

Estimation of infrastructure performance models using state-space specifications of time series models

Chih-Yuan Chu ^a, Pablo L. Durango-Cohen ^{b,*}

^a Graduate Student Researcher, Department of Civil and Environmental Engineering and Transportation Center, Northwestern University, 2145 Sheridan Road, A316, Evanston, IL 60208-3109, USA

^b Department of Civil and Environmental Engineering and Transportation Center, Northwestern University, 2145 Sheridan Road, A335, Evanston, IL 60208-3109, USA

Received 9 December 2005; received in revised form 28 November 2006; accepted 29 November 2006

Abstract

We consider state-space specifications of autoregressive moving average models (ARMA) and structural time series models as a framework to formulate and estimate inspection and deterioration models for transportation infrastructure facilities. The framework provides a rigorous approach to exploit the abundance and breadth of condition data generated by advanced inspection technologies. From a managerial perspective, the framework is attractive because the ensuing models can be used to forecast infrastructure condition in a manner that is useful to support maintenance and repair optimization, and thus they constitute an alternative to Markovian transition probabilities. To illustrate the methodology, we develop performance models for asphalt pavements. Pressure and deflection measurements generated by pressure sensors and a falling weight deflectometer, respectively, are represented as manifestations of the pavement's elasticity/load-bearing capacity. The numerical results highlight the advantages of the two classes of models; that is, ARMA models have superior data-fitting capabilities, while structural time series models are parsimonious and provide a framework to identify components, such as trend, seasonality and random errors. We use the numerical examples to show how the framework can accommodate missing values, and also to discuss how the results can be used to evaluate and select between inspection technologies.

© 2006 Elsevier Ltd. All rights reserved.

Keywords: Infrastructure performance modeling; State-space models; Time series analysis; Advanced inspection technologies; Missing data; Maintenance optimization

1. Introduction

Resource allocation decisions concerning the design, preservation and improvement of transportation infrastructure facilities are evaluated and selected based on their long-term economic consequences. The

* Corresponding author. Tel.: +1 847 491 4008; fax: +1 847 491 4011.

E-mail addresses: jameschu@northwestern.edu (C.-Y. Chu), pdc@northwestern.edu (P.L. Durango-Cohen).

evaluation involves processing data related to current condition¹ and using them to forecast the effect of these decisions on future condition. The economic consequences are then estimated by assuming a correspondence between infrastructure condition and costs. Condition forecasts are generated with performance models, which in this paper correspond to statistical expressions that relate condition to a set of explanatory variables such as design characteristics, traffic loading, environmental factors, and history of maintenance activities. The social and economic importance of the aforementioned decisions has, over the last 40 years, served as motivation for the development of numerous performance models.

Performance models are estimated using data from panels of facilities. Panel data consist of two components: cross-section data describing the differences between the facilities that comprise the panel, i.e., heterogeneity, and time series data describing the evolution of individual facilities over time. Most data sources are unbalanced in that they typically have extensive cross-section data and limited time series data. For example, Madanat et al. (1997) develop a bridge-deck performance model using data collected between 1978 and 1988 for 2602 bridges in the state of Indiana. More than 80% of the bridges were inspected three or fewer times. No bridge in the data set was inspected more than seven times.² To exploit this structure, existing performance modeling approaches are static, meaning that condition is represented as a function of contemporaneous explanatory variables and random error terms that describe the cross-sectional differences in the panel.

The motivation for the work described herein is that technologies such as sensors, laser or ultrasonic probes, and satellite-imaging, are becoming increasingly useful for collecting condition data and the factors that cause their deterioration. In the performance modeling context, advanced technologies permit frequent and comprehensive inspections, and therefore, are capable of providing long and detailed historical data for individual facilities, i.e., time series data. This serves as motivation for our work, which consists of the development of dynamic performance models that are capable of exploiting the information contained in the sequence of observations/measurements for individual facilities.

Another relevant characteristic of advanced inspection technologies is that they are capable of simultaneously evaluating and measuring multiple factors and distresses, thus, providing a complete and complex picture of a facility's condition. The work described herein builds on the framework of Ben-Akiva and Ramaswamy (1993), who introduce state-space models to rigorously address the problem of forecasting condition when multiple technologies are used to simultaneously collect various types of condition data. The key feature of the approach is that a facility's condition is represented by latent/unobservable variables that capture the ambiguity that exists in defining, and consequently in measuring infrastructure condition. State-space models are used to capture and estimate simultaneous processes, i.e., inspection, deterioration and maintenance. In particular, measurements are related to the latent condition through a *measurement model* that accounts for systematic and random errors in the inspection process, as well as for the relationships between different technologies and measurements. Latent performance models also include a *structural model* that describes the relationship between a set of explanatory variables and infrastructure condition. Empirical studies by Ben-Akiva and Ramaswamy (1993) and by Ben-Akiva and Gopinath (1995) have shown that latent performance models are appropriate to generate condition forecasts of transportation infrastructure, i.e., the goodness-of-fit measures are better than those reported using other statistical methods.

A salient feature of state-space specifications of time series models is that they provide a statistically rigorous approach to estimate the parameters that drive the maintenance and repair (M&R) optimization model of Durango-Cohen (2006). This means that in addition to providing a different approach for performance modeling, the proposed framework is attractive (from a managerial perspective) because the ensuing models can be used to support the allocation of resources for the preservation of transportation facilities. It also

¹ Information about current infrastructure condition is obtained by evaluating and measuring surface distresses and structural properties. Examples of surface distresses on transportation infrastructure include rut depth, type and extent of cracking, and extent of surface patching, on pavements; and cracking, spalling, and chloride contamination on bridge decks. Deflection and strain are manifestations of a pavement's elasticity/load-bearing capacity.

² Seminal data collection efforts for (in-service) pavements are described in Paterson (1987). Overall, their structure is similar to that of the bridge data set in Indiana.

means that the work described herein can be compared to the estimation of Markovian transition probabilities³ that are used to drive M&R optimization models formulated as (latent) Markov Decision Processes (cf. Madanat and Ben-Akiva (1994)). Indeed, the framework presented in Durango-Cohen (2006) addresses critical computational limitations that make the latent MDP impractical when multiple technologies are used simultaneously to measure (different) indicators of condition. These limitations follow from the fact that the state variables used to represent the measurements in MDP formulations are defined over discrete and ordinal sets. In addition to the computational problems, this assumption is unattractive as performance indicators are increasingly measured by automated inspection methods, which are capable of providing precise continuous readings (as opposed to ratings typically provided by human inspectors).

This paper complements our earlier work by providing a general class of statistical models that can be used to represent the evolution of infrastructure condition, and that can be included in the aforementioned framework to support M&R decision-making. Specifically, we describe and compare two classes of time series models: Autoregressive Moving Average (ARMA) models and structural time series models. Through an empirical study, we illustrate how the proposed methodology can be used to process (complete or incomplete) condition data gathered simultaneously using advanced technologies. The data consist of deflection and pressure measurements for an asphalt pavement located on a closed-loop test track that is run by the Minnesota Road Research Program, MnROAD, the Road Research Division of the Minnesota Department of Transportation. The results highlight the advantages of each of the two types of models. ARMA models provide superior data-fitting capabilities, and structural time series provide parsimonious models that enable the identification of meaningful data components, e.g., trend, seasonality and random errors. We also use the numerical results to illustrate how the proposed framework can be used to evaluate the capabilities of various (combinations of) inspection technologies.

The remainder of the paper is organized as follows. Section 2 reviews state-space specifications of time series models, and describes the two classes of models considered in this paper: ARMA and structural time series models. Section 3 provides an overview of the approach that we use for model estimation, selection and diagnosis. We also describe important details about our implementation. Section 4 details the empirical study used to illustrate the methodology. A summary of the contributions of the paper, as well as our conclusions are presented in Section 5.

2. Time series models in state-space form

Here, we describe our approach to model infrastructure performance as a time series in state-space form. The general form of a state-space model is shown in Eqs. (1) and (2).

$$X_{t+1} = g_t X_t + h_t A_t + \epsilon_{t+1} \quad (1)$$

$$Z_t = A_t X_t + \xi_t \quad (2)$$

where for periods $t = 1, 2, \dots, T$ it is assumed that

$$E(\epsilon_t) = 0 \quad (3)$$

$$\text{Var}(\epsilon_t) = \Sigma_\epsilon \quad (4)$$

$$E(\xi_t) = 0 \quad (5)$$

$$\text{Var}(\xi_t) = \Sigma_\xi \quad (6)$$

$$E(\epsilon_t \xi_t) = 0 \quad (7)$$

where the variables, parameters and random error terms in the model are as follows:

X_t : $d \times 1$ vector used to represent the system's/facility's state/condition at the start of period t . As in Ben-Akiva and Gopinath (1995) the vector might, for example, include a component to represent

³ For a review of this literature, the reader is referred to Mishalani and Madanat (2002) and the references therein.

functional performance and a component to represent structural fitness. The additional components in the vector can also be used to include lagged dependent variables to account for the effect of the history of the process on the system's deterioration.

- A_t : $m \times 1$ vector of exogenous factors that may include maintenance activities, environmental factors, traffic loading, etc.
- Z_t : (with components denoted $z_t^{(i)}$) is a $k \times 1$ vector used to represent the distress measurements.
- g_t, h_t : respectively, contain the parameters that describe the effect of the state vectors and explanatory variables on the state vector. Their dimensions are $d \times d$ and $d \times m$.
- A_t : is a $k \times d$ matrix describing the relationship between measurements and latent condition.
- ϵ_t, ξ_t : represent random error terms. They are assumed to follow Normal Distributions with finite second moments. Σ_ϵ and Σ_ξ are used to denote their covariance matrices. The error terms and covariance matrices are of dimensions $d \times 1$, $k \times 1$, $d \times d$ and $k \times k$, respectively. They satisfy the properties presented in Eqs. (3)–(7).

Eq. (1) is the system equation and Eq. (2) is the measurement error model. Eq. (1) governs system's dynamics, and captures the effect of explanatory variables. The components of the equation depend on the nature of the system, and on the formulation that is adopted. The components considered herein are discussed in detail both below, as well as in the subsequent section. For practical reasons, Σ_ϵ is frequently assumed diagonal meaning that the elements in the state vector are independent.

Eq. (2) captures systematic and random errors in the inspection process. Often distresses corresponding to different physical characteristics are collected, and therefore, the number of measurements is larger than the number of latent condition variables (i.e. $k > d$). In such situations, the measurement error model captures the relationships between the measurements and accounts for the fact that they are imperfect surrogates of the latent condition variables. To reduce the number of parameters that require estimation, we follow [Humplick \(1992\)](#) where the measurement errors are attributed to the inspection technologies even though, in general, they could also depend on factors such as the facility being inspected, the equipment operator or the nature and magnitude of the underlying distress. If multiple measurements are used, the correlations between the distress measurement errors are captured by the covariance matrix of ξ_t, Σ_ξ . Σ_ξ could be either diagonal or non-diagonal depending on the properties of inspection technologies and their errors.

As stated above, the structure of general state-space models depends on the assumptions that are used to specify the state vector components. In this paper, we consider two approaches leading to two classes of time series models: ARMA models and structural time series models. In ARMA models, the state vector is expressed as a function of lagged dependent variables and random error terms. The objective in this type of formulation is to obtain a model that provides good fit-to-data. In structural time series models, the state vector is decomposed into components representing meaningful data structures such as trend, seasonality, random errors, etc. Lagged dependent variables can also be included, making structural time series more general than ARMA models. In contrast, to ARMA models, the specification of structural time series models requires judgments about the structures that are present in the sequence of observations. Additional information about the two classes of models is provided in the following subsections.

2.1. Formulation of ARMA(p, q) models

An ARMA(p, q) model in state-space form can be represented as follows:

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \cdots + \phi_p x_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \cdots + \theta_q \epsilon_{t-q}, \quad t = 1, 2, \dots, T \quad (8)$$

$$z_t^{(i)} = \lambda_i x_t + \xi_t^{(i)}, \quad i = 1, 2, \dots, k, \quad t = 1, 2, \dots, T \quad (9)$$

Eq. (8) is the system equation and Eq. (9) is the measurement error model. We let $d \equiv \max\{p, q + 1\}$, $\phi_i = 0$ if $i > p$, and $\theta_j = 0$ if $j > q$. As shown in [Durbin and Koopman \(2001\)](#), these definitions can be used to rewrite the above model in the general state-space form as is shown in Eqs. (10)–(12). To simplify the presentation, we exclude the explanatory variables

$$X_t = \begin{bmatrix} x_t \\ x_t^{(2)} \\ x_t^{(3)} \\ \vdots \\ x_t^{(d-1)} \\ x_t^{(d)} \end{bmatrix} = \begin{bmatrix} \phi_1 & 1 & 0 & \cdots & 0 & 0 \\ \phi_2 & 0 & 1 & \cdots & 0 & 0 \\ \phi_3 & 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \phi_{d-1} & 0 & 0 & \cdots & 0 & 1 \\ \phi_d & 0 & 0 & \cdots & 0 & 0 \end{bmatrix} \begin{bmatrix} x_{t-1} \\ x_{t-1}^{(2)} \\ x_{t-1}^{(3)} \\ \vdots \\ x_{t-1}^{(d-1)} \\ x_{t-1}^{(d)} \end{bmatrix} + \begin{bmatrix} 1 \\ \theta_1 \\ \theta_2 \\ \vdots \\ \theta_{d-2} \\ \theta_{d-1} \end{bmatrix} \epsilon_t, \tag{10}$$

where

$$\Sigma_\epsilon = \begin{bmatrix} \sigma_\epsilon^2 & 0 & 0 & \cdots & 0 & 0 \\ 0 & \theta_1^2 \sigma_\epsilon^2 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \theta_2^2 \sigma_\epsilon^2 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \theta_{d-2}^2 \sigma_\epsilon^2 & 0 \\ 0 & 0 & 0 & \cdots & 0 & \theta_{d-1}^2 \sigma_\epsilon^2 \end{bmatrix},$$

$$Z_t = \begin{bmatrix} z_t^{(1)} \\ z_t^{(2)} \\ \vdots \\ z_t^{(k-1)} \\ z_t^{(k)} \end{bmatrix} = \begin{bmatrix} \lambda_1 & 0 & 0 & \cdots & 0 & 0 \\ \lambda_2 & 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \lambda_{k-1} & 0 & 0 & \cdots & 0 & 0 \\ \lambda_k & 0 & 0 & \cdots & 0 & 0 \end{bmatrix} X_t + \begin{bmatrix} \zeta_t^{(1)} \\ \zeta_t^{(2)} \\ \vdots \\ \zeta_t^{(k-1)} \\ \zeta_t^{(k)} \end{bmatrix}, \tag{11}$$

where

$$\Sigma_\zeta = \begin{bmatrix} \sigma_{\zeta_t^{(1)}}^2 & \sigma_{\zeta_t^{(1)} \zeta_t^{(2)}} & \sigma_{\zeta_t^{(1)} \zeta_t^{(3)}} & \cdots & \sigma_{\zeta_t^{(1)} \zeta_t^{(k-1)}} & \sigma_{\zeta_t^{(1)} \zeta_t^{(k)}} \\ \sigma_{\zeta_t^{(2)} \zeta_t^{(1)}} & \sigma_{\zeta_t^{(2)}}^2 & \sigma_{\zeta_t^{(2)} \zeta_t^{(3)}} & \cdots & \sigma_{\zeta_t^{(2)} \zeta_t^{(k-1)}} & \sigma_{\zeta_t^{(2)} \zeta_t^{(k)}} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \sigma_{\zeta_t^{(k-1)} \zeta_t^{(1)}} & \sigma_{\zeta_t^{(k-1)} \zeta_t^{(2)}} & \sigma_{\zeta_t^{(k-1)} \zeta_t^{(3)}} & \cdots & \sigma_{\zeta_t^{(k-1)}}^2 & \sigma_{\zeta_t^{(k-1)} \zeta_t^{(k)}} \\ \sigma_{\zeta_t^{(k)} \zeta_t^{(1)}} & \sigma_{\zeta_t^{(k)} \zeta_t^{(2)}} & \sigma_{\zeta_t^{(k)} \zeta_t^{(3)}} & \cdots & \sigma_{\zeta_t^{(k)} \zeta_t^{(k-1)}} & \sigma_{\zeta_t^{(k)}}^2 \end{bmatrix}$$

$$z_t^{(1)} = [\lambda_1 \ 0 \ 0 \ 0 \ 0 \ \cdots \ 0] X_t + \zeta_t^{(1)}. \tag{12}$$

The first element in the state vector, x_t , represents the true facility condition at time t . Eq. (11) is the measurement error model for k distress measurements (Eq. (12) is for the special case of a single measurement). As discussed above, the structure of Σ_ζ depends on the assumptions regarding the correlations between inspection technologies. The parameters to be estimated are $\phi_1, \dots, \phi_d, \theta_1, \dots, \theta_{d-1}, \sigma_\epsilon^2, \lambda_1, \dots, \lambda_k, \Sigma_\zeta$. In the estimation process, the true condition is generally set to be equal to one of the condition measurements, a reference measurement. This is done by setting $\lambda_i = 1$ for the reference measurement. As stated earlier, ϵ_t and ζ_t are assumed to follow Normal Distributions with finite second moments. Without loss of generality, we assume they have zero means.

2.2. Formulation of structural time series models

This section follows the presentation in Harvey (1990) and Durbin and Koopman (2001). Structural time series models consist of components that capture (deterministic or stochastic) trend, seasonality, and random error. We use μ_t and β_t to denote trend and slope at time t . Similar to intercept and slope terms in regression models, trend and slope components represent the long-term movements of the series. The relationship between trend and slope is described in Eqs. (13) and (14). The existence of η_t and ζ_t indicate that both trend and slope could be stochastic and change over time. Note that when η_t and ζ_t are set to zero, the model exhibits a deterministic linear trend, that is, $\mu_t = \alpha + \beta t$, where α and β are constant coefficients. γ_t is the seasonal component at time t . As shown in Eq. (15), seasonality components are formulated such that their sum for a year equals zero for deterministic cases, or a

random error ω_t for stochastic cases. Lagged dependent variables up to order p can also be included to capture the dependence on the history of the process. This gives Eq. (16) describing facility condition as a function of lagged dependent variables up to order p , trend, seasonality, and random error. This equation has the unattractive feature that the state-vector, X_t , is expressed as a function of the contemporaneous arguments, μ_t and γ_t . This is easily fixed by replacing the results of Eqs. (13)–(15) into Eq. (16) to obtain Eq. (17). The special case with no lagged dependent variables in the model is obtained by setting $\phi_1 = \phi_2 = \dots = \phi_p = 0$ for periods $t = 1, \dots, T$. The measurement error equation of structural time series with k measurements is identical to Eq. (9).

Trend: $\mu_t = \mu_{t-1} + \beta_{t-1} + \eta_t$ (13)

Slope: $\beta_t = \beta_{t-1} + \zeta_t$ (14)

Seasonality: $\sum_{j=0}^{s-1} \gamma_{t-j} = \omega_t$ (15)

System equation:

$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \dots + \phi_p x_{t-p} + \mu_t + \gamma_t + \epsilon_t, \quad t = p + 1, \dots, T$ (16)

$x_t = \phi_1 x_{t-1} + \dots + \phi_p x_{t-p} + (\mu_{t-1} + \beta_{t-1} + \eta_t) + (\omega_t - \gamma_{t-1} - \gamma_{t-2} - \dots - \gamma_{t-s+1}) + \epsilon_t$ (17)

In Eq. (18), we rewrite the system equation in the form of the general state-space model. $\epsilon_t, \eta_t, \zeta_t$, and ω_t , are assumed independent Normally distributed random variables with zero mean and finite variance. To simplify the presentation, we exclude the explanatory variables. As with ARMA models, the latent condition is generally set to be equal to a reference measurement. Thus, the parameters to be estimated in structural time series formulations are $\phi_1, \dots, \phi_p, \sigma_\epsilon^2, \sigma_\eta^2, \sigma_\zeta^2, \sigma_\omega^2, \lambda_1, \dots, \lambda_k, \Sigma_\xi$.

$$X_t = \begin{bmatrix} x_t \\ x_{t-1} \\ \vdots \\ x_{t-p+1} \\ \mu_t \\ \beta_t \\ \gamma_t \\ \gamma_{t-1} \\ \gamma_{t-2} \\ \vdots \\ \gamma_{t-s+2} \end{bmatrix} = \begin{bmatrix} \phi_1 & \phi_2 & \dots & \phi_p & 1 & 1 & -1 & -1 & -1 & \dots & -1 \\ 1 & 0 & \dots & 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 \\ \hline & & & & 1 & 1 & & & & & \\ \mathbf{0} & & & & 0 & 1 & & & & & \mathbf{0} \\ \hline & & & & & & -1 & -1 & -1 & \dots & -1 \\ & & & & & & 1 & 0 & 0 & \dots & 0 \\ & & \mathbf{0} & & & & 0 & 1 & 0 & \dots & 0 \\ & & & & & & \vdots & \vdots & \vdots & \ddots & \vdots \\ & & & & & & 0 & 0 & 0 & \ddots & 0 \end{bmatrix} \begin{bmatrix} x_{t-1} \\ x_{t-2} \\ \vdots \\ x_{t-p} \\ \mu_{t-1} \\ \beta_{t-1} \\ \gamma_{t-1} \\ \gamma_{t-2} \\ \gamma_{t-3} \\ \vdots \\ \gamma_{t-s+1} \end{bmatrix} + \begin{bmatrix} \eta_t + \omega_t + \epsilon_t \\ 0 \\ \vdots \\ 0 \\ \eta_t \\ \zeta_t \\ \omega_t \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix},$$

$$\text{where } \Sigma_\epsilon = \begin{bmatrix} \sigma_\eta^2 + \sigma_\omega^2 + \sigma_\epsilon^2 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \\ \hline & & & \sigma_\eta^2 & 0 \\ & & & 0 & \sigma_\zeta^2 \\ \hline & & & & & \sigma_\omega & 0 & 0 & \dots & 0 \\ & & & & & 0 & 0 & 0 & \dots & 0 \\ & & \mathbf{0} & & & 0 & 0 & 0 & \dots & 0 \\ & & & & & \vdots & \vdots & \vdots & \ddots & \vdots \\ & & & & & 0 & 0 & 0 & \ddots & 0 \end{bmatrix}.$$

To conclude this subsection, we discuss the relationship between the formulations considered herein and ARIMA (autoregressive integrated moving average) models developed by Box and Jenkins, which is the most common approach for the analysis of time series. The Box–Jenkins approach is based on the assumption of stationary data. For non-stationary data, differencing techniques must be applied to obtain stationary data. However, the elimination of trends or seasonality by differencing may be a drawback when they are of interest, as may be the case when modeling the performance of transportation facilities. Structural time series models can be reduced to ARIMA models. For example, a BSM with monthly seasonal terms can be reduced to MA(13) after first and 12th order differencing (in order to remove trend and monthly seasonality). We emphasize that this reduction is not attractive because the feature of structural time series models is the meaningful components and these components will not exist for observation after the data are transformed to obtain an ARIMA model. Further details are presented in [Harvey \(1990\)](#).

3. Model estimation, selection and diagnosis

One of the advantages of presenting different types of time series models as instances of the general state-space form is that it simplifies the presentation and implementation of the procedure to estimate the model parameters. This section outlines the procedure we implemented. One of the critical steps involves evaluating single-step prediction errors and their variances, for which we implement the square root filter algorithm. Our implementation of the square root filter handles missing observations/measurements and is presented in [Appendix A](#). Finally, we describe the approach that we use to evaluate, select and diagnose models for statistical adequacy.

3.1. Loglikelihood function

We adopt Maximum Likelihood Estimation to estimate the parameters due to the asymptotic properties of the ensuing estimates. In time series the observations are assumed to follow a joint distribution, $f(X_1, X_2, \dots, X_{T-1}, X_T)$. The assumption required for MLE to estimate time series models is that ξ_t and ϵ_t follow Normal distributions. It follows that the conditional distribution of X_t given data up to $t - 1$ is also Normal. Based on the above assumption, the loglikelihood function of a state-space model can be formulated as Eq. (19). A detailed derivation can be found in [Harvey \(1990\)](#). We maximize the loglikelihood function with the non-linear unconstrained optimization routine in MATLAB. As shown in the equation, the evaluation of the function depends on values of V_t and F_t . These values are generated by the square root filter described below. The reason that the calculation starts at time $d + 1$ is explained in [Appendix A](#).

$$\log \mathcal{L} = -\frac{T-d}{2} \log 2\pi - \frac{1}{2} \sum_{t=1+d}^T \log \det(F_t) - \frac{1}{2} \sum_{t=1+d}^T V_t' F_t^{-1} V_t, \quad (19)$$

where V_t is the one step prediction error at time t , F_t is the prediction error variance (PEV) at time t and d is the cardinality of the state vector.

The square root filter is a variant of Kalman filter, which is an algorithm to calculate the optimal estimators of the state vector for state-space specifications, and also to compute V_t and F_t in the loglikelihood function. However, covariance matrices are not guaranteed positive-definite in the standard Kalman Filter. This, in addition to numerical instability and rounding errors that appear in the standard Kalman Filter creates difficulties both in evaluating and maximizing the loglikelihood function. These disadvantages motivate the implementation of the square root filter. Instead of using the original covariance matrices, the square root filter decomposes them into lower-triangular matrices (square root of covariance matrices), and recursively computes the square roots of matrices. This transformation not only guarantees positive definite covariance matrices, but also provides more significant digits for calculation. Therefore, it is considered numerically stable and robust. The disadvantage of this algorithm is the additional computational effort. For the details of the algorithm, readers are referred to [Morf and Kailath \(1975\)](#). A modification of the algorithm that can handle missing observations is presented in [Appendix A](#).

3.2. Model selection and diagnosis

Generally, the prediction error decreases as the number of model parameters increases. To avoid “overfitting” in the process of selecting adequate model specifications in the empirical study, we use Akaike’s Information Criterion (AIC) in Eq. (20) to measure the trade-off between prediction error and number of explanatory variables. Smaller values of AIC are preferred

$$\text{AIC} = \frac{-2 \times \log \mathcal{L} + 2 \times (n + d)}{T}, \quad (20)$$

where n represents the number of parameters in the model.

We note that autocorrelation and partial autocorrelation functions are commonly used for selecting between ARMA models; however, this type of analysis does not apply to time series models in state-space form. Instead, we propose to use backward elimination, where models are refined by removing the (statistically) insignificant parameters from the model specification. This technique is used in the empirical study presented in Section 4.

In addition to AIC, which examines data-fitting, models are further diagnosed by generating in-sample and out-of-sample predictions. In-sample predictions are made from $t = d + 1$ to $t = T$ and compared with the observations. This task studies a model’s capability of reproducing observations and capturing the underlying mechanism generating the data sequence. Out-of-sample predictions start from $t = T + 1$ and the length of prediction is arbitrary. This is the most powerful test for a time series model. Generally, the same length of observations are not used for estimation in order to compare out-of-sample prediction and actual data; however, due to data limitations, we decided not to use this technique in the empirical study.

After a model is selected, we further diagnose the model by examining the residuals, i.e., the differences between one-step predictions and observations. We conduct the residual analysis to check that serial correlations of the residuals and these residuals are expected to be independent and identically distributed.

4. Empirical study

In this section, we use the methodology described in the previous section to estimate performance models for asphalt pavements. The data for the study consist of deflection and pressure measurements from cell 33, a Superpave test cell located on a closed-loop test track that is run by MnROAD, the road research division of the Minnesota DOT. Superpave (SUPERior PERforming Asphalt PAVEMENTS) refers to the criteria used for designing and building hot-mix asphalt (HMA). These pavements are designed for performance under extreme environmental conditions and heavy truck loadings. Additional details are available in [Palmquist et al. \(2002\)](#) and [Zerfas \(2003\)](#). The design of the cell is 4-in. HMA, 12-inch class-6 base, and clay subgrade. The asphalt binder used is polymer modified PG 58-34. The construction finished in September 2, 1999 and the traffic loading started on September 30, which implies that the cell was almost new in the beginning of data collection.

Two measurements, deflection and pressure, are chosen to illustrate the methodology. The descriptions of these two measurements appears below. Measurements such as rut depth and roughness are more commonly used in pavement performance modeling. However, as stated earlier time series modeling requires long, detailed historical data. Our choice reflects the fact that deflection and pressure are the most complete measurements in MnROAD database. Both measurements were collected in the four year period from November 1999 ($t = 1$) through October 2003 ($t = T = 48$).

4.1. Deflection

Pavement surface deflection is the primary means of evaluating a flexible pavement’s structural fitness/load-bearing capacity. It is the most commonly used indicator of structural condition by transportation agencies [Gramling, 1994](#). [Golabi and Pereira \(2003\)](#), for example, selected deflection as performance indicators in the pavement management system they developed for Portugal. Deflection is measured as a pavement surface’s vertical deflected distance as a result of an applied (static or dynamic) load. It is collected by impact/impulse load response, where an impact load device delivers a transient impulse/load to the pavement surface.

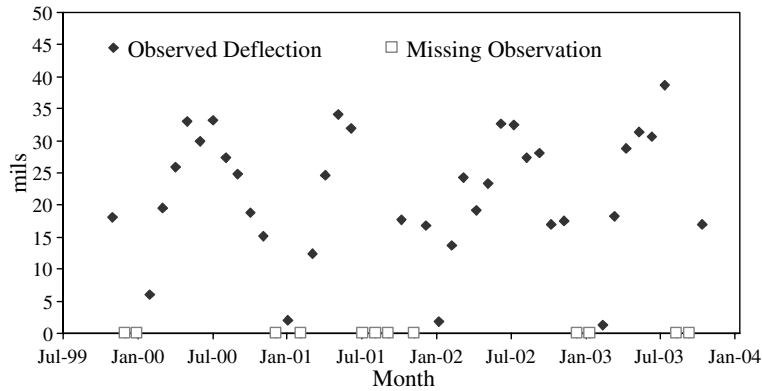


Fig. 1. Monthly average deflection measurements.

The induced pavement response (deflection basin) is measured by a series of sensors. The most common type of tool used to induce the load is a falling weight deflectometer (FWD).

Deflection can be used to estimate the modulus of elasticity of pavement layers, as well as the effective structural number, which represents the remaining structural capacity of the pavement section. Deflection measurements are relatively low during the winter and high during the summer. Thus, in practice, deflection measurements are adjusted to a single reference temperature before use in order to control for seasonal effects [AASHTO, 1993](#). In this empirical study, we use the unadjusted values in order to explore the capability of the proposed specifications to capture seasonal patterns.

The data used in our analysis consist of the monthly average deflection measurements (in thousandths of an inch – *mils*) induced by a FWD. The data are plotted in [Fig. 1](#). The frequency of data has important effect on model formulation and will be explained later. Considering that maintenance policies of transportation infrastructure are hardly made more frequently than monthly, we adopt monthly average data in the empirical study. However, the methodology is capable of different frequencies of data; in fact, data availability is more likely to constrain the analysis. Twelve entries in the data set were missing and are represented as empty squares in the figure.

4.2. Pressure

Vertical pressure data are used to determine the vertical stress distribution in the base and subgrade layers of a pavement. Pressure measurements in MnROAD are collected by inducing a load generated by a truck (that has five axles). The pressure measurements (in *millivolts*) are collected with a *dynamic soil pressure cell sensor* set on top of the subgrade. Pressure measurements can be used to calculate strain measurements

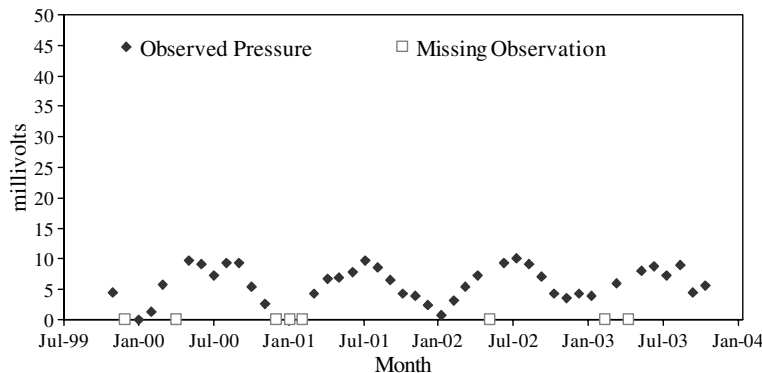


Fig. 2. Monthly average pressure measurements.

approximately via modulus of elasticity. Moreover, since strain is commonly used to represent pavement structural strength, pressure data constitute another important indicator of a pavement’s structural performance. Like deflection, stress is affected by temperature and an adjustment is commonly used in practice. Another reason for choosing the measurement is that the sensor is still fully functional after four years of operation while many other sensors installed in the pavement have failed. Although the data collected by sensors are automatic and frequent, the nature of pavement facility makes the replacement of failed sensors very difficult.

In our study, we use the monthly average of the pressure measurements induced by the first axle of loaded trucks. The data are presented in Fig. 2. Eight missing entries are represented as empty squares in the figure.

4.3. Estimation results

4.3.1. ARMA time series models

The data in Figs. 1 and 2 exhibit clear seasonality; therefore, it is necessary to consider ARMA models of order 13 or higher for the undifferenced monthly data. To simplify the estimation of the model and the interpretation of the results, we consider AR models instead of ARMA models in the analysis. We set deflection to be the reference measurement, henceforth denoted with subscript 1. We use backward elimination principle to refine the resulting models. The backward elimination principle is to start the estimation with a higher order and eliminate the insignificant parameters to obtain reasonable models.

Several models were considered and the best results were obtained for an AR model of order 13. To illustrate the model selection process, we begin by estimating the general AR(13) model, labeled AR-1. The least

Table 1
Estimation results for ARMA models with two distress measurements

Parameters	AR-1		AR-6		AR-7	
	Est.	t-Stat.	Est.	t-Stat.	Est.	t-Stat.
σ_{ξ}^2	-7.26E-04	-3.13E-03	-1.96E-06	-5.50E-06	1.08E-06	2.46E-06
$\sigma_{\xi}^{(l)}$	-3.98	-6.79	3.62	6.37	3.57	6.43
$\sigma_{\xi}^{(m)}$	0.52	3.75	-0.37	-1.83	-0.38	-1.81
$\sigma_{\xi}^{(mm)}$	-0.45	-5.37	0.70	6.66	0.74	6.75
ϕ_1	-0.28	-1.49	-	-	-	-
ϕ_2	0.33	1.95	-0.18	-3.06	-0.16	-3.09
ϕ_3	-0.26	-2.29	0.34	3.55	0.39	4.15
ϕ_4	0.69	3.30	0.06	1.61	-	-
ϕ_5	-0.16	-0.84	-	-	-	-
ϕ_6	-1.10	-6.28	-0.80	-13.20	-0.83	-14.45
ϕ_7	0.71	2.86	0.51	5.59	0.55	7.82
ϕ_8	-0.03	-0.14	-	-	-	-
ϕ_9	0.09	0.50	0.27	4.13	0.28	4.07
ϕ_{10}	0.55	2.85	-	-	-	-
ϕ_{11}	-0.28	-2.19	0.18	3.94	0.15	3.77
ϕ_{12}	0.22	2.33	-	-	-	-
ϕ_{13}	0.54	4.03	0.62	8.32	0.62	7.89
λ_2	0.26	21.08	0.26	22.66	0.27	22.58
LL	-112.71		-116.66		-117.90	
AIC	5.99		5.94		5.95	
n	18		13		12	

σ_{ξ}^2 represents the variance of the system equation. The covariance matrix of measurement errors are defined by $\Sigma_{\xi} = \begin{bmatrix} (\sigma_{\xi^{(1)}})^2 & \sigma_{\xi^{(1)}\xi^{(2)}} \\ \sigma_{\xi^{(2)}\xi^{(1)}} & (\sigma_{\xi^{(2)}})^2 \end{bmatrix} = \Sigma_{\xi}^L \Sigma_{\xi}^{L'} = \begin{bmatrix} (\sigma_{\xi}^{(l)})^2 & \sigma_{\xi}^{(l)} \sigma_{\xi}^{(m)} \\ \sigma_{\xi}^{(l)} \sigma_{\xi}^{(m)} & (\sigma_{\xi}^{(m)})^2 + (\sigma_{\xi}^{(mm)})^2 \end{bmatrix}$, where $\Sigma_{\xi}^L = \begin{bmatrix} \sigma_{\xi}^{(l)} & 0 \\ \sigma_{\xi}^{(m)} & \sigma_{\xi}^{(mm)} \end{bmatrix}$. As defined throughout the empirical study, the first measurement is deflection and the second is pressure.

Bold entries indicate the parameters are insignificant at 95% level of confidence. The AICs of AR-2 to AR-5 are 5.97, 6.05, 6.01, and 5.97, respectively.

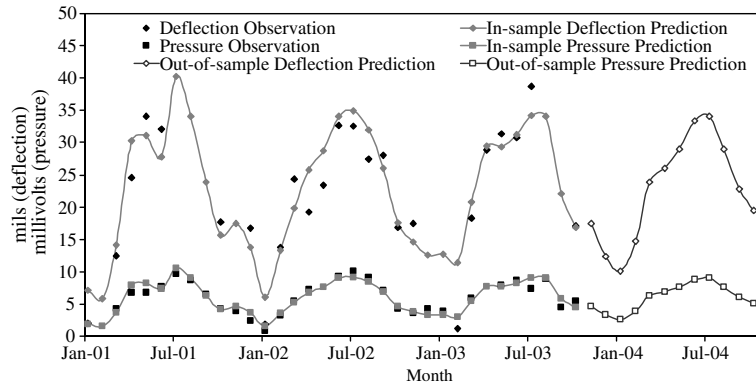


Fig. 3. In-sample and out-of-sample prediction for AR-6 model.

significant ϕ parameter is set to zero and then AR-2 is estimated (t -statistics with magnitude greater than 2 indicate that the associated parameter is significant at the 95% confidence level). AR-3 through AR-7 are generated in the same way. Since all of the ϕ parameters are significant, we terminate the elimination process with AR-7. The parameter estimates and goodness-of-fit measures for all variations are presented in Table 1. We observe that AR-6 results in the best AIC. In other words, AR-6 has the best tradeoff of the number of parameters and model fit. The standardized residuals of both measurements from AR-6 model are further tested using sample autocorrelation function. Sample autocorrelation function shows the residuals are independent and identically distributed with $N(0,1/T)$ since over 95% of the function values are within $\pm 1.96\sqrt{T}$. Hence, the preferred AR-6 is statistically satisfactory.

The final model is presented in Eqs. (21) and (22). Eq. (21) represents the evolution of the pavement over time. The fact that the variance of error term is negligible indicates that all randomness in the data is attributed to the inspection technologies. Eq. (22) is the measurement error model. $z^{(1)}$ is the deflection measurement and $z^{(2)}$ is the pressure measurement. ξ is the measurement error vector of the two technologies and Σ_ξ is the covariance. From the equation, the measurement error of deflection is much greater than pressure, which also can be seen from Fig. 3. Since the two measurements have different magnitudes, variance of measurement is not sufficient to compare technologies and we further use coefficient of determination to choose the better technology. The result shows that R^2 of deflection is 0.83 while R^2 of pressure measurements is 0.89. It means that pressure sensor not only has smaller measurement error but provides better fit than FWD. This information might be useful for technology selection as suggested in Ben-Akiva and Ramaswamy (1993).

$$x_t = -0.18x_{t-2} + 0.34x_{t-3} + 0.06x_{t-4} - 0.80x_{t-6} + 0.51x_{t-7} + 0.27x_{t-9} + 0.18x_{t-11} + 0.62x_{t-13} + \epsilon_t, \epsilon_t \sim N(0, 3.85E - 12) \tag{21}$$

$$\begin{bmatrix} z_t^{(1)} \\ z_t^{(2)} \end{bmatrix} = \begin{bmatrix} 1.00 \\ 0.26 \end{bmatrix} x_t + \begin{bmatrix} \xi_t^{(1)} \\ \xi_t^{(2)} \end{bmatrix}, \quad \Sigma_\xi = \begin{bmatrix} 13.14 & -1.33 \\ -1.33 & 0.62 \end{bmatrix} \tag{22}$$

Further, the model is examined by the 33-month prediction (from February 2001 ($t = 16$) to October 2003 ($t = 48$)). From Fig. 3, the model reproduce both observations adequately, which indicates that the model captures the data evolution well. In addition, 12-month out-of-sample prediction (from November 2003 to October 2004) is done. The length of prediction is chosen because M&R decisions are generally planned for on an annual basis. From the figure, the out-of-sample prediction is consistent with the history of data and the pattern seems reasonable.

4.3.2. Structural time series models

As described earlier, the data sets show clear seasonality; therefore, it is beneficial to adopt structural time series models. To identify the components of the data set, four structural time series models are estimated. The results are listed in Table 2. The first model, BSM-1, incorporates trend, seasonality, random error, one-step, and two-step lagged dependent variables. The estimation shows that the variances of trend, slope, and random

Table 2
Estimation results for basic structural model (BSM) with two-step lagged dependent variable

Parameters	BSM-1		BSM-2		BSM-3		BSM-4	
	Est.	<i>t</i> -Stat.	Est.	<i>t</i> -Stat.	Est.	<i>t</i> -Stat.	Est.	<i>t</i> -Stat.
σ_{ξ}^2	3.22E-11	5.98E-06	4.91E-08	-2.08E-04	-0.71	-1.57E-06	0.24	1.13E-03
σ_{η}^2	3.16E-13	-2.03E-06	–	–	1.27E-14	3.75E-07	0.01	0.02
σ_{ζ}^2	6.65E-14	-9.63E-06	–	–	4.62E-19	2.25E-08	5.78E-03	0.58
σ_{ω}^2	8.17E-12	4.86E-06	–	–	0.71	-0.61	4.87E-10	1.07E-05
$\sigma_{\xi}^{(l)}$	-3.59	-6.91	3.59	6.91	-3.57	-5.55	4.86	0.11
$\sigma_{\xi}^{(m)}$	-0.05	-0.12	0.05	0.12	-0.14	-0.34	0.77	0.05
$\sigma_{\xi}^{(m)}$	1.13	7.36	1.13	7.36	1.16	5.90	1.15	0.54
ϕ_1	0.93	8.34	0.93	8.34	0.18	1.15	–	–
ϕ_2	-0.67	-3.37	-0.67	-3.37	–	–	–	–
λ_2	0.26	22.43	0.26	22.43	0.26	22.72	0.26	24.24
LL	-130.66		-130.66		-132.73		-140.32	
AIC	6.49		6.36		6.53		6.80	
<i>n</i>	10		7		9		8	

σ_{ξ}^2 represents the variance of the system equation; σ_{η}^2 is the variance component of the trend equation; σ_{ζ}^2 is the variance component of the slope equation and σ_{ω}^2 is the variance component of the seasonality equation.

The covariance matrix of measurement errors are defined by $\Sigma_{\xi} = \begin{bmatrix} (\sigma_{\xi^{(1)}})^2 & \sigma_{\xi^{(1)}\xi^{(2)}} \\ \sigma_{\xi^{(2)}\xi^{(1)}} & (\sigma_{\xi^{(2)}})^2 \end{bmatrix} = \Sigma_{\xi}^L \Sigma_{\xi}^{L'} = \begin{bmatrix} (\sigma_{\xi^{(l)}})^2 & \sigma_{\xi^{(l)}\xi^{(m)}} \\ \sigma_{\xi^{(m)}\xi^{(l)}} & (\sigma_{\xi^{(m)}})^2 + (\sigma_{\xi^{(m)}})^2 \end{bmatrix}$, where $\Sigma_{\xi}^L = \begin{bmatrix} \sigma_{\xi^{(l)}} & 0 \\ \sigma_{\xi^{(m)}} & \sigma_{\xi^{(m)}} \end{bmatrix}$. As defined throughout the empirical study, the first measurement is deflection and the second is pressure.

Bold entries indicate the parameters are insignificant at 95% level of confidence.

errors are insignificant. Two approaches are used to refine BSM-1. The first approach is to replace stochastic trend and seasonality with deterministic ones. The results show that removing the stochastic components provides almost identical loglikelihood values. Thus, BSM-2 is preferred with a smaller number of parameters.

The second approach is to reduce the number of lagged dependent variables. The two-step lagged dependent variable is removed from BSM-1 and the one-step lagged dependent variable of BSM-3 turns out not to be significant. The fact that the one-step lagged dependent variable is not significant can be explained by the inclusion of the “slope” variable in the model, that is, their effect overlaps due to the linearity of the state space model. Also, BSM-3 is worse than BSM-1 in terms of goodness-of-fit. Further, the one-step lagged dependent variable is removed from BSM-3 and BSM-4 is estimated. Again, BSM-4 has worse AIC than other models. Comparing BSM-1 to BSM-3 and BSM-4 in Table 2, virtually all the variances of the random error terms are smaller, which indicates that BSM-1 captures the transition of the system better and the model itself becomes “less stochastic”. In other words, the lagged dependent variables in the model explain a larger part of the variance in the data.

In summary, BSM-2 is preferred among BSM models in terms of AIC. The beauty of the structural time series models is the meaningful components of the series, which are shown in Fig. 4. In the figure, trend, slope, and seasonality are clearly decomposed. The seasonality matches the seasonal change of the reference measurements reasonably. The information of seasonality can be extracted from actual data and external temperature adjustment factors are not required. Further, since differencing is not used, the component remains, whether it is of interest or not. However, the important component of a structural time series model, the trend, is harder to interpret from this data set. The negligible slope matches the fact that both the measurements are not increasing or decreasing except for the seasonal change. In other words, the section did not deteriorate over the duration of the experiment. This is explained by the fact that the pavement section was designed to withstand environmental and traffic loading extremes. Note that if structural deterioration of pavement were to exist, this specification would be able to detect it. Moreover, the residuals of both measurements from BSM-3 model are further tested using sample autocorrelation function. The test shows that the residuals are independent and identically distributed, which indicates that the preferred BSM-2 model is statistically satisfactory.

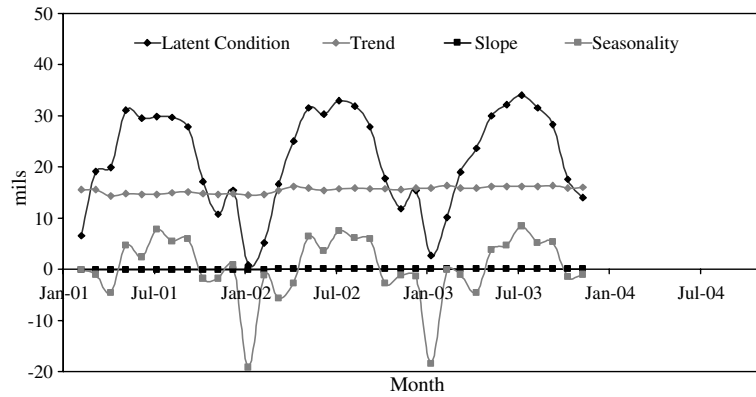


Fig. 4. Decomposition of latent condition for BSM-3 model.

The preferred model is presented in Eqs. (23)–(27). Again, randomness is entirely attributed to measurement errors. Note that the measurement error of deflection is still higher than that of pressure. However, the deflection has higher R^2 than pressure ($0.76 > 0.71$). It is expected that the R^2 values of BSM models are lower than those of ARMA models given the smaller number of parameters. However, the much worse data-fitting of BSM model for pressure could be explained by the fact that pressure does not follow presumed structures closely. Thus, for structural time series model, the best technology for measuring pavement condition would be FWD, which contradicts the result in ARMA models. Hence, in this empirical study, the preferred technology depends on the model specification, and therefore the test is inconclusive.

$$x_t = 0.93x_{t-1} - 0.67x_{t-2} + \mu_t + \gamma_t + \epsilon_t, \quad \epsilon_t \sim N(0, 4.91E - 08) \tag{23}$$

$$\mu_t = \mu_{t-1} + \beta_{t-1} + \eta_{t-1}, \quad \eta_t \sim N(0, 0) \tag{24}$$

$$\beta_t = \beta_{t-1} + \zeta_{t-1}, \quad \zeta_t \sim N(0, 0) \tag{25}$$

$$\gamma_t + \gamma_{t-1} + \gamma_{t-2} + \dots + \gamma_{t-s+1} = \omega_t, \quad \omega_t \sim N(0, 0) \tag{26}$$

$$\begin{bmatrix} z_t^{(1)} \\ z_t^{(2)} \end{bmatrix} = \begin{bmatrix} 1.00 \\ 0.26 \end{bmatrix} x_t + \begin{bmatrix} \xi_t^{(1)} \\ \xi_t^{(2)} \end{bmatrix}, \quad \Sigma_\xi = \begin{bmatrix} 12.91 & 0.16 \\ 0.16 & 1.28 \end{bmatrix}. \tag{27}$$

Fig. 5 also shows that the model makes condition forecasts adequately. The model is examined by the 33-month in-sample and 12-month out-of-sample prediction. The forecasts clearly are consistent with the structures of the model specification. For example, the seasonality patterns are identical for the near three-year in-sample prediction. On the other hand, the seasonality patterns show greater variety in the ARMA example in Fig. 3.

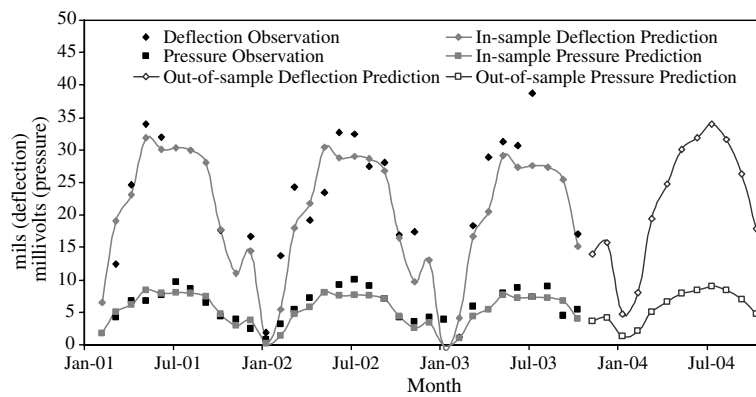


Fig. 5. In-sample and out-of-sample prediction for BSM-2 model.

5. Conclusions and discussion

We consider state-space specifications of ARMA and structural time series models as a framework to develop performance models for transportation infrastructure facilities. The work is motivated by the increased use of advanced technologies, e.g., sensors and radar, to evaluate and measure distresses on transportation facilities, as well as the factors that cause their degradation. In particular and in contrast to existing performance modeling approaches, the proposed framework exploits the information contained in the sequence of observations/measurements for individual facilities. From a managerial perspective, an attractive feature of the proposed framework is that the ensuing performance models fit the maintenance optimization framework of [Durango-Cohen \(2006\)](#), which means that they can be used to forecast infrastructure condition in a manner that is useful to support decision-making. Moreover and because the framework is consistent with the latent performance modeling approach of [Ben-Akiva and Ramaswamy \(1993\)](#), the capabilities of the inspection process, i.e., errors and the relationships between various indicators/measurements, can be estimated and properly accounted for in the forecasting process.

To illustrate the methodology, we present an empirical study where we develop and compare performance models for an asphalt pavement. The models in the study capture the evolution of a latent variable, i.e., the pavement's elasticity/load-bearing capacity, that manifests itself through pressure and deflection measurements collected with a pressure sensor and falling weight deflectometer, respectively. The results highlight the characteristics of each of the two types of models. ARMA models provide superior data-fitting capabilities, but the model parameters, including their magnitudes and signs, have no physical meaning. In contrast, structural time series provide parsimonious models that enable the identification of meaningful data components, e.g., trend, seasonality and random errors. We also use the numerical results to illustrate how the estimation procedure can be easily updated to account for missing observations/measurements, and how the proposed framework can be used to evaluate the capabilities of various (combinations of) inspection technologies.

An attractive feature of time series models in state space form is that intervention analysis can be used to estimate the effect of maintenance activities. Data limitations preclude this analysis in the current paper, but the application is presented in [Chu and Durango-Cohen \(2005\)](#).

Two relevant limitations of the framework described herein are that (i) the estimation procedure can be interpreted as the characterization of a stochastic process from a single realization, and (ii) constant explanatory factors, whose effect does not change over time, e.g., structural design, cannot be included because their effect cannot be identified from time series data. The first limitation precludes formulating general inferences about the underlying process, meaning that predictions are only warranted for identical facilities that deteriorate under identical conditions, if at all. The first limitations can be addressed by extending the framework to estimate models for panels of facilities as is shown in [Chu and Durango-Cohen \(2006\)](#).

Unfortunately, the second limitation is more troublesome. Even though the effect of constant explanatory variables, e.g., structural design, soil characteristics, etc., should be identifiable when panel data are used, the parameter estimates can, in fact, be biased due to the linearity of the state-space model. For example, when a constant explanatory variable is included in A_t , its effect and that of the random error term ϵ_t overlap in Eq. (1), and consequently, the parameter identification becomes difficult. [Chu and Durango-Cohen \(2006\)](#) further discuss this limitation, and how it can be “finessed” by employing nonlinear data transformations that do not compromise the linearity of state-space models.

Acknowledgement

The authors gratefully acknowledge Benjamin Worel and Bruce Chadbourn at MnROAD, who have provided extensive data and technical assistance for the study.

Appendix A. Square root filter

The general treatment of missing observations in state-space models is discussed in [Durbin and Koopman \(2001\)](#). This appendix describes our modification of the square root filter to cope with missing observations/measurements. The main idea is that when observations are missing at time t , the measurement vector Z_t is

separated into Z_t^* that consists of the observed measurements and $Z_t^\#$ that consists of the missing measurements. Therefore, the filter is modified to works with Z_t^* , whose dimensionality changes over time. Obviously, when all measurements are observed, $Z_t = Z_t^*$. To automatically separate vectors or matrices into observed and missing parts, the matrix W_t^* is introduced in the algorithm. W_t^* is a subset of the identity matrix $I_{k \times k}$. When the i th measurement is missing, the i th row of $I_{k \times k}$ is removed from the matrix. The remaining rows of the identity matrix constitute the matrix W_t^* .

The complete algorithm is presented below. In the presentation, we assume basic familiarity with the standard Kalman Filter, which can be found in, for example, [Harvey \(1990\)](#). To initialize the filter, values X_1 and P_1 are required. We adopt the diffuse initial values approach; that is, $X_1 = 0$ and $P_1 = kI$ where k is a large number and $I_{d \times d}$ is a $d \times d$ identity matrix. This implies that the knowledge of initial values is unavailable and they are calculated using the first d period(s) of observations. It also explains why the same time periods of the prediction errors are not counted in Eq. (19). Note that when models with different d values are compared, the largest value among them is used.

As mentioned above, the filter computes square roots of the covariance matrices, as shown in Eqs. (28)–(30), where P_t is the covariance matrix of state vector at time t .

$$\Sigma_\xi = \widetilde{\Sigma}_\xi \widetilde{\Sigma}'_\xi \quad (28)$$

$$\Sigma_\epsilon = \widetilde{\Sigma}_\epsilon \widetilde{\Sigma}'_\epsilon \quad (29)$$

$$P_t = \widetilde{P}_t \widetilde{P}'_t. \quad (30)$$

The complete algorithm is as follows: Square root filter

$$X_1 = 0 \quad (31)$$

$$\widetilde{P}_1 = \sqrt{k} I_{d \times d} \quad (32)$$

For $t = 1, \dots, T$:

$$\text{Generate } W_t^* \quad (33)$$

$$Z_t^* = W_t^* Z_t \quad (34)$$

$$\widehat{Z}_t^* = (W_t^* A) X_t \quad (35)$$

$$V_t^* = Z_t^* - \widehat{Z}_t^* \quad (36)$$

$$V_t^\# = 0 \quad (37)$$

Calculate $\widetilde{\Sigma}_\xi^*$ by applying Cholesky decomposition on

$$\Sigma_\xi^* = \widetilde{\Sigma}_\xi^* \widetilde{\Sigma}_\xi^{*'} = W_t^* (\widetilde{\Sigma}_\xi \widetilde{\Sigma}'_\xi) W_t^{*'} \quad (38)$$

$$U_t = \begin{bmatrix} (W_t^* A) \widetilde{P}_t & \widetilde{\Sigma}_\xi^* & 0 \\ g \widetilde{P}_t & 0 & \widetilde{\Sigma}_\epsilon \end{bmatrix} \quad (39)$$

$$\text{Calculate } U_t^\dagger \text{ by applying } QR \text{ decomposition on } U_t' = Q U_t^{\dagger'} \quad (40)$$

$$\text{Obtain } U_{1,t}^\dagger, U_{2,t}^\dagger, \text{ and } U_{3,t}^\dagger \text{ by partitioning } U_t^\dagger = \begin{bmatrix} U_{1,t}^\dagger & 0 & 0 \\ U_{2,t}^\dagger & U_{3,t}^\dagger & 0 \end{bmatrix},$$

$$\text{where } U_{1,t}^\dagger, U_{2,t}^\dagger, \text{ and } U_{3,t}^\dagger \text{ are } k \times k, d \times k, \text{ and } d \times d \text{ matrices.} \quad (41)$$

If observations are fully or partially available at time t then

$$F_t^* = U_{1,t}^\dagger U_{1,t}^{\dagger'} \quad (42)$$

$$X_{t+1} = g X_t + h A_t + U_{2,t}^\dagger U_{3,t}^{\dagger'}{}^{-1} V_t^* \quad (43)$$

End If

If the observations are fully missing at time t then

$$X_{t+1} = g X_t + h A_t \quad (44)$$

End If

$$\widetilde{P}_{t+1} = U_{3,t}^{\dagger'} \quad (45)$$

End For Loop

Eqs. (31) and (32) are used to initialize the filter at $t = 1$. The algorithm then processes W_t^* , Z_t^* , \widehat{Z}_t^* , and V_t as shown in Eqs. (33)–(37). Note that for the measurements that are missing, the prediction errors $V_t^\#$ are assumed zero as Eq. (37) shows, which implies that these measurements are not used to update the state vector. In Eq. (38) Σ_t^* is calculated using Cholesky decomposition.⁴ This step is reduced to Eq. (28) and can be skipped if all measurements are observed.

Eq. (40) makes use of the fact that Q is an orthogonal matrix to compute $U_t U_t' = (U_t^\dagger Q')(Q U_t^\dagger) = U_t^\dagger U_t^{\dagger'}$, where U_t^\dagger is a lower triangular matrix. F_t and P_t are obtained with $U_t U_t'$ (Eq. (39)) and $U_t^\dagger U_t^{\dagger'}$ (Eq. (41)). Note that QR decomposition⁵ is not used in the original square root filter. In Morf and Kailath (1975), U_t^\dagger is derived by multiplying U_t by any orthogonal matrix Q such that $Q'Q = I_{2d+k \times 2d+k}$. To provide an unambiguous procedure for implementation, this step is replaced by QR decomposition that gives a unique orthogonal matrix for calculating U_t^\dagger in Eq. (40). Eqs. (42)–(45) are obtained by replacing the original components in the standard Kalman filter with U_t^\dagger 's. Note that Eq. (44) implies that the prediction error variance is extremely large ($F_t^{*-1} = 0$) when no measurement is observed. Thus, no information is available for updating the state vector.

In terms of evaluating loglikelihood function when missing observations exist, V_t^* and F_t^* substitute V_t and F_t in the function. In the special case when all observations are missing for a time period, the dimensionality of the measurement vector reduces to zero and that particular time period is ignored.

References

- AASHTO, 1993. AASHTO Guide for Design of Pavement Structures 1993. American Association of State Highway and Transportation Officials, Washington, DC.
- Ben-Akiva, M., Gopinath, D., 1995. Modeling infrastructure performance and user costs. *Journal of Infrastructure Systems* 1 (1), 33–43.
- Ben-Akiva, M., Ramaswamy, R., 1993. An approach for predicting latent infrastructure facility deterioration. *Transportation Science* 27 (2), 174–193.
- Chu, C.Y., Durango-Cohen, P.L., 2005. Estimation of latent infrastructure performance models using time series analysis. *ASCE Journal of Transportation Engineering* (under review).
- Chu, C., Durango-Cohen, P.L., 2006. Estimation of dynamic performance models for transportation infrastructure using panel data: An application of multivariate time series analysis. *Transportation Research Part B* (under review).
- Durango-Cohen, P.L., 2006. A time series analysis framework for transportation infrastructure management. *Transportation Research Part B*, in press, doi:10.1016/j.trb.2006.08.002.
- Durbin, J., Koopman, S.J., 2001. *Time Series Analysis by State Space Methods*. Oxford University Press, New York, NY.
- Golabi, K., Pereira, P., 2003. Innovative pavement management and planning system for road network of Portugal. *Journal of Infrastructure Systems* 9 (2), 75–80.
- Gramling, W.L., 1994. *Current Practices in Determining Pavement Condition*. National Academy Press, Washington, DC.
- Harvey, A.C., 1990. *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge University Press, New York, NY.
- Humplick, F., 1992. Highway pavement distress evaluation: modeling measurement error. *Transportation Research Part B* 26B (2), 135–154.
- Madanat, S., Ben-Akiva, M., 1994. Optimal inspection and repair policies for infrastructure facilities. *Transportation Science* 28 (1), 55–61.
- Madanat, S.M., Karlaftis, M.G., McCarthy, P.S., 1997. Probabilistic infrastructure deterioration models with panel data. *Journal of Infrastructure Systems* 3 (1), 4–9.
- Mishalani, R.G., Madanat, S.M., 2002. Computation of infrastructure transition probabilities using stochastic duration models. *Journal of Infrastructure Systems* 8 (4), 139–148.
- Morf, M., Kailath, T., 1975. Square-root algorithms for least-squares estimation. *IEEE Transactions on Automatic Control* 20 (4), 487–497.
- Palmquist, D., Worel, B., Zervas, W., 2002. Mn/ROAD Hot-mix Asphalt Mainline Test Cell Condition Report. Office of Materials and Road Research, MnDOT, MN.
- Paterson, W.D.O., 1987. *Road Deterioration and Maintenance Effects: Models for Planning and Management*. The Johns Hopkins University Press, Baltimore, MD.
- Zervas, W.J., 2003. 2002 Superpave Report. Office of Materials and Road Research, MnDOT, MN.

⁴ Given a symmetric positive definite matrix A , the Cholesky decomposition is the lower triangular matrix L such that $A = LL'$.

⁵ Given a matrix A , the QR decomposition is a decomposition such that $A = QR$, where Q is an orthogonal matrix ($Q'Q = I$) and R is an upper triangular matrix.