation of parameters and the prediction of future pavement distress.

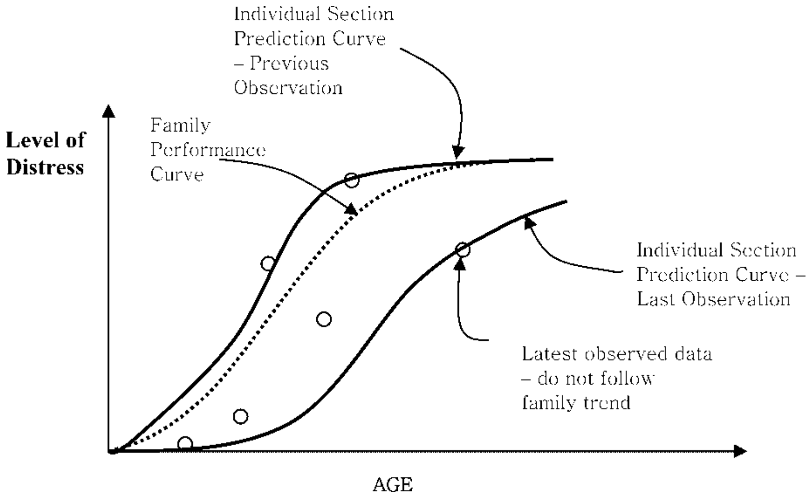

Figure 5. TxDOT method 2: potential problems with individual pavement section prediction with known date of construction.

Section 4 applies the proposed method to pavement distress data from several sites in Texas, and Section 5 concludes.

2. Model

When there are observational errors in the model, Equation (2) may represent the median distress level, rather than the observed distress level, and needs to be written as

$$\mu_t = \alpha \mathrm{e}^{-(\chi\rho/\mathrm{AGE}_t)^\beta}, \tag{3}$$

where μ_t is the median level of distress at time AGE_t. Now, we need a model for the random variation of L_t about μ_t. We adopt a model

$$L_t = \alpha \mathrm{e}^{-(\chi\rho/\mathrm{AGE}_t)^\beta \mathrm{e}^{\varepsilon_t}}, \tag{4}$$

where $\varepsilon_t \sim N\,(0, \Sigma)$, for the observational errors.

As mentioned in Section 1, the individual measurements of pavement distress such as longitudinal cracking can vary considerably over time not only becuase of real measurement errors but also because of inherent variability in cracking from year to year within a section of pavement. We introduce a model that can account for both sources of variability. To accommodate the dynamic nature of pavement conditions over the years, we allow χ to evolve over time, i.e. replace χ with χ_t, and assume that χ_t depends on the past values. Then, the median distress level of a given pavement section can be described by a sequence of curves

$$\mu_t = \alpha \mathrm{e}^{-(\chi_t\rho/\mathrm{AGE}_t)^\beta}, \quad t = 1, 2, \dots. \tag{5}$$

Applying the log-transformation on both sides of Equation (4) results in

$$\log(L_t) = \log(\alpha) - \left(\frac{\chi_t\rho}{\mathrm{AGE}_t}\right)^\beta \mathrm{e}^{\varepsilon_t}. \tag{6}$$

After rearranging terms and applying the log-transformation one more time, we end up with

$$\log\{\log(\alpha) - \log(L_t)\} = \beta \log(\rho) + \beta \log(\chi_t) - \beta \log(\mathrm{AGE}_t) + \varepsilon_t \tag{7}$$

which is linear in $\log(\chi_t)$. Let $Y_t = \log\{\log(\alpha) - \log(L_t)\}$, $X_t = \log(\mathrm{AGE}_t)$, and $\lambda_t = \log(\chi_t)$. We model the time evolution of the transformed projection parameter λ_t as an autoregressive (AR) process of order 1, AR(1), i.e.

$$\lambda_t = \lambda_{t-1}\phi + \nu_t. \tag{8}$$

We also assume that both errors, ε_t and ν_t, follow normal distributions with unknown variances. Then, we can write the model in the dynamic linear model [10] or state-space model [9] form as

$$\begin{aligned} &\text{Observation equation: } Y_t = \beta \log(\rho) - \beta X_t + \beta\lambda_t + \varepsilon_t, \quad \varepsilon_t \sim N(0, \Sigma). \\ &\text{State equation: } \lambda_t = \lambda_{t-1}\phi + \nu_t, \quad \nu_t \sim N(0, V). \\ &\text{Initial information: } (\lambda_1 | D_0) \sim N(0, M), \end{aligned} \tag{9}$$

where D_0 represents all the available relevant starting information (at time 0) that is used to form initial views about the future, $M = V/(1-\phi^2)$, and the error sequences ε_t and ν_t are independent and mutually independent. Note that the mean of $\lambda_t (= \log(\chi_t))$ is assumed to be 0. This is equivalent to assuming that the median of χ_t is 1. This assumption can be justified by our prior expectation that the performance curves would be centered around the family curve.

Assume, we have n observations in the data. We wish to estimate the unknown parameters, $\lambda_1, \ldots, \lambda_n$, Σ, ϕ, and V, and predict the future values of λ, for example, λ_{n+1} (equivalently χ_{n+1}). Bayesian inferences are made based on the joint posterior distribution for the parameters:

$$\text{Posterior} \propto \text{likelihood} \times \text{prior}$$

From the observation equation in model (9), the likelihood $f(\mathbf{Y}|\cdots)$ can be written as

$$f(\mathbf{Y}|\cdots) = (2\pi\Sigma)^{-n/2} \exp\left\{-\frac{1}{2\Sigma}\sum_{t=1}^{n}(Y_t - \beta\log(\rho) + \beta X_t - \beta\lambda_t)^2\right\}. \tag{10}$$

We use '$|\ldots$' to denote conditioning on all other variables. The Bayesian model specification is completed with the specification of priors. For a prior distribution $p(\cdot)$, here we assume

$$p(\Sigma, \phi, V, \lambda_1, \ldots, \lambda_n) = p(\Sigma)p(\phi)p(V)p(\lambda_1, \ldots, \lambda_n|\phi, V), \tag{11}$$

abusing notation with the multiple use of $p(\cdot)$ to represent any density.

Note that the state equation in (9) implies

$$p(\lambda_1, \ldots, \lambda_n|\phi, V) = (2\pi)^{-n/2}M^{-1/2}\exp\left(-\frac{\lambda_1^2}{2M}\right)V^{-(n-1)/2}\exp\left\{-\frac{1}{2V}\sum_{t=2}^{n}(\lambda_t - \lambda_{t-1}\phi)^2\right\}, \tag{12}$$

where $M = V/(1-\phi^2)$. For the autoregression coefficient ϕ we use uniform prior, $p(\phi) = \mathbf{I}(0 < \phi < 1)$, according to our prior belief that λ_t is positively correlated. For the variance parameters Σ and V, conventional inverse gamma priors are chosen. Namely,

$$\Sigma^{-1} \sim \text{Gamma}(a_0, b_0), \tag{13}$$

which specifies for Σ^{-1} a prior mean and variance of a_0/b_0 and a_0/b_0^2. Likewise,

$$V^{-1} \sim \text{Gamma}(c_0,\ d_0). \tag{14}$$

Care must be taken in selecting the hyperparameters, $a_0, b_0, c_0,$ and d_0, because the results of the analyses depend on those parameters. The choice of hyperparameters is discussed as part of the analysis of the data in Section 4.

3. Estimation and prediction of future pavement condition

Under the priors in Section 2, the joint posterior density of $\lambda_1, \ldots, \lambda_n, \Sigma,\ \phi,$ and V is proportional to

$$\begin{aligned}
&f(\mathbf{Y}|\cdots)p(\Sigma)p(\phi)p(V)p(\lambda_1,\ldots,\lambda_n|\phi, V)\\
&= (2\pi\Sigma)^{-n/2}\exp\left\{-\frac{1}{2\Sigma}\sum_{t=1}^{n}(Y_t - \beta\log(\rho) + \beta X_t - \beta\lambda_t)^2\right\}\\
&\quad\times \frac{b_0^{a_0}}{\Gamma(a_0)}\left(\frac{1}{\Sigma}\right)^{a_0+1}\exp\left(-\frac{b_0}{\Sigma}\right)\times \mathbf{I}(0<\phi<1)\times\frac{d_0^{c_0}}{\Gamma(c_0)}\left(\frac{1}{V}\right)^{c_0+1}\exp\left(-\frac{d_0}{V}\right)\\
&\quad\times (2\pi)^{-n/2}M^{-1/2}\exp\left(-\frac{\lambda_1^2}{2M}\right)V^{-(n-1)/2}\exp\left\{-\frac{1}{2V}\sum_{t=2}^{n}(\lambda_t - \lambda_{t-1}\phi)^2\right\}.
\end{aligned} \tag{15}$$

Posterior inferences on the parameters $\lambda_1, \ldots, \lambda_n$, Σ, ϕ, and V require high-dimensional integration of Equation (15), which is apparently analytically intractable in this case. Also, the complexity of Equation (15) makes a direct simulation from it difficult. We employ Markov chain Monte Carlo (MCMC) methods [2,4,7] to obtain posterior summaries of the parameters $\lambda_1, \ldots, \lambda_n$, Σ, ϕ, and V. In our MCMC algorithm, one sweep consists of four updating procedures:

(a) updating $\lambda_1, \ldots, \lambda_n$

(b) updating Σ

(c) updating V

(d) updating ϕ.

Samples of $\lambda_1, \ldots, \lambda_n$ can be obtained using forward filtering backward sampling algorithm [10], which is based on the use of Kalman filtering [3]. The full conditional distributions for Σ and V are inverse gamma distributions, so obtaining samples from them is simple. The full conditional distribution of ϕ is not given as the functional form, so we use the Metropolis–Hastings algorithm [4]. The details are given in Appendix A.

Using our state-space model, we can also produce short-term predictions (e.g. one-step ahead prediction). The estimate of the state vector λ_n summarizes the information from the past (up to time n) that is needed to forecast the future. Given samples from the posterior distribution of λ_n, we can simulate from the distribution of λ_{n+1} using the state equation. The samples for λ_{n+1} can then be back-transformed to χ_{n+1} using the relationship $\chi_{n+1} = e^{\lambda_{n+1}}$, and the median values of the future observations given current information is obtained as

$$\mu_{n+1} = \alpha e^{-(\chi_{n+1}\rho/\text{AGE}_{n+1})^{\beta}}. \tag{16}$$

Remark 1 Model (4) can easily be extended to account for the unknown δ by replacing AGE_t by $\text{AGE}_t - \delta$. Under a uniform prior $\delta \sim U(\theta_1, \theta_2)$ that is independent of the priors on the other parameters, δ can be updated by adding a move, (e) updating δ, after moves (a)–(d). Because the full conditional distribution of δ is not recognizable as having the functional form of any standard density, a Metropolis–Hastings update is natural (as for the full conditional distribution of ϕ). The details are given in Appendix A. In Section 4, real examples for non-zero δ are provided.

4. Application to real data

The methods are applied to the longitudinal cracking data obtained from three Strategic Highway Research Program (SHRP) Long-Term Pavement Performance (LTPP) GPS sites in Texas. The values for TxDOT performance coefficients α, β, and ρ, are given as 500, 0.90, and 19.06, respectively, for all those sites.

We set the hyperparameters a_0 and c_0 of the inverse gamma prior distributions for Σ and V equal to 2 as a default. This value for a_0 and c_0 reflects the lack of precise knowledge on the values of Σ and V. There has not been a universal agreement on the best way of choosing the scale parameter of the inverse gamma distribution. The scale parameters b_0 and d_0 are equal to the prior means of Σ and V in our case (with specification of $a_0 = c_0 = 2$). Recall that Σ and V describe the two different sources of variability in our model, where Σ is related to a random perturbation in the measurement process and V is related to the dynamic nature of the model. Roughly speaking, a large value of Σ implies large measurement errors in the observations (deviations between the observation and the individual curves), and a large value of V implies a big curve-to-curve variability (dynamic changes in the performance curve). Thus, a large value of b_0 is preferred when a measurement process is presumed to involve a large measurement error, and a large value of d_0 is preferred if it is believed that there is a big curve-to-curve variability. In practice, it is not easy to predetermine the size of these variabilities. Instead, the deviations of the observations about the family curve can be used as a guideline. Large variability of the observations around the family curve is associated with either a big measurement error, or a dynamic change of the χ values over time, or both.

A value of d_0 can be elicited from engineering experts as follows. Our previous experience, on analyzing pavements throughout the USA, is that χ typically varies between 0.5 and 1.5, though it can be as small as 0.25 or as large as 3 for some extreme cases. As a rule-of-thumb, we use the approximate relationship between the range (maximum value–minimum value) and the variance, variance $\approx$ (range/4)^2. Note that V is the variance of errors in the transformed χ variable, $\lambda = \log(\chi)$, in our model. From $0.5 \le \chi \le 1.5$, the range of λ is obtained as $\log(0.5) \le \lambda \le \log(1.5)$, and this subsequently leads to a candidate value for d_0 as,

$$d_0 = \left\{ \frac{\log(1.5) - \log(0.5)}{4} \right\}^2 = 0.0754.$$

From $0.25 \le \chi \le 3$, a candidate value for d_0 is given as,

$$d_0 = \left\{ \frac{\log(3) - \log(0.25)}{4} \right\}^2 = 0.3859.$$

Depending on our prior belief about χ values (a large value of d_0 is preferred if it is believed that there is a big curve-to-curve variability), either $d_0 = 0.0754$ or $d_0 = 0.3859$ is used selectively in our data analysis. It should be remembered, however, that one may try other plausible values.

For b_0, the deviations of observations about the family curve are used as a basis for predetermining the error variance in λ because the individual curve coefficients are unknown until they are estimated. The average squared differences between the distress values and the family curve in a transformed scale can be used as its default value, i.e.

$$b_0 = \frac{\sum_{t=1}^{n}[\log\{\log(\alpha) - \log(L_t)\} - \log\{\log(\alpha) - \log(f_t)\}]^2}{n}, \tag{17}$$

where $f_t = \alpha e^{-(\rho/\mathrm{AGE}_t)^\beta}$. Note that this might be considered as an empirical Bayes approach because the data is used to specify the hyperpameter value b_0. This may, however, be replaced with relevant expert information on the size of the measurement error (from a source external to the data) or the information from historical data, whichever is available.

In applying our methods to several longitudinal cracking datasets, we collect the posterior samples of size 5000 after a burn-in period of 500,000 by subsampling every 100th from the 500,000 subsequent values. We monitored trace plots of parameters $\lambda_1, \ldots, \lambda_n, \lambda_{n+1}, \chi_1, \ldots, \chi_n, \chi_{n+1}, \mu_{n+1}, \Sigma, V$, and ϕ, to ensure the chain has converged to the area of high posterior density by the end of the burn-in period. We also inspected the autocorrelation function plots of posterior samples for those parameters, though we do not present any of those plots in the article due to limited space. No significant serial correlation was observed from those autocorrelation function plots.

We first apply our method on the dataset presented in Figure 2, site 481049. Appendix B contains the dataset from this site as well as the datasets from two additional sites. For illustration purposes, we pretend that we have only one ($n = 1$), two ($n = 2$), three ($n = 3$), or four observations ($n = 4$) in the dataset, and see how the projection parameter value χ_{n+1} is adjusted from the initial guess of 1 as a new observation is entered. Table 1 contains the posterior summaries for model parameters, $\lambda_{n+1}, \chi_{n+1}, \mu_{n+1}, \Sigma, V$, and ϕ, as n changes. The parameters of main interest in our case are χ_{n+1} and μ_{n+1}. Note that in computing $\mu_{N+1} = \alpha e^{-(\chi_{n+1}\rho/\mathrm{AGE}_{N+1})^\beta}$ where N is the total number of observations in the dataset (e.g. $N = 4$ in Figure 2), the value for AGE_{N+1} is obtained by adding the mean difference in $\mathrm{AGE}_1, \ldots, \mathrm{AGE}_N$ to AGE_N. Because two choices for the hyperparameter d_0 are available, the MCMC analysis is run for both values of d_0 to determine how sensitive the results are to these choices. The parameter that depends most heavily on the value of d_0 seems to be the variance of the state vectors, V, as expected. For the key parameters, χ_{n+1} and μ_{n+1}, however, the posterior medians do not change much between the two choices of d_0, although the posterior distributions obtained under $d_0 = 0.3859$ are more variable (resulting in wider credible intervals) than those obtained under $d_0 = 0.0754$. Table 1 illustrates how important it is to incorporate good engineering judgment in the analyses when the same size is extremely small. Prior influence diminishes in general as the sample size increases. This diminishing prior influence is conspicuous for the posterior summaries of V. As a matter of fact, the difference almost monotonically decreases as the sample size increases for all the posterior summaries of V. This is probably because the hyperparameter for the prior distribution of V is being changed in this sensitivity analysis. It should be noted, however, that for small sample sizes the differences of estimates of parameters obtained with different priors need not converge monotonically to zero.

To assess how well prediction from the suggested model works, the 80% prediction intervals on the future observed distress level at AGE_{n+1} were also obtained and compared with the observed distress level L_{n+1} ($n = 1, 2$, or 3) in the data although none of those intervals are presented here due to limited space. As a matter of fact, the 80% prediction intervals always contain the 80% credible intervals on μ_{n+1} presented in Table 1. It turns out that all but one 80% credible interval (that is for the case $n = 2$, $d_0 = 0.0754$) capture the observed L_{n+1}, and even for that case the 80% prediction interval (which is wider than the 80% credible interval) captures L_{n+1} (as does the 90% credible interval).

Table 1. Summaries of posterior distributions for λ_{n+1}, χ_{n+1}, μ_{n+1}, Σ, V, and ϕ with two choices of hyperparameter value d_0 based on the increasing number of observations at Site 481049.

n	Hyperparameters	Posterior summary	λ_{n+1}	χ_{n+1}	μ_{n+1}	Σ	V	ϕ
1	$a_0 = c_0 = 2$	Mean	−0.0745	0.9630	67.8106	0.0375	0.0604	0.4593
	$b_0 = 0.0422$	Median	−0.0656	0.9365	63.0578	0.0243	0.0401	0.4566
	$d_0 = 0.0754$	SD	0.2691	0.2937	31.8777	0.0598	0.0734	0.2715
		IQR	0.3032	0.2830	35.7593	0.0263	0.0413	0.4597
		LCI	−0.3867	0.6793	33.5098	0.0107	0.0186	0.0891
		UCI	0.2302	1.2589	106.0174	0.0715	0.1153	0.8428
	$a_0 = c_0 = 2$	Mean	−0.0933	1.0755	78.5455	0.0381	0.2769	0.4345
	$b_0 = 0.0422$	Median	−0.0949	0.9095	66.5439	0.0244	0.1878	0.4158
	$d_0 = 0.3859$	SD	0.5280	2.0467	59.7809	0.0487	0.3409	0.2655
		IQR	0.6143	0.5707	72.9249	0.0269	0.1918	0.4418
		LCI	−0.7046	0.4943	14.9854	0.0109	0.0874	0.0812
		UCI	0.5193	1.6809	155.7122	0.0759	0.5262	0.8160
2	$a_0 = c_0 = 2$	Mean	0.0467	1.0854	92.5881	0.0399	0.0576	0.4115
	$b_0 = 0.0461$	Median	0.0415	1.0423	90.4272	0.0275	0.0410	0.3853
	$d_0 = 0.0754$	SD	0.2575	0.3405	33.7162	0.0454	0.0609	0.2651
		IQR	0.2992	0.3137	41.4246	0.0281	0.0397	0.4342
		LCI	−0.2524	0.7769	52.4525	0.0121	0.0191	0.0714
		UCI	0.3479	1.4161	134.5094	0.0777	0.1073	0.8013
	$a_0 = c_0 = 2$	Mean	0.0852	1.2450	94.1389	0.0389	0.2292	0.4141
	$b_0 = 0.0461$	Median	0.0837	1.0874	84.6227	0.0267	0.1726	0.3869
	$d_0 = 0.3859$	SD	0.4986	0.9699	60.7847	0.0436	0.2086	0.2657
		IQR	0.5652	0.6254	75.3099	0.0273	0.1490	0.4412
		LCI	−0.5040	0.6041	25.4519	0.0118	0.0834	0.0733
		UCI	0.6577	1.9304	175.3472	0.0747	0.4214	0.8016
3	$a_0 = c_0 = 2$	Mean	0.1090	1.1605	95.1581	0.0668	0.0592	0.4524
	$b_0 = 0.0854$	Median	0.0948	1.0994	93.8485	0.0491	0.0424	0.4411
	$d_0 = 0.0754$	SD	0.2771	0.3561	36.4967	0.0615	0.0633	0.2707
		IQR	0.3241	0.3605	45.3954	0.0479	0.0413	0.4588
		LCI	−0.2097	0.8108	49.9931	0.0223	0.0196	0.0882
		UCI	0.4492	1.5671	140.1272	0.1279	0.1116	0.8329
	$a_0 = c_0 = 2$	Mean	0.1545	1.3143	94.8143	0.0622	0.2162	0.4342
	$b_0 = 0.0854$	Median	0.1463	1.1576	86.6875	0.0457	0.1646	0.4162
	$d_0 = 0.3859$	SD	0.4822	0.7630	60.4033	0.0652	0.1962	0.2658
		IQR	0.5737	0.6851	76.5119	0.0439	0.1368	0.4398
		LCI	−0.4245	0.6541	25.2711	0.0208	0.0839	0.0848
		UCI	0.7375	2.0907	175.2175	0.1160	0.3857	0.8118
4	$a_0 = c_0 = 2$	Mean	−0.0122	1.0202	130.7160	0.0563	0.0545	0.3892
	$b_0 = 0.0721$	Median	−0.0106	0.9895	128.9800	0.0441	0.0408	0.3600
	$d_0 = 0.0754$	SD	0.2509	0.2972	37.5323	0.0446	0.0488	0.2581
		IQR	0.2886	0.2871	45.2155	0.0392	0.0385	0.4115
		LCI	−0.3025	0.7389	86.5136	0.0203	0.0188	0.0646
		UCI	0.2760	1.3179	176.3866	0.1039	0.1044	0.7736
	$a_0 = c_0 = 2$	Mean	−0.0495	1.0554	139.2976	0.0526	0.1923	0.3924
	$b_0 = 0.0721$	Median	−0.0426	0.9583	134.0369	0.0385	0.1549	0.3624
	$d_0 = 0.3859$	SD	0.4533	0.5469	64.9710	0.0472	0.1410	0.2550
		IQR	0.5504	0.532	86.4201	0.0368	0.1205	0.4120
		LCI	−0.5948	0.5517	61.0761	0.0179	0.0809	0.0703
		UCI	0.4775	1.6120	224.4435	0.1005	0.3389	0.7681

Notes: 1. SD stands for the posterior standard deviation; 2. IQR stands for the interquartile range; 3. LCI and UCI stand for the lower limit and upper limit of the 80% credible interval.

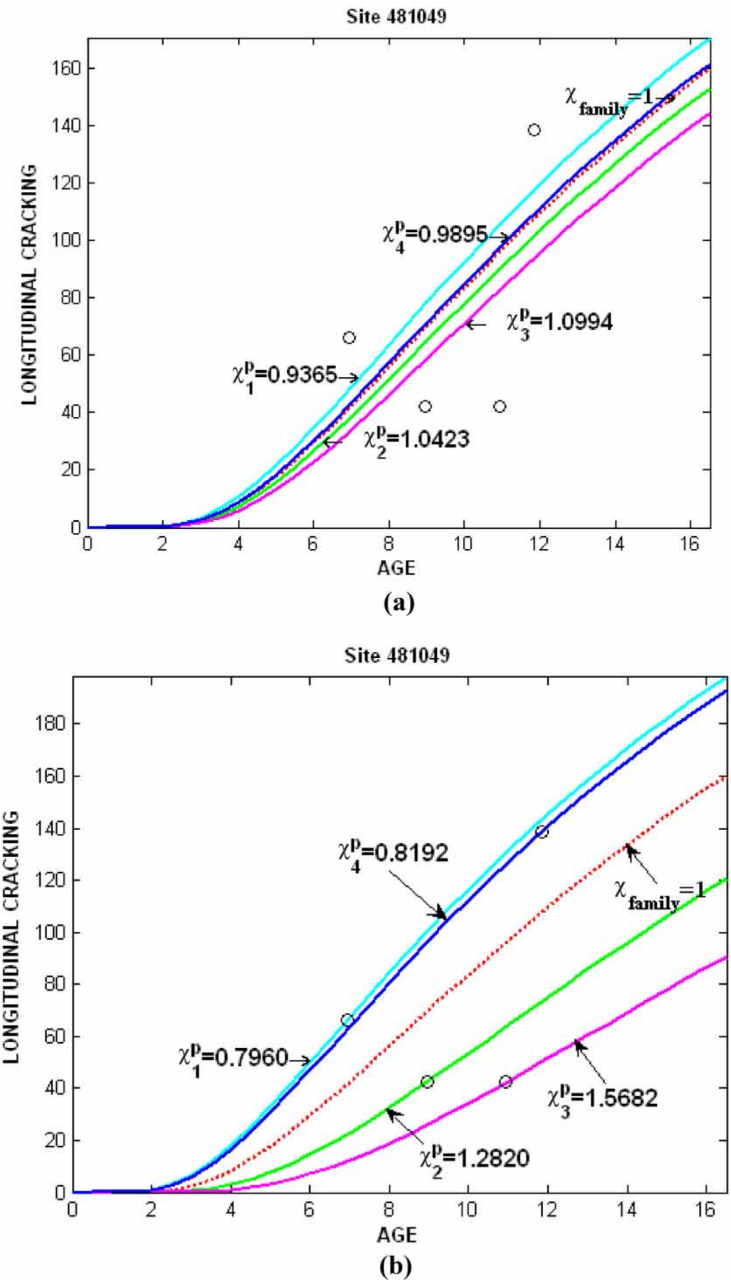

Figure 6. Changes in prediction curves: (a) proposed method; (b) TxDOT method 2.

Figure 6 shows how the projection parameter and the corresponding prediction curve changes as a new observation enters. For comparison, prediction curves from the proposed method and prediction curves obtained from TxDOT method 2 are presented in Figures 6a and 6b, respectively. Let us introduce a new notation χ_n^p to denote the projection parameter based on n observations (χ_{n+1}). Then in the figures χ_1^p, χ_2^p, χ_3^p, and χ_4^p represent the projection parameter values when we have one, two, three, and four observations, respectively, in the dataset. Note that the family curve with the projection parameter 1 is used as a prediction curve before we observe any data. The posterior median of χ_{n+1} obtained with a hyperparameter $d_0 = 0.0754$ in Table 1 is used as an estimate for the projection parameter in Figure 6a for each sample size. It can be observed from Figure 6b that there is a huge variability among the prediction curves, which immediately implies large uncertainty in prediction. On the other hand, prediction curves in Figure 6a show much less variability. Also, recall that all the 80% prediction intervals that accompanied each prediction curve in Figure 6a captured the observation at AGE_{n+1}, whereas no interval predictions can be provided for the TxDOT method 2 (Figure 6b). Although we considered $n = 1, 2,$ or 3 cases for purposes of illustration, in practice we are not interested in projecting the past, but in projecting the future (e.g. next year's pavement condition) using all of the N observations in the data, i.e. $n = N$. Thus, the curve corresponding to $\chi_N^p(\chi_4^p)$ will be of final interest, and it can be used for projecting next year's pavement condition. Note that there is a huge difference between the proposed method and the TxDOT method 2 in the estimated projection parameter for the last curve.

The proposed method is illustrated with the dataset from a second site, Site 482108 (Appendix B). Figure 7 shows the observations from the site with the family curve and the prediction curves based on the observations up to time n ($n = 1, 2, 3, 4, 5$) superimposed.

Table 2 contains the posterior summaries for model parameters as n changes. The posterior median of χ_{n+1} obtained with a hyperparameter $d_0 = 0.0754$ in Table 2 is used as an estimate for the projection parameter for each curve obtained based on the observations up to time n ($n = 1, 2, 3, 4,$ and 5) in Figure 7. Again, the MCMC analysis is run for both values of d_0. It can be observed from Table 2 that in general the difference in the posterior summaries between inferences based on $d_0 = 0.0754$ and $d_0 = 0.3859$ decreases as the sample size increases. Also, note that the posterior medians of the key parameters, χ_{n+1} and μ_{n+1}, are almost the same for the two choices of d_0 although the posterior distributions obtained under $d_0 = 0.3859$ are more diffuse (resulting in wider credible intervals) than those obtained under $d_0 = 0.0754$. As before,

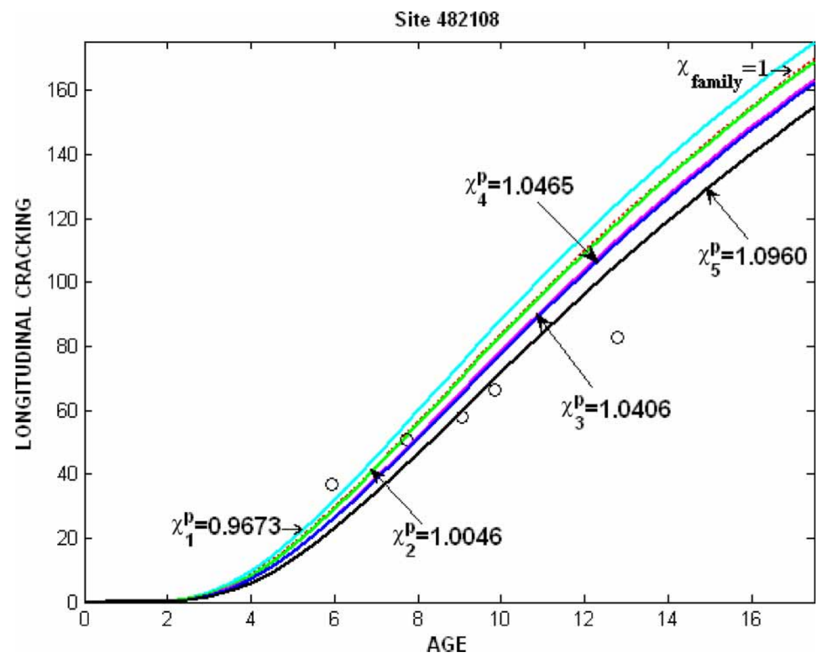

Figure 7. Changes in prediction curves with $\delta = 0$ for Site 482108.

Table 2. Summaries of posterior distributions for λ_{n+1},χ_{n+1}, μ_{n+1}, Σ, V, and ϕ with two choices of hyperparameter value d_0 based on the increasing number of observations at Site 482108.

n	Hyperparameters	Posterior summary	λ_{n+1}	χ_{n+1}	μ_{n+1}	Σ	V	ϕ
1	$a_0 = c_0 = 2$	Mean	−0.0364	0.9934	48.3908	0.0071	0.0532	0.4345
	$b_0 = 0.0080$	Median	−0.0333	0.9673	44.8585	0.0046	0.0366	0.4175
	$d_0 = 0.0754$	SD	0.2416	0.2840	24.2161	0.0090	0.0592	0.2680
		IQR	0.2680	0.2594	26.1438	0.0051	0.0369	0.4504
		LCI	−0.3101	0.7334	23.3619	0.0020	0.0171	0.0790
		UCI	0.2323	1.2615	76.3371	0.0138	0.1016	0.8160
	$a_0 = c_0 = 2$	Mean	−0.0434	1.0962	58.2085	0.0078	0.2607	0.4224
	$b_0 = 0.0080$	Median	−0.0336	0.9669	44.8954	0.0048	0.1789	0.4077
	$d_0 = 0.3859$	SD	0.5074	0.8653	51.9985	0.0128	0.3463	0.2637
		IQR	0.5875	0.5740	57.2257	0.0052	0.1745	0.4363
		LCI	−0.6249	0.5353	9.1912	0.0021	0.0840	0.0751
		UCI	0.5278	1.6952	121.2530	0.0148	0.4780	0.8007
2	$a_0 = c_0 = 2$	Mean	0.0036	1.0256	72.5366	0.0035	0.0420	0.4213
	$b_0 = 0.0041$	Median	0.0046	1.0046	70.2062	0.0023	0.0313	0.4049
	$d_0 = 0.0754$	SD	0.2067	0.2219	25.2850	0.0043	0.0402	0.2635
		IQR	0.2432	0.2448	30.1242	0.0024	0.0281	0.4323
		LCI	−0.2475	0.7807	43.3209	0.0010	0.0154	0.0773
		UCI	0.2488	1.2825	104.5444	0.0068	0.0760	0.8060
	$a_0 = c_0 = 2$	Mean	0.0054	1.0085	79.3831	0.0040	0.1996	0.4246
	$b_0 = 0.0041$	Median	0.0049	1.0049	70.1689	0.0024	0.1485	0.4085
	$d_0 = 0.3859$	SD	0.4540	0.6179	52.9944	0.0063	0.1914	0.2686
		IQR	0.5250	0.5344	64.4107	0.0027	0.1294	0.4545
		LCI	−0.5195	0.5948	21.4809	0.0011	0.0746	0.0716
		UCI	0.5283	1.6960	146.8542	0.0078	0.3613	0.8081
3	$a_0 = c_0 = 2$	Mean	0.0406	1.0622	78.3481	0.0047	0.0367	0.4205
	$b_0 = 0.0060$	Median	0.0398	1.0406	76.5134	0.0032	0.0286	0.3998
	$d_0 = 0.0754$	SD	0.1980	0.2172	25.2210	0.0052	0.0311	0.2648
		IQR	0.2391	0.2497	30.8198	0.0032	0.0231	0.4316
		LCI	−0.1989	0.8197	48.1290	0.0015	0.0147	0.0744
		UCI	0.2850	1.3298	109.9062	0.0090	0.0666	0.8076
	$a_0 = c_0 = 2$	Mean	0.0374	1.1287	84.2318	0.0053	0.1613	0.4299
	$b_0 = 0.0060$	Median	0.0441	1.0451	75.9618	0.0034	0.1263	0.4140
	$d_0 = 0.3859$	SD	0.4084	0.5078	50.6273	0.0075	0.1314	0.2671
		IQR	0.4846	0.5095	62.0807	0.0037	0.0987	0.4503
		LCI	−0.4508	0.6371	28.4439	0.0015	0.0665	0.0781
		UCI	0.5101	1.6654	149.4853	0.0099	0.2835	0.8126
4	$a_0 = c_0 = 2$	Mean	0.0510	1.0701	112.7384	0.0052	0.0315	0.4319
	$b_0 = 0.0074$	Median	0.0454	1.0465	112.4527	0.0037	0.0254	0.4119
	$d_0 = 0.0754$	SD	0.1824	0.2006	26.8452	0.0051	0.0221	0.2641
		IQR	0.2286	0.2408	34.3795	0.0036	0.0189	0.4344
		LCI	−0.1658	0.8472	79.2652	0.0017	0.0138	0.0823
		UCI	0.2790	1.3218	145.5268	0.0099	0.0553	0.8141
	$a_0 = c_0 = 2$	Mean	0.0501	1.1259	116.1156	0.0060	0.1375	0.4272
	$b_0 = 0.0074$	Median	0.0450	1.0460	112.5195	0.0040	0.1122	0.4114
	$d_0 = 0.3859$	SD	0.3673	0.4479	51.6897	0.0082	0.0948	0.2666
		IQR	0.4557	0.4822	68.1474	0.0041	0.0837	0.4495
		LCI	−0.3931	0.6750	51.9393	0.0018	0.0620	0.0778
		UCI	0.5084	1.6626	182.8470	0.0111	0.2380	0.8115
5	$a_0 = c_0 = 2$	Mean	0.0968	1.1229	124.9250	0.0098	0.0325	0.4750
	$b_0 = 0.0163$	Median	0.0917	1.0960	124.6475	0.0075	0.0268	0.4774
	$d_0 = 0.0754$	SD	0.1948	0.2257	29.5518	0.0082	0.0218	0.2650
		IQR	0.2463	0.2718	38.2306	0.0066	0.0198	0.4317
		LCI	−0.1424	0.8672	88.1992	0.0036	0.0145	0.1054
		UCI	0.3386	1.4030	162.2421	0.0181	0.0559	0.8395

(*Continued*)

Table 2. Continued.

n	Hyperparameters	Posterior summary	λ_{n+1}	χ_{n+1}	μ_{n+1}	Σ	V	ϕ
	$a_0 = c_0 = 2$	Mean	0.0994	1.1799	126.9695	0.0114	0.1267	0.4405
	$b_0 = 0.0163$	Median	0.1018	1.1072	123.0704	0.0082	0.1065	0.4284
	$d_0 = 0.3859$	SD	0.3623	0.4545	52.7327	0.0176	0.0807	0.2673
		IQR	0.4504	0.5013	69.4308	0.0077	0.0728	0.4516
		LCI	−0.3493	0.7052	61.5198	0.0039	0.0594	0.0835
		UCI	0.5482	1.7302	196.4291	0.0210	0.2097	0.8199

Notes: 1. SD stands for the posterior standard deviation; 2. IQR stands for the interquartile range; 3. LCI and UCI stand for the lower limit and upper limit of the 80% credible interval.

the diminishing influence of prior is most apparent for the posterior summaries of V. For the other parameters, the difference in the posterior summaries resulting from different d_0 values does not necessarily monotonically decrease, although in general it shows a decreasing pattern as mentioned previously. Although not presented in the paper, all the 80% prediction intervals on the future observed distress level at AGE_{n+1} capture the observed L_{n+1}. (As a matter of fact, even the 80% credible intervals for μ_{n+1} given in Table 2 capture the observed L_{n+1} for $n = 1, 2, 3,$ and 4.)

For the datasets represented in Figures 6 and 7, the value of a horizontal shift factor, δ, has been assumed to be zero. Figure 8a shows the data obtained at a third site, Site 481094 with the family curve ($\chi = 1$) having $\delta = 0$ superimposed, which shows the need for incorporating a non-zero δ value in addition to adjusting χ values. In practice, the value of δ is often estimated by an engineer's judgment. In that case, the data can be adjusted to reflect the non-zero δ by substracting the estimated δ from AGE_t, and the MCMC method can be applied without any modification. Alternatively, the method can easily be generalized to account for the unknown δ with an additional updating step for δ as described in Remark 1 of Section 3.

The hyperparameters for the prior for δ are θ_1 and θ_2. There are several ways to set the lower bound θ_1 and the upper bound θ_2. For example, the lower bound θ_1 is always ≥ 0 and can be further refined to be greater than AGE_{t_0} if the distress level is observed to be zero at AGE_{t_0}. The upper bound θ_2 typically needs to be less than AGE_{t_1} where AGE_{t_1} is the year when the first non-zero distress is observed. If there is no observed zero distress after construction, θ_1 can be taken as 0 and θ_2 can be taken as AGE_{t_1}. For example, for Site 481049 (in Figure 8), the first non-zero distress level is observed at year 14.7945 (Appendix B). Thus, a prior distribution for δ can be taken as $\delta \sim U(0, 14.7945)$. We take the posterior median as an estimate for δ because the posterior distribution based on MCMC samples for δ seems to be truncated at zero. Table 3 contains the posterior summaries of model parameters as n changes for two different values of d_0. Again, it can be observed that influence of d_0 diminishes as the sample size increases with the most evident diminishing effect on the posterior summaries of V. Also, note that the posterior medians of the key parameters, χ_{n+1} and μ_{n+1}, are very close for the two choices of d_0 although the posterior distributions obtained under $d_0 = 0.3859$ are more diffuse than those obtained under $d_0 = 0.0754$. Although not presented in the paper, all the 80% prediction intervals on the future observed distress level at AGE_{n+1} ($n = 1, 2, 3, 4, 5, 6$) capture the observed L_{n+1}. (As a matter of fact, even the 80% credible intervals for μ_{n+1} given in Table 3 capture the observed L_{n+1} for all but one case ($n = 5$, $d_0 = 0.0754$).)

Figure 8b shows the prediction curve with the estimated projection parameter based on all seven observations and the family curve with the estimated δ value. The posterior medians

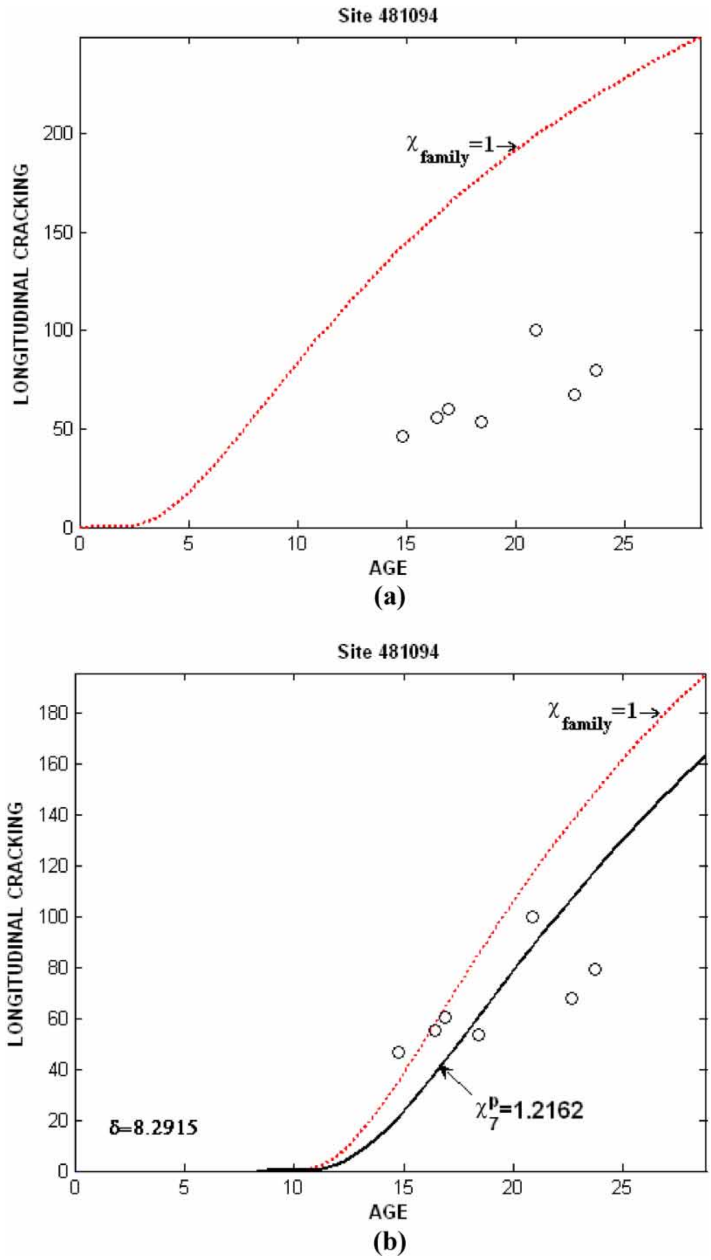

Figure 8. Longitudinal cracking data at Site 481094: (a) family curve with $\delta = 0$: (b) family curve and the prediction curve with an estimated $\delta = 8.2915$.

of χ_{N+1} and δ obtained with a hyperparameter $d_0 = 0.3859$ are used as the estimates for χ_N^p and δ, respectively. The resulting estimates appear to be consistent with engineering expert judgments.

The changes in prediction curves over time seem to be reasonable. It can be observed from Figure 6a that a prediction curve does not change significantly from the family curve when the data points are randomly scattered around the family curve, which can be contrasted with the TxDOT method (Figure 6b) that moves a curve each time a new data value is entered to go through the new observation. As opposed to the TxDOT method, our approach attempts to capture the systematic change of the system only. If the observations consistently fall under or above the family curve, or gradually move to one direction, then the prediction curves will reflect the systematic trend, and eventually be changed toward that direction (Figures 7 and 8).

Table 3. Summaries of posterior distributions for λ_{n+1}, χ_{n+1}, μ_{n+1}, Σ, V, and ϕ with two choices of hyperparameter value d_0 based on the increasing number of observations at Site 481094.

n	Hyperparameters	Posterior summary	δ	λ_{n+1}	χ_{n+1}	μ_{n+1}	Σ	V	ϕ
1	$a_0 = c_0 = 2$	Mean	6.5028	0.0228	1.0880	72.6702	0.0649	0.0655	0.4835
	$b_0 = 0.0703$	Median	6.7721	0.0217	1.0220	65.2906	0.0415	0.0438	0.4848
	$d_0 = 0.0754$	SD	2.6921	0.3428	0.5013	43.0665	0.0826	0.0796	0.2823
		IQR	3.7825	0.3606	0.3714	54.1201	0.0463	0.0457	0.4922
		LCI	2.5908	−0.3477	0.7063	24.2625	0.0177	0.0187	0.0926
		UCI	9.8126	0.4017	1.4943	129.4474	0.1261	0.1270	0.8791
	$a_0 = c_0 = 2$	Mean	6.3029	0.0134	1.2297	83.2746	0.0649	0.3041	0.4623
	$b_0 = 0.0703$	Median	6.3711	0.0209	1.0212	67.0661	0.0413	0.2048	0.4495
	$d_0 = 0.3859$	SD	3.2658	0.6208	0.9922	69.6665	0.0887	0.5236	0.2733
		IQR	5.1590	0.7162	0.7529	90.5128	0.0436	0.2048	0.4640
		LCI	1.7316	−0.6990	0.4971	9.2443	0.0181	0.0927	0.0951
		UCI	10.5942	0.7333	2.0820	177.4044	0.1220	0.5614	0.8433
2	$a_0 = c_0 = 2$	Mean	6.8957	0.0364	1.0860	79.9475	0.0643	0.0574	0.4852
	$b_0 = 0.0884$	Median	7.2561	0.0284	1.0288	74.0851	0.0451	0.0403	0.4857
	$d_0 = 0.0754$	SD	2.4254	0.3011	0.3643	39.8052	0.0841	0.0596	0.2804
		IQR	3.1509	0.3261	0.3393	48.4872	0.0439	0.0395	0.4776
		LCI	3.4320	−0.3014	0.7398	34.9447	0.0205	0.0184	0.0961
		UCI	9.7101	0.3917	1.4794	129.9265	0.1199	0.1069	0.8686
	$a_0 = c_0 = 2$	Mean	6.5356	0.0585	1.2447	87.1791	0.0696	0.2395	0.4675
	$b_0 = 0.0884$	Median	6.7778	0.0619	1.0639	74.4858	0.0477	0.1741	0.4623
	$d_0 = 0.3859$	SD	3.1032	0.5483	1.0667	64.3187	0.0772	0.2348	0.2699
		IQR	4.6529	0.6274	0.6754	82.2738	0.0487	0.1580	0.4516
		LCI	2.0760	−0.5912	0.5537	15.8077	0.0214	0.0857	0.0993
		UCI	10.4618	0.6990	2.0117	171.9877	0.1326	0.4386	0.8458
3	$a_0 = c_0 = 2$	Mean	7.2382	0.0471	1.0934	93.3328	0.0552	0.0486	0.4971
	$b_0 = 0.0941$	Median	7.5759	0.0332	1.0337	90.2776	0.0417	0.0358	0.5065
	$d_0 = 0.0754$	SD	2.2203	0.2837	0.3614	36.4450	0.0485	0.0449	0.2748
		IQR	2.7169	0.3290	0.3445	45.3183	0.0374	0.0327	0.4675
		LCI	4.1651	−0.2759	0.7589	50.5929	0.0199	0.0174	0.1109
		UCI	9.7393	0.3884	1.4746	139.8575	0.1013	0.0899	0.8679
	$a_0 = c_0 = 2$	Mean	6.7777	0.0631	1.2330	101.0088	0.0631	0.2025	0.4752
	$b_0 = 0.0941$	Median	7.0723	0.0563	1.0580	90.1906	0.0471	0.1553	0.4759
	$d_0 = 0.3859$	SD	2.9381	0.5191	1.3055	64.3260	0.0687	0.1961	0.2745
		IQR	4.2056	0.6008	0.6545	83.1777	0.0439	0.1337	0.4663
		LCI	2.4915	−0.5462	0.5792	27.9932	0.0220	0.0780	0.0967
		UCI	10.4058	0.6789	1.9717	186.7776	0.1149	0.3629	0.8540
4	$a_0 = c_0 = 2$	Mean	7.5357	0.0975	1.1495	111.5760	0.0677	0.0487	0.5062
	$b_0 = 0.1305$	Median	7.9575	0.0753	1.0782	109.5890	0.0521	0.0364	0.5151
	$d_0 = 0.0754$	SD	2.2586	0.2822	0.3642	37.7321	0.0551	0.0437	0.2828
		IQR	2.6893	0.3240	0.3543	48.0109	0.0443	0.0341	0.4899
		LCI	4.4452	−0.2213	0.8015	65.0345	0.0265	0.0172	0.1049
		UCI	10.0143	0.4523	1.5719	158.7127	0.1220	0.0911	0.8817
	$a_0 = c_0 = 2$	Mean	7.1340	0.1306	1.2901	114.7738	0.0782	0.1917	0.4835
	$b_0 = 0.1305$	Median	7.5203	0.1216	1.1293	107.5281	0.0590	0.1499	0.4863
	$d_0 = 0.3859$	SD	2.8125	0.4859	0.7999	63.7678	0.0721	0.1560	0.2702

(*Continued*)

Table 3. Continued.

n	Hyperpara meters	Posterior summary	δ	λ_{n+1}	χ_{n+1}	μ_{n+1}	Σ	V	ϕ
		IQR	3.8618	0.5899	0.6803	85.6076	0.0534	0.1214	0.4500
		LCI	3.0636	−0.4442	0.6413	37.6412	0.0284	0.0772	0.1066
		UCI	10.5087	0.7065	2.0268	199.9082	0.1436	0.3443	0.8530
5	$a_0 = c_0 = 2$	Mean	7.7625	0.0682	1.1084	132.8141	0.0577	0.0431	0.5000
	$b_0 = 0.1239$	Median	8.1133	0.0573	1.0590	131.6961	0.0454	0.0337	0.5070
	$d_0 = 0.0754$	SD	2.0631	0.2587	0.3150	37.0456	0.0477	0.0353	0.2796
		IQR	2.3502	0.2950	0.3138	46.8000	0.0363	0.0284	0.4735
		LCI	5.0505	−0.2307	0.7940	87.6677	0.0238	0.0168	0.1079
		UCI	9.9832	0.3857	1.4707	179.7108	0.1019	0.0781	0.8849
	$a_0 = c_0 = 2$	Mean	7.2732	0.0902	1.2256	136.3470	0.0666	0.1658	0.4838
	$b_0 = 0.1239$	Median	7.7039	0.0798	1.0830	132.3429	0.0520	0.1331	0.4827
	$d_0 = 0.3859$	SD	2.7589	0.4481	1.2487	62.6790	0.0585	0.1302	0.2704
		IQR	3.6282	0.5419	0.6015	84.0458	0.0442	0.1025	0.4523
		LCI	3.1641	−0.4442	0.6413	57.2225	0.0258	0.0716	0.1090
		UCI	10.5053	0.6421	1.9005	218.2729	0.1187	0.2895	0.8554
6	$a_0 = c_0 = 2$	Mean	7.9597	0.1491	1.2119	128.9494	0.0745	0.0462	0.5289
	$b_0 = 0.1684$	Median	8.3961	0.1235	1.1314	128.5827	0.0610	0.0366	0.5484
	$d_0 = 0.0754$	SD	2.2179	0.2872	0.3829	39.4118	0.0515	0.0354	0.2855
		IQR	2.4950	0.3597	0.4163	50.3162	0.0460	0.0311	0.4964
		LCI	4.7408	−0.1851	0.8311	79.1973	0.0324	0.0177	0.1181
		UCI	10.2744	0.5211	1.6839	177.8632	0.1306	0.0858	0.9022
	$a_0 = c_0 = 2$	Mean	7.5101	0.1960	1.3556	128.7606	0.0826	0.1607	0.5111
	$b_0 = 0.1684$	Median	7.9601	0.1805	1.1979	124.7306	0.0669	0.1320	0.5234
	$d_0 = 0.3859$	SD	2.7846	0.4516	0.7557	61.8379	0.0612	0.1027	0.2723
		IQR	3.7245	0.5685	0.6941	84.6866	0.0520	0.0993	0.4501
		LCI	3.3051	−0.3532	0.7024	51.2761	0.0342	0.0727	0.1196
		UCI	10.6801	0.7645	2.1480	211.7196	0.1432	0.2771	0.8727
7	$a_0 = c_0 = 2$	Mean	7.9804	0.1932	1.2729	136.0641	0.0770	0.0459	0.5709
	$b_0 = 0.1944$	Median	8.4924	0.1546	1.1672	136.2073	0.0646	0.0363	0.6055
	$d_0 = 0.0754$	SD	2.3509	0.3009	0.4311	41.2154	0.0495	0.0365	0.2800
		IQR	2.6595	0.3756	0.4520	53.5058	0.0463	0.0300	0.4650
		LCI	4.6410	−0.1538	0.8575	82.8752	0.0352	0.0180	0.1517
		UCI	10.4201	0.5985	1.8194	187.5491	0.1317	0.0828	0.9241
	$a_0 = c_0 = 2$	Mean	7.7768	0.2130	1.3680	136.5745	0.0840	0.1525	0.5225
	$b_0 = 0.1944$	Median	8.2915	0.1957	1.2162	132.5762	0.0691	0.1273	0.5407
	$d_0 = 0.3859$	SD	2.7627	0.4424	0.6768	61.8143	0.0558	0.0950	0.2740
		IQR	3.6610	0.5599	0.7007	85.4779	0.0529	0.0917	0.4570
		LCI	3.6953	−0.3288	0.7198	58.7344	0.0373	0.0707	0.1210
		UCI	10.8773	0.7814	2.1846	219.1312	0.1454	0.2590	0.8782

Notes: 1. SD stands for the posterior standard deviation; 2. IQR stands for the interquartile range; 3. LCI and UCI stand for the lower limit and upper limit of the 80% credible interval.

5. Summary and conclusions

We have developed a Bayesian data analysis tool for periodically updating the expected pavement performance and demonstrated the method on three examples. The methods were applied to the longitudinal cracking data obtained from several SHRP, LTPP, GPS sites in Texas. In our implementation of MCMC, we showed how engineers could provide useful information that leads to reasonable choices for priors. This is important because in typical cases the sample size is very small. We conjectured that the large variability of the observations around the family curve is associated with either a big measurement error or a dynamic change of the χ values over time, or both. We used this variability to set a prior for the variance of measurement error. We then used engineering judgment to set the prior for the variability of χ by having the engineers specify the range of anticipated values for χ. The prior was set using a well-known applied statistics approximation variance $\approx (\text{range}/4)^2$. Once the priors were set, an MCMC calculation was performed, and each of a sequence of updated projections was given and discussed. The third example in this paper demonstrated the need to use a shift parameter δ in the model to account for the fact that road deterioration happened well after the road segment monitoring began. We showed how to obtain a prior and estimates for the shift parameter δ, and most important, how the projection parameter, χ_{n+1}, can be estimated.

The paper demonstrates that modern Bayesian MCMC technology can be used in important applications of roadway maintenance to improve pavement distress predictions greatly, thus leading to better use of stressed highway and road repair budgets. As each observation of pavement distress is made, the Bayesian method updates the predicted future distress within an acceptable engineering and statistical framework. As mentioned earlier, a majority of states (36 of 39 states as in [1]) use distress data for planning road repairs. The methods presented provide a framework for updating pavement distress predictions by highway departments across the USA.

Acknowledgements

The authors wish to thank the Texas Department of Transportation (TxDOT) for financial support (Grant 0-4186) for this work. The contents of this paper reflect the views of the authors who are responsible for the opinions, findings, and conclusions presented herein. The contents do not necessarily reflect the official view or policies of the TxDOT. This paper does not constitute a standard, specification, or regulation.

References

[1] M. Fellers and K. Zimmerman, *Pavement Management System Business Process Review: Market Survey*, California Department of Transportation, Sacramento, CA, 2005.
[2] W.R. Gilks, S. Richardson, and D.J. Spiegelhalter, *Markov Chain Monte Carlo in Practice*, Chapman and Hall, London, 1996.
[3] R.J. Meinhold and N.D. Singpurwalla, *Understanding the Kalman filter*, Am. Stat. 37 (1983), pp. 123–127.
[4] C.P. Robert and G. Casella, *Monte Carlo Statistical Methods,* 2nd edn. Springer-Verlag, New York, 2004.
[5] R.E. Smith, B. Mukerjee, M. Zulyaminayn, C. Pilson, T. Dossey, and B. McCullough, *Integration of network- and project-level performance models for TxDOT PMIS*, Res. Rep. 1727-2, Texas Transportation Institute, College Station, TX, USA, 2001.
[6] B.E. Stampley, B. Miller, R.E. Smith, and T. Scullion, *Pavement management information system concepts, equations, and analysis models*, Res. Rep. 1989–1, Texas Transportation Institute, College Station, TX, 1993.
[7] L. Tierney, *Markov chains for exploring posterior distributions*, *Ann. Stat.* 22 (1994), pp. 1701–1762.
[8] TxDOT, *Pavement Management Information System Rater's Manual for Fiscal Year 2007*, Texas Department of Transportation, 2006.
[9] W.S. Wei, *Time Series Analysis*, Addison-Wesley, New York, 1990.
[10] M. West and J. Harrison, Bayesian Forecasting and Dynamic Models, 2nd edn. Springer-Verlag, New York, 1997.

Appendix A Details in MCMC algorithm

(a) Updating $\lambda_1, \ldots, \lambda_n$

To sample from the full conditional posterior $\pi(\lambda_1, \ldots, \lambda_n | \cdots)$, we sequentially simulate the individual vectors $\lambda_1, \ldots, \lambda_n$ as follows:

(1) Sample λ_n from $N(m_n, C_n)$ where m_n and C_n are obtained from the Kalman filtering recurrences

$$m_t = a_t + \mathrm{e}_t K_t,$$

$$C_t = R_t - K_t^2 Q_t,$$

with

$$
\begin{aligned}
a_t &= m_{t-1}\phi, \\
R_t &= \phi^2 C_{t-1} + V, \\
f_t &= a_t \beta, \\
Q_t &= \beta^2 R_t + \Sigma, \\
K_t &= \frac{\beta R_t}{Q_t}, \\
e_t &= Y_t - \beta \log(\rho) + \beta X_t - f_t.
\end{aligned}
$$

(2) For each $t = n-1, L, n-2, 1$, sample λ_t from $N(h_t, H_t)$ where $h_t = m_t + (\lambda_{t+1} - a_{t+1})B_t$, $H_t = C_t - B_t^2 R_{t+1}$, $B_t = \phi C_t / R_{t+1}$, and λ_{t+1} is the value just sampled.

(b) Updating Σ

The full conditional posterior distribution for Σ is

$$\Sigma^{-1} | \cdots \sim \text{Gamma}\left(a_0 + \frac{1}{2}n, b_0 + \frac{1}{2}d\right),$$

where $d = \sum_{t=1}^{n}(Y_t - \beta \log(\rho) + \beta X_t - \beta\lambda_t)^2$. This can be easily sampled using a Gibbs sampler.

(c) Updating V

The full conditional posterior distribution for V is

$$V^{-1} | \cdots \sim \text{Gamma}\left(c_0 + \frac{1}{2}n, d_0 + \frac{1}{2}\{(1-\phi^2)\lambda_1^2 + G\}\right),$$

where $G = \sum_{t=2}^{n}(\lambda_t - \lambda_{t-1}\phi)^2$. Again, this can be easily sampled using the Gibbs sampler.

(d) Updating ϕ

The full conditional posterior density for ϕ, $\pi(\phi|L)$, is proportional to

$$g(\phi) f_{\text{nor}}(\phi|\tau, \mathrm{T})\mathbf{I}(0 < \phi < 1),$$

where f_{nor} is the normal density function with $T = V \sum_{t=2}^{n} \lambda_{t-1}^2$, $\tau = \sum_{t=2}^{n}(\lambda_t\lambda_{t-1}) / \sum_{t=2}^{n} \lambda_{t-1}^2$, $g(\phi) = M^{-1/2}\exp(-\lambda_1^2/2M)$, $M = V/1-\phi^2$, and $\mathbf{I}(0 < \phi < 1) = \prod_{k=1}^{q} \mathbf{I}(0 < \phi_k < 1)$. We use the truncated normal distribution $f_{\text{nor}}(\phi|\tau, \mathrm{T})\mathbf{I}(0 < \phi < 1)$ as a proposal distribution for (independent proposal) and accept the proposal ϕ^* with probability

$$\min\left\{1, \frac{g(\phi^*)}{g(\phi)}\right\}.$$

(e) Updating δ

The full conditional posterior density for δ, $\pi(\delta|\cdots)$, is proportional to

$$\mathbf{I}(\theta_1 < \delta < \theta_2)\exp\left[-\frac{1}{2\Sigma}\left\{\sum_{t=1}^{n}(\beta\log(\text{AGE}_t - \delta))^2 - 2\sum_{t=1}^{n}\beta\log(\text{AGE}_t - \delta)(\beta\log(\rho) + \beta\lambda_t - Y_t)\right\}\right].$$

We use the uniform distribution, $U(\theta_1, \theta_2)$, as a proposal distribution for δ (independent proposal) and accept the proposal δ^* with probability

$$\min\left\{1, \frac{g(\delta^*)}{g(\delta)}\right\},$$

where

$$g(\delta) = \exp\left[-\frac{1}{2\Sigma}\left\{\sum_{t=1}^{n}(\beta\log(\mathrm{AGE}_t-\delta))^2 - 2\sum_{t=1}^{n}\beta\log(\mathrm{AGE}_t-\delta)(\beta\log(\rho)+\beta\lambda_t-Y_t)\right\}\right].$$

Appendix B Datasets from sites 481049, 482108, and 481094

Sites	Age (in years)	Distress (in linear feet per 100 feet of road segment)
481049	6.9507	66.4
	8.9479	42.3
	10.9452	2.3
	11.8301	138.5
482108	5.9342	36.6142
	7.7425	50.7218
	9.0493	57.8084
	9.8521	66.4042
	12.7863	82.8740
481094	14.7945	46.5879
	16.4110	55.4462
	16.9068	60.4331
	18.4384	53.4777
	20.8959	100.0656
	22.6822	67.7165
	23.7041	79.5276