Condition-Dependent Maintenance Effectiveness in Dynamic Performance Models for Transportation Infrastructure

James C. Chu[1]

Abstract: Dynamic performance models that combine performance prediction and maintenance effectiveness are required for state-of-the-art optimization techniques such as optimal control. Because records of maintenance effectiveness depend on facility condition, nonlinear models are necessary to include interactions between variables to account for this dependence and estimate condition-dependent maintenance effectiveness. Therefore, this paper proposes a procedure for estimating nonlinear dynamic performance models that capture interactions between variables using panel data. The relationships between maintenance effectiveness and current facility condition and between structural design and traffic impact were found to be polynomial in a numerical example of highway pavements. It was also demonstrated that imposing physical constraints on the maintenance effectiveness based on external data sources generated more reasonable models when self-selected samples were used for estimation. **DOI: 10.1061/(ASCE)IS.1943-555X.0000092.**

CE Database subject headings: Infrastructure; Life cycles; Predictions; Maintenance; Rehabilitation; Flexible pavements; Pavement overlays.

Author keywords: Infrastructure; Life cycles; Dynamic models; Predictions; Maintenance; Rehabilitation; Flexible pavements; Pavement overlays.

Introduction

Maintenance and rehabilitation (M&R) resource allocation plans for transportation infrastructure systems such as pavement and bridge networks are evaluated and selected based on their long-term economic consequences. The evaluation of these plans involves predicting their effect on infrastructure condition in the future and generating condition predictions with performance prediction models. Motivated by the importance of the M&R plans, researchers have developed numerous performance models over the last 50 years. One of the most significant reasons for developing a dynamic performance model is that optimal control using continuous state variables, which is a state-of-the-art optimization framework, requires dynamic models to generate optimal polices and rules for maintenance and rehabilitation plans (Madanat and Ben-Akiva 1994; Ouyang and Madanat 2006; Durango-Cohen 2007).

As advocated by Lytton (1987), including maintenance in performance models is attractive for supporting maintenance decision making in transportation infrastructure management. Obviously, modeling maintenance effectiveness correctly is crucial for making accurate performance predictions and M&R decisions. Labi and Sinha (2003) categorize the effects of maintenance into performance jumps and deterioration rate reductions based on their impact on facility condition. Performance jumps refer to transitory condition improvements, whereas deterioration rate reductions refer to permanent changes in deterioration rates after maintenance. The empirical models that quantify performance jumps include N.D. Lea International (1995) and Paterson (1990). Livneh (1998), Labi (2004), and Chu and Durango-Cohen (2008b) provide evidence that deterioration rates change after maintenance. In this research, maintenance effectiveness is defined as a performance jump because condition-dependent maintenance effectiveness is largely recognized in the literature as a performance jump (Al-Mansour et al. 1994; Paterson 1987, 1990; Watanabe et al. 1987). However, it should be acknowledged that the dependence of deterioration rate change on premaintenance condition is also possible.

Time series data containing both infrastructure performance progression and maintenance history must be used to estimate performance models that incorporate maintenance effectiveness. Estimating such models is not trivial because two dependences must be properly considered in a single performance model: the dependence of facility condition on maintenance history and the dependence of maintenance effectiveness on facility condition. In statistics, the dependence of current condition on historical condition is called serial dependence. Models that include this serial dependence of facility condition are called dynamic models or time series models in different fields. Allowing serial dependence means that the first dependence, the dependence of facility condition on maintenance history, can be estimated by analyzing facility conditions before and after each maintenance activity.

Commonly used regression models can be used as dynamic models by including lagged dependent variables. However, in the literature of infrastructure performance modeling, performance models and maintenance effectiveness have rarely been considered simultaneously in a regression model. Archilla and Madanat (2000, 2001) and Archilla (2006) represent some of the few attempts to consider maintenance in a (static) performance model (i.e., a model ignoring serial dependence). The studies developed performance

[1]Assistant Professor, Dept. of Civil Engineering, National Central Univ., No. 300, Jhongda Rd., Jhongli City, Taoyuan County 32001, Taiwan. E-mail: jameschu@ncu.edu.tw

Note. This manuscript was submitted on August 15, 2010; approved on December 29, 2011; published online on January 5, 2012. Discussion period open until August 1, 2013; separate discussions must be submitted for individual papers. This paper is part of the ***Journal of Infrastructure Systems***, Vol. 19, No. 1, March 1, 2013. © ASCE, ISSN 1076-0342/2013/1-85-98/$25.00.

models that include pavement rut depth as facility condition and overlay as maintenance. The assumption made in these studies was that when a section of pavement is overlaid, the rut depth will be reduced to a value similar to that of new pavement. In other words, instead of modeling the dependence of maintenance effectiveness on facility condition, the facility condition after the maintenance action was assumed to be constant to eliminate one of the two dependences. Alternatively, Ben-Akiva and Ramaswamy (1993) adopted a simultaneous equation model containing facility condition as a function of maintenance and maintenance as a function of facility condition. Maintenance and facility condition are both endogenous variables in such a model specification.

Chu and Durango-Cohen (2008a, b) are some of the studies using dynamic models to include maintenance in performance models. Note that those studies rely on linear dynamic models, which assume that variables have no interaction; therefore, the second dependence described previously, the dependence of maintenance effectiveness on facility condition, cannot be estimated in linear dynamic models. As a result, they are limited to estimating a constant maintenance effectiveness and are incapable of estimating maintenance effectiveness as a function of facility condition. The constant maintenance effectiveness is biased because the maintenance effectiveness is dependent on the premaintenance condition; this dependence has been demonstrated in various studies, e.g., Al-Mansour et al. (1994), Paterson (1987, 1990), and Watanabe et al. (1987). Therefore, nonlinear dynamic models that involve the interactions between variables to estimate maintenance effectiveness in conjunction with the premaintenance condition are necessary. If the dependence of maintenance effectiveness on facility condition is estimated, the current facility condition can be predicted by historical facility conditions and in response to maintenance effectiveness as a function of premaintenance condition. In such a model specification, facility condition is the dependent variable and maintenance application is an explanatory variable. Note that infrastructure performance models are particularly useful with maintenance application as an explanatory variable. Such models can be used to evaluate various M&R strategies to support maintenance optimization, which is one of the contributions of this research.

Motivated by this background, this paper describes a procedure for estimating nonlinear dynamic performance models that capture interactions between variables. Single-facility and multifacility nonlinear dynamic model formulations are first explained. The procedure for estimating their parameters is described later in the Appendix. Then, a numerical example to show the capability of capturing interactions between important variables (e.g., condition-dependent maintenance effectiveness) is demonstrated. Finally, the conclusion and areas for future research are presented.

Methodology

In this paper, dynamic models are formulated using state-space specifications. The state-space specification is a unified framework for describing a wide range of dynamic systems, including well-known autoregressive moving-average (ARMA) models and structural time series models. Once a dynamic model is transformed into state-space specifications, it can be estimated efficiently using a Kalman filter (Harvey 1990). Other advantages of state-space specifications over traditional time series analysis are that exogenous variables such as maintenance and environmental factors are straightforward to include, and measurement error and missing values can be easily dealt with (Chu and Durango-Cohen 2007, 2008a).

This section describes the relevant notation, definitions, and formulations for nonlinear dynamic models in state-space specifications. The estimation procedure for these models is described in the Appendix with illustrative examples.

Nonlinear Dynamic Models Using Single-Facility Data

Because the state-space specification for a single facility can be very complex on its own, introducing the complete formulation for panel data is sometimes impractical. This subsection defines the nonlinear state-space specification of dynamic models for a single facility in full detail, and later subsections describe the model for panel data in condensed form without providing complete formulations to save space.

A single-facility dynamic model is shown in Eqs. (1) and (2), which constitute a state-space model. Eq. (1) is the system equation, which governs the system's dynamics and captures the effect of explanatory variables. Eq. (2) is the measurement error equation, which captures random errors in the inspection process:

$$X_t = F(g_{t-1}, X_{t-1}, h_{t-1}, A_{t-1}, \omega_{t-1}) \tag{1}$$

$$Z_t = H(\Lambda_t, X_t, \xi_t) \qquad (t = 1, 2, \ldots, T) \tag{2}$$

where the variables, parameters, and random error terms in the model are explained briefly as follows: X_t: d-dimensional *state vector* representing the unobservable condition of infrastructure facility i at the start of period t, where $X_t = [x_t^{(1)}, \ldots, x_t^{(d)}]'$; A_t: s-dimensional vector of exogenous explanatory variables, e.g., structural design, history of maintenance and rehabilitation activities, environmental factors, and traffic loading, where $A_t = [A_t^{(1)}, \ldots, A_t^{(s)}]'$; Z_t: k-dimensional *observation vector* representing the set of observable condition measurements or indicators collected for facility i in period t, where $Z_t = [z_t^{(1)}, \ldots, z_t^{(k)}]'$. This vector may include measurements of distress such as pavement cracking, rutting, or raveling. The vector may also include subjective ratings or aggregate condition indices; g_t, h_t: contain the parameters that describe the transition of state vectors and the effect of explanatory variables on the state vector, respectively, where $g_t = [g_t^{(1)}, \ldots, g_t^{(p)}]'$ and $h_t = [h_t^{(1)}, \ldots, h_t^{(r)}]'$. The dimensions of the two vectors are p and r, respectively; Λ_t: dk-dimensional vector of parameters describing the relationship between measurements and condition, where $\Lambda_t = [\Lambda_t^{(1)}, \ldots, \Lambda_t^{dk}]'$; ω_t, ξ_t: represent random error terms assumed to be normally distributed with finite second moments, where $\omega_t = [\omega_t^{(1)}, \ldots, \omega_t^{(d)}]'$ and $\xi_t = [\xi_t^{(1)}, \ldots, \xi_t^{(k)}]'$. The error terms have dimensions $d \times 1$ and $k \times 1$. Σ_{ω_t} and Σ_{ξ_t} are their covariance matrices with dimensions $d \times d$ and $k \times k$, respectively. They satisfy $E(\omega_t) = 0$, $\mathrm{Var}(\omega_t) = \Sigma_{\omega_t}$, $E(\xi_t) = 0$, $\mathrm{Var}(\xi_t) = \Sigma_{\xi_t}$, and $E(\omega_t \xi_t) = 0$. Note that the assumption of normality is not a requirement for modeling a dynamic system. However, the assumption is made here because it is necessary for the model estimation procedure described later.

In what follows, three illustrative examples are discussed to demonstrate possible formulations that can be considered using dynamic models in state-space specifications. The first example is the linear dynamic model for a facility that has a single condition variable with first-order autoregression given by Eqs. (3) and (4). The unobservable condition of the facility is $x_t^{(1)}$ and the imperfect measurement of the condition is $z_t^{(1)}$. First-order autoregression implies that g_t contains only an autoregressive parameter $g_t^{(1)}$ in the system equation [Eq. (3)]. Assuming $A_{t-1}^{(1)}$ is the design strength of the facility, $A_{t-1}^{(2)}$ is the number of operating hours of the facility or traffic applications on a section of pavement, and $A_{t-1}^{(3)}$ is the unit of maintenance applied during time $t-1$. The gain in condition due to the design strength at time $t-1$ can be represented by $h_{t-1}^{(1)}A_{t-1}^{(1)}$, where $h_{t-1}^{(1)}$ is a parameter that represents the condition gain for a unit of design strength. Similarly, the deterioration of facility condition due to operation can be captured by the $h_{t-1}^{(2)}A_{t-1}^{(2)}$ term, where $h_{t-1}^{(2)}$ is a parameter representing the negative impact of an hour of operation or a unit of traffic on facility condition. The condition change caused by the maintenance can then be represented by $h_{t-1}^{(3)}A_{t-1}^{(3)}$, where $h_{t-1}^{(3)}$ is a parameter that indicates the condition change for a unit of maintenance. Clearly, the linear formulation of F is limited in predicting facility performance because the interactions between facility condition, exogenous variables, and errors are limited to being additive. It also means that only the dependence of facility condition on maintenance history can be modeled; the dependence of maintenance effectiveness on facility condition cannot be included. The final component in the system equation is an additive random error term. Note that, although a more complicated relationship can be assumed for the measurement process, an additive inspection error between the measurement and the true condition is assumed in Eq. (4) because this research focuses on the modeling of maintenance effectiveness in the system equation, and the measurement procedure is simplified for presentation. Also, the formulation for the measurement equation will remain the same and will not be repeated for the next two examples.

$$\begin{aligned} X_t = [x_t^{(1)}] &= [F_1(g_{t-1}, X_{t-1}, h_{t-1}, A_{t-1}, \omega_{t-1})] \\ &= [g_{t-1}^{(1)}x_{t-1}^{(1)} + h_{t-1}^{(1)}A_{t-1}^{(1)} + h_{t-1}^{(2)}A_{t-1}^{(2)} + h_{t-1}^{(3)}A_{t-1}^{(3)} + \omega_{t-1}^{(1)}] \end{aligned} \quad (3)$$

$$Z_t = [z_t^{(1)}] = [H_1(\Lambda_t, X_t, \xi_t)] = [\Lambda_t^{(1)}x_t^{(1)} + \xi_t^{(1)}] \quad (4)$$

As stated in the introduction, a nonlinear model is required to consider more realistic relationships between variables, i.e., the dependence between the maintenance effectiveness and the current facility condition. In this example, the assumption of linear formulation is relaxed and the function F is allowed to be nonlinear, as Eq. (5) shows. Due to the nonlinear nature of the formulations, interactions between variables can be incorporated into the models. For example, the impact of facility operation on its condition may be a function of the design strength. That is, when the design is strong, the impact due to operations should be small, and vice versa. Therefore, the impact for a unit of operation during time $t-1$ can be represented as $(h^{(1)} + h^{(2)}A_{t-1}^{(1)} + h^{(3)}(A_{t-1}^{(1)})^2)$, where a second-order polynomial relationship is assumed and $h^{(1)}$, $h^{(2)}$, $h^{(3)}$ are parameters representing this dependence. It follows that the total impact due to the facility operations during time $t-1$ should be $(h^{(1)} + h^{(2)}A_{t-1}^{(1)} + h^{(3)}(A_{t-1}^{(1)})^2)A_{t-1}^{(2)}$. The maintenance effectiveness could also be a function of the facility condition. Then the condition-dependent maintenance effectiveness during time $t-1$ can be expressed as $(h^{(4)} + h^{(5)}x_{t-1}^{(1)} + h^{(6)}(x_{t-1}^{(1)})^2)$, where the polynomial relationship is assumed again, and $h_{t-1}^{(4)}$, $h_{t-1}^{(5)}$, $h_{t-1}^{(6)}$ are parameters of this dependence. As defined previously, $A_{t-1}^{(3)}$ is the unit of maintenance applied during time $t-1$, and the total condition change caused by the maintenance in time period $t-1$ can then be represented by $(h^{(4)} + h^{(5)}x_{t-1}^{(1)} + h^{(6)}(x_{t-1}^{(1)})^2)A_{t-1}^{(3)}$. The term shows that a nonlinear relationship between the facility condition and the exogenous explanatory variable can be modeled. As a result, the dependence of facility condition on maintenance history and the dependence of maintenance effectiveness on facility condition can be modeled simultaneously in nonlinear dynamic models. Note that the error terms in the nonlinear functions are not necessarily additive but can have any form:

$$\begin{aligned} X_t = [x_t^{(1)}] &= [g^{(1)}x_{t-1}^{(1)} + (h^{(1)} + h^{(2)}A_{t-1}^{(1)} + h^{(3)}(A_{t-1}^{(1)})^2)A_{t-1}^{(2)} \\ &\quad + (h^{(4)} + h^{(5)}x_{t-1}^{(1)} + h^{(6)}(x_{t-1}^{(1)})^2)A_{t-1}^{(3)} + \omega_{t-1}^{(1)}] \end{aligned} \quad (5)$$

The third example demonstrates the possibility of adding external constraints to the formulation to obtain models that are more reasonable. One of the possible applications is as follows. To obtain adequate data for estimating condition-dependent maintenance effectiveness, maintenance should be applied under the full range of facility conditions, and maintenance records must be collected from these applications. However, for economic reasons, maintenance is rarely done when a facility is still in good condition. Therefore, in practice, only a limited range of data is available and the model parameter is not applicable to the range of facility conditions that is not observed; this is called *selectivity bias* in the literature (Madanat and Mishalani 1998). In this case, external information can be incorporated into the model estimation if concrete prior knowledge or trustworthy and independent sources of data are available (Ben-Akiva and Lerman 1985). For example, the maximum maintenance effectiveness of a maintenance action might be available in external reports. The information could be added to the estimation procedure as an additional constraint. Therefore, the first constraint that can be included in the model is that when the facility condition $(x_{t-1}^{(1)})$ is 0, maintenance effectiveness in Eq. (5), i.e., $(h^{(4)} + h^{(5)}x_{t-1}^{(1)} + h^{(6)}(x_{t-1}^{(1)})^2)$, is Δ_1, where Δ_1 is the maximum maintenance effectiveness found in the literature for the same situation. Similarly, maintenance actions usually result in no improvement to a brand new facility; therefore, the second constraint that can be considered is that when the facility condition $(x_{t-1}^{(1)})$ is Δ_2, maintenance effectiveness is 0, where Δ_2 is the facility condition for a newly constructed facility. Adding the two constraints into Eq. (5) leads to Eq. (6). Again, the condition-dependent maintenance effectiveness can be modeled and estimated with nonlinear dynamic models. In addition, the model formulations are flexible enough to incorporate the external sources of information to supplement the original data. As a result, the proposed methodology has the potential to estimate more satisfactory models in practice:

$$\begin{aligned} X_t = [x_t^{(1)}] &= \Bigg[g^{(1)}x_{t-1}^{(1)} + (h^{(1)} + h^{(2)}A_{t-1}^{(1)} + h^{(3)}(A_{t-1}^{(1)})^2)A_{t-1}^{(2)} \\ &\quad + \left(\Delta_1 + h^{(5)}x_{t-1}^{(1)} - \frac{\Delta_1 + h^{(5)}\Delta_2}{(\Delta_2)^2}(x_{t-1}^{(1)})^2\right)A_{t-1}^{(3)} + \omega_{t-1}^{(1)}\Bigg] \end{aligned} \quad (6)$$

All three of the preceding formulations and other alternative nonlinear relationships between variables will appear later in a numerical example. Actual data will be used to estimate various formulations of performance models to demonstrate how the proposed model can be used to estimate performance models with condition-dependent maintenance effectiveness.

Nonlinear Dynamic Model Using Panel Data

To cover the case of multiple facilities, a new subscript i is added to identify the facility, where $i = 1, \ldots, N$, resulting in $X_{i,t}$, $A_{i,t}$, $Z_{i,t}$, $\omega_{i,t}$, and $\xi_{i,t}$. Many assumptions about heterogeneity among facilities are possible for multiple-facility performance models (Chu and Durango-Cohen 2008a). The assumption used in this paper is that data from the facilities are instances of the same stochastic process, i.e., different surveys of the same stochastic process. Therefore, the system equation of a nonlinear dynamic model using panel data is based on Eq. (7), which implies that heterogeneity among facilities does not exist and that the specification has a single set of parameters shared by all facilities. Note that the assumption is chosen to focus on the estimation of condition-dependent maintenance effectiveness and to avoid the penalty of the large number of parameters required to consider heterogeneity. It is acknowledged that the consideration of heterogeneity improves the properties of the parameter estimators, and assuming that facilities are instances of the same stochastic process might cause biased parameters if heterogeneity is substantial in the data (Prozzi and Madanat 2003). The measurement equation is based on Eq. (8). Eqs. (9) and (10) are condensed to a nonlinear state-space model of the dynamics of a panel of N facilities. The functions $\mathbf{F}$ and $\mathbf{H}$ are collections of the functions for individual facilities. The vectors given previously are collected in $\mathbf{X_t}$, $\mathbf{Z_t}$, $\mathbf{A_t}$, $\mathbf{\Omega_t}$, and $\mathbf{\Xi_t}$. That is, $\mathbf{X_t} = [X'_{1,t}, \ldots, X'_{N,t}]$, $\mathbf{Z_t} = [Z'_{1,t}, \ldots, Z'_{N,t}]$, $\mathbf{A_t} = [A'_{1,t}, \ldots, A'_{N,t}]$, $\mathbf{\Omega_t} = [\omega'_{1,t}, \ldots, \omega'_{N,t}]'$, and $X_t = [\xi'_{1,t}, \ldots, \xi'_{N,t}]'$:

$$\mathbf{X_t} = \begin{bmatrix} X_{1,t} \\ X_{2,t} \\ \vdots \\ X_{N-1,t} \\ X_{N,t} \end{bmatrix} = \begin{bmatrix} F(g_{t-1}, X_{1,t-1}, h_{t-1}, A_{1,t-1}, \omega_{1,t-1}) \\ F(g_{t-1}, X_{2,t-1}, h_{t-1}, A_{2,t-1}, \omega_{2,t-1}) \\ \vdots \\ F(g_{t-1}, X_{N-1,t-1}, h_{t-1}, A_{N-1,t-1}, \omega_{N-1,t-1}) \\ F(g_{t-1}, X_{N,t-1}, h_{t-1}, A_{N,t-1}, \omega_{N,t-1}) \end{bmatrix} \tag{7}$$

$$\mathbf{Z_t} = \begin{bmatrix} Z_{1,t} \\ \vdots \\ Z_{N,t} \end{bmatrix} = \begin{bmatrix} H(\Lambda_t, X_{1,t}, \xi_{1,t}) \\ \vdots \\ H(\Lambda_t, X_{N,t}, \xi_{N,t}) \end{bmatrix} \tag{8}$$

$$\mathbf{X_t} = \mathbf{F}(g_{t-1}, \mathbf{X_{t-1}}, h_{t-1}, \mathbf{A_{t-1}}, \mathbf{\Omega_{t-1}}) \tag{9}$$

$$\mathbf{Z_t} = \mathbf{H}(\Lambda_t, \mathbf{X_t}, \mathbf{\Xi_t}) \tag{10}$$

$\Sigma_{\mathbf{\Omega}_t}$ and Σ_{Ξ_t} are calculated using Eqs. (11) and (12). In this paper, the elements in the state vectors are assumed to be independent to reduce the number of parameters and the computational burden. Therefore, covariances of ω_t (Σ_{ω_t}) are diagonal for all facilities. It follows that $\Sigma_{\mathbf{\Omega}_t}$ is also diagonal. Furthermore, because the assumption is made that the data from the facilities are instances of the same stochastic process, the covariances of ω_t are identical for all facilities. Similarly, the covariances of ξ_t (Σ_{ξ_t}) capture the relationships between the measurement errors for a facility. To reduce the number of parameters requiring estimation, Σ_{Ξ_t} is assumed to be block diagonal, which means that the measurement errors for different facilities are independent. Moreover, the covariances of ξ_t are identical for all facilities because of the assumption that measurement errors are attributed only to inspection technologies. This assumption is reasonable and widely accepted in engineering. However, if there is any reason to believe that the measurement errors are facility-specific, the assumption of identical covariances for measurement errors can be relaxed. Note that the number of samples available for model estimation is only $k \times N \times T$. However, the total number of elements in Σ_{ω_t} is $dN \times dN$ and the number for Σ_{ξ_t} is $kN \times kN$. In the panel data of transportation infrastructure, the number of facilities N is usually larger than the number of time periods T, which implies that the numbers of elements in the covariance matrices would be significantly larger than that of available samples. Therefore, the assumption of (block-)diagonal covariance matrices would be necessary in practice to reduce the number of parameters and estimate the model. Finally, the estimation procedure used to estimate the parameters of nonlinear dynamic models in state-space specifications for panel data is described in the appendix:

$$\Sigma_{\mathbf{\Omega}_t} = \mathrm{Var}(\mathbf{\Omega}_t) = \mathrm{Var}\left(\begin{bmatrix} \omega_{1,t} \\ \vdots \\ \omega_{N,t} \end{bmatrix} \right) = \begin{bmatrix} \Sigma_{\omega_t} & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \Sigma_{\omega_t} \end{bmatrix} = I_N \otimes \Sigma_{\omega_t} \tag{11}$$

where $\otimes$ denotes a Kronecker product:

$$\Sigma_{\Xi_t} = \mathrm{Var}(\Xi_t) = \mathrm{Var}\left(\begin{bmatrix} \xi_{1,t} \\ \vdots \\ \xi_{N,t} \end{bmatrix} \right) = \begin{bmatrix} \Sigma_{\xi_t} & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \Sigma_{\xi_t} \end{bmatrix} = I_N \otimes \Sigma_{\xi_t} \tag{12}$$

Empirical Study

This section describes the data source for the empirical study and the estimation results of the nonlinear dynamic performance models.

Data Source

The empirical study used the AASHO Road Test data set (Highway Research Board 1962). The AASHO Road Test was conducted between 1956 and 1961 near Ottawa, Illinois. The objectives of the test included evaluating the impact of structural design (surface, base, and subbase thicknesses) and traffic loading (axle applications, load, and configuration) on the performance of flexible pavements. The test was carefully designed to investigate the effects of the important factors and interactions between them in a comprehensive manner. More specifically, the test tracks consisted of six loops with two lanes each. Each lane was segmented into sections containing different combinations of pavement layer thicknesses. Sections of the two lanes in Loop 1 were used as a control and not subjected to traffic. All sections of the other 10 lanes were subjected to approximately the same number of axle applications but different axle loadings and configurations during the test. In general, the traffic loading and pavement thickness increased as the loop number increased. See Highway Research Board (1962) for more details of the test design.

The pavement serviceability, distress measurements, and other factors such as the number of traffic loadings and temperature were recorded biweekly. A total of 56 measurements were made for each variable and indexed by $t = 1, \ldots, 56$ in the model specifications. For example, $t = 1$ indicates the data collected on November 3, 1958, and $t = 56$ indicates the data collected on December 1, 1960. The actual dates of the data collection can be found in the Highway Research Board (1962). Because maintenance effectiveness is the main focus of the research, pavement sections are categorized into two groups according to their need for maintenance. The categorization is useful for discussing the different purposes of these pavements in the model estimation. Of the 332 sections, 125 did not fail during the test and thus did not require any major maintenance intervention. As a result, they represent pavement designs that are strong relative to applied traffic loads and thus have no need for maintenance. On the other hand, 207 of the sections failed during the test. Clearly, these sections require maintenance and represent weak pavement designs with respect to applied traffic loads. Of these failed sections, 120 failed very early in the test (on average, 9 months after the start). They were not maintained, and measurements were not collected for these sections after failure, i.e., their records contain missing values as time series data. To eliminate the influence of missing values, the pavement sections whose data were not recorded during the whole period are not selected for analysis. The exclusion of the sections with incomplete records causes potential sampling bias. However, this approach is justified because the time series data collected for these sections are limited (18 observations per section on average). In addition, the early failure of sections indicates that their pavement designs were highly inadequate and are unlikely to be used in practice. Although the pavement sections that failed very early are excluded from the data set, the other 87 failed sections whose failures occurred late in the test can still be included in the data set. Because their designs are more adequate and more likely to be used in practice, they were maintained after failure and then reopened to traffic. Collection of measurements continued after maintenance, which means that these pavements have complete records for the whole test period test, and they do not suffer from the problem of missing values. In fact, including these pavements is valuable for the example because they not only constitute the samples of weaker pavement designs, but they also provide data samples for the estimation of maintenance action effectiveness. In addition to the aforementioned early failures mentioned, 24 sections from Lane 2 of Loop 1 were not available for this study. As a result, 188 pavement sections are considered in this example. The selected pavement sections were indexed by i, where $i = 1, \ldots, 188$, and the pavement condition in the data was represented by the present serviceability index (PSI). The PSI is a widely accepted indicator of a pavement's serviceability or functional performance. The average PSI values for the inner and outer wheel paths of each pavement section were used in this paper. The theoretical range of PSI is 0.0 to 5.0, and the actual range of the observations in the data set was 0.1–4.7. The index is computed by

$$\mathrm{PSI} = 5.03 - 1.91 \cdot \log_{10}(1 + \mathrm{SV}) - 0.002139 \cdot \mathrm{RD}^2 - 0.01 \cdot \sqrt{\mathrm{CR} + \mathrm{PA}}$$

where SV is the slope variance (adimensional), RD is the rut depth in millimeters, and CR and PA correspond to cracking and patching in square meters per 1,000 square meters surface area.

The maintenance strategy in the test was to overlay the pavement when the PSI reached 1.5. The detailed history is recorded for each overlay application, including the date, thickness, and PSI values before and after overlay. In the 188 pavements used for estimation, 101 of them were not overlaid and the remaining ones (87) received exactly one overlay during the test. The overlay thickness used in the test ranged from 2 to 4 in.; however, the majority of them (66 out of 87) were 3 in. The exogenous explanatory variables and their notations considered in this study were as follows:

Variable SN_i

The variable SN_i is structural number of section i. A pavement's structural number serves as a proxy for its structural design and is a function of the pavement surface, base, and subbase thicknesses. The formula calibrated by the Highway Research Board (1962) and used here is

$$\mathrm{SN} = 0.44D_1 + 0.14D_2 + 0.11D_3$$

where D_1, D_2, and D_3 represent the thickness in inches of surface, base, and subbase, respectively. The thickness of the surface layer clearly provides more strength than the thicknesses of the other two layers. The range of SN in the data set was 0.44 to 5.66. Note that the data for reestimating the function of structural numbers are available for the research. Alternative functions for calculating structural numbers can be found in Archilla and Madanat (2000, 2001) and Prozzi and Madanat (2003, 2004). However, the function originally calibrated by the Highway Research Board (1962) was used to simplify the estimation and to allow direct comparison of the previous works on dynamic performance models such as Chu and Durango-Cohen (2008a).

Variable $\mathrm{TRF}_{i,t}$

The variable $\mathrm{TRF}_{i,t}$ is the seasonal weighted traffic loading applied to section i during time period t measured in 10^5 equivalent single-axle loads (ESALs). Traffic and weathering are the critical factors in pavement deterioration. Load applications were transformed to ESALs according to the axle loads and axle configurations. ESALs were further adjusted by a seasonal weighting function to account for environmental factors that make pavement more or less vulnerable to deterioration due to traffic. The seasonal weighting function that captures the effects of ambient temperature and frost depth by measuring the pavement deflection at the time of loading was established by the Highway Research Board (1962). The range of the function is 0.00 to 4.44; in general, the values of function are highest in spring due to the freeze-thaw effect and lowest in winter due to the high stiffness of the pavement. The complete function is not presented here due to space limitations. The range of unweighted traffic is 0.0 to 2.71 in the data and that of the weighted traffic ($\mathrm{TRF}_{i,t}$) is 0.00 to 9.64, both in 10^5 ESALs. Note that many studies have proposed methods to modify traffic impact as a function of environmental factors and axle configurations (Archilla and Madanat 2000, 2001; Archilla 2006; Prozzi and Madanat 2003, 2004; Small and Winston 1988). However, the function calibrated by the Highway Research Board (1962) was used to focus on maintenance-related variables and to compare with the related studies of dynamic performance models, e.g., Chu and Durango-Cohen (2008a). Modified estimation methods or estimates obtained by the aforementioned methods listed will be considered in future research to improve estimation results.

Variable $\mathrm{OVR}_{i,t}$

The variable $\mathrm{OVR}_{i,t}$ is the indicator variable where $\mathrm{OVR}_{i,t} = 1$ when an overlay is applied on section i between time t and $t + 1$, and 0 otherwise. A dummy variable of overlay application is chosen instead of overlay thickness because the variability in the records of overlay thickness is rather low, as described earlier.

Key Observations

Three key observations reported by the Highway Research Board (1962) will be useful later in the model estimation and validation. The first observation is that the average PSI of newly constructed pavements was 4.2. This value is the upper limit of the pavement condition in the analysis that follows. The second key observation is that the application of overlays was difficult due to the cold weather and short pavement sections, and the average PSI value of the pavement immediately after overlay application was only 3.4.

The third observation involves the maintenance strategy of the AASHO Road Test. Even though the strategy in that test was to overlay the pavement when the PSI reached 1.5, the actual PSI before the overlay was actually applied varied somewhat. This occurred because one wheel path failed faster than the other, so the overlays were applied before reaching an average PSI of 1.5, and overlays were delayed due to maintenance crew schedules. As a result, the PSI values of conditions before the overlay ranged from 0.0 to 2.2. When these data are used for estimating maintenance

J. Infrastruct. Syst. 2013.19:85-98.

effectiveness, the implication is that the data were collected only from pavement in relatively poor condition. The estimated effectiveness is only certain for pavement in this range of PSI values, and the results could be incorrect for pavement in different conditions. Madanat and Mishalani (1998) referred to this bias as *selectivity bias* because the samples for estimation were already self-selected. This problem, if it exists, would lead to biased estimates for maintenance effectiveness and cause M&R optimization frameworks to generate erroneous solutions. Madanat and Mishalani (1998) assumed that the bias could be corrected by the preference for selecting maintenance activities of transportation agencies. Recently, Prozzi and Hong (2010) considered the same problem with average treatment effect modeling also using the AASHO Road Test data. In this paper, an alternative approach that imposes constraints on the situations that were not observed is adopted to address this data limitation.

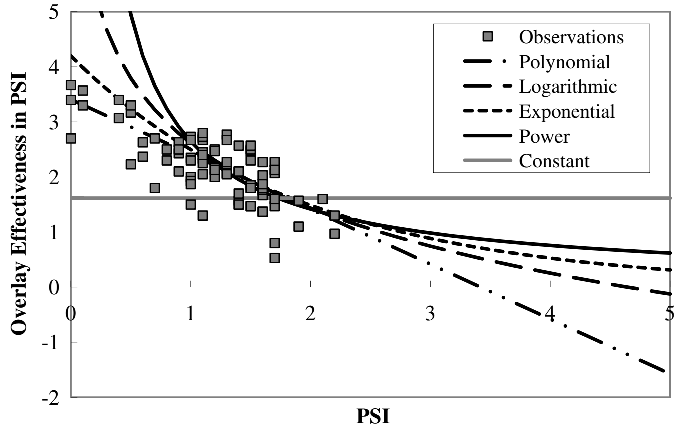

Fig. 1. Relationship between pavement condition and maintenance effectiveness (linear model and nonlinear model without constraints)

Estimation Results and Discussion

Linear Models

For the preliminary study, the linear dynamic performance model specification proposed by Chu and Durango-Cohen (2008a) is given by Eqs. (13) and (14). The unobservable true condition of pavement section i at time t is $x_{i,t}^{(1)}$, and $z_{i,t}^{(1)}$ is the PSI of section i at time t, which is an indicator of the true condition. The measurement equation implies that PSI is the imprecise measurement for the true condition, and thus $\Lambda_t = [\Lambda_t^{(1)}] = [1]$. The specification of the measurement equation remains the same for the remainder of this paper and is not further mentioned. There are two more important points in the model formulation. First, the model formulations are time-dependent so far for generality. In infrastructure management, time-dependent models are useful to capture the rate change of deterioration due to the application of maintenance activities (Chu and Durango-Cohen 2008b). However, because this change of rate is not the main focus of the research, the models are assumed to be *time-homogeneous* or *time-invariant* in this numerical example, i.e., the parameters are not indexed by time. This assumption is necessary to reduce the parameters to a reasonable number and also to reduce the computational effort required for model estimation. Second, because there is only one state variable and one measurement for each facility, the superscripts in $x_{i,t}^{(1)}$, $z_{i,t}^{(1)}$, $\omega_{i,t}^{(1)}$, and $\xi_{i,t}^{(1)}$ are dropped henceforth for simplicity.

Eqs. (15) and (16) describe the estimated model. The structural design (SN) in Eq. (15) has a direct additive effect on the pavement condition. This implies that a pavement section has constant differences of condition from other pavement sections throughout the deterioration process due to the difference in structural number. A negative TRF parameter indicates that traffic negatively impacts the pavement, and a positive OVR parameter indicates that the overlay improves the pavement condition, both of which are reasonable. The result also shows that the PSI improvement would be 1.616, indicated by the horizontal line in Fig. 1, which shows the values of PSI improvements. Note that the values of PSI improvements are reported without considering the effect of other variables (TRF and SN) in this paper to focus on the analysis for maintenance effectiveness. This is reasonable because the effect of nonoverlay variables in 2 weeks is negligible compared to the impact of a major maintenance event. The estimate of maintenance effectiveness appears biased when compared to the average value of the PSI improvement observations in the data set (2.170). A possible explanation is that dependence may exist among $x_{i,t-1}$, SN, and OVR. More specifically, maintenance effectiveness depends on pavement conditions, and pavement conditions depend on structural design. These factors are omitted in a linear model, resulting in the bias.

The order of autoregression (the value of d) was also tested. Models with higher orders of autoregression were estimated (results not included here), and the autoregressive parameters were highly insignificant at the 95% level, meaning that these specifications were inappropriate. Therefore, first-order autoregression was sufficient for the data set:

$$\begin{aligned} x_{i,t} = g^{(1)}x_{i,t-1} + \cdots + g^{(d)}x_{i,t-d} + h^{(1)}\mathrm{SN}_{i,t} + h^{(2)}\mathrm{TRF}_{i,t-1} \\ + h^{(3)}\mathrm{OVR}_{i,t-1} + \omega_{i,t-1}, \omega_{i,t} \sim N(0, \sigma_\omega^2) \end{aligned} \tag{13}$$

$$z_{i,t} = x_{i,t} + \xi_{i,t}, \xi_{i,t} \sim N(0, \sigma_\xi^2) \tag{14}$$

$$\begin{aligned} x_{i,t} = 0.986x_{i,t-1} + 0.00568\mathrm{SN}_{i,t} - 0.210\mathrm{TRF}_{i,t-1} \\ + 1.616\mathrm{OVR}_{i,t-1} + \omega_{i,t-1}, \omega_{i,t} \sim N(0, 0.184^2) \end{aligned} \tag{15}$$

$$z_{i,t} = x_{i,t} + \xi_{i,t}, \xi_{i,t} \sim N(0, 0.113^2) \tag{16}$$

Nonlinear Models without Constraints

Two nonlinear phenomena were considered in the system equation for this example. The first is the effect of structural design on traffic impact (SN − TRF). With a nonlinear model, structural design can have a nonlinear relationship with traffic impact and thus also have an indirect effect on the pavement condition. In other words, the impact of a unit of traffic depends on the structural design, which is a more reasonable and relaxed assumption. This study assumed four specifications of the nonlinear effect of structural design on traffic impact to demonstrate the flexibility of the methodology. The selection is rather arbitrary, but it will be shown later that satisfactory specifications can be found among these four commonly used specifications:

- Power specification: $h^{(1)}\mathrm{SN}_{i,t}^{h^{(2)}}\mathrm{TRF}_{i,t-1}$
- Logarithmic specification: $h^{(1)}\log(h^{(2)}\mathrm{SN}_{i,t})\mathrm{TRF}_{i,t-1}$
- Exponential specification: $h^{(1)}\exp(h^{(2)}\mathrm{SN}_{i,t})\mathrm{TRF}_{i,t-1}$
- Polynomial specification: $(h^{(1)} + h^{(2)}\mathrm{SN}_{i,t} + h^{(3)}(\mathrm{SN}_i)^2)\mathrm{TRF}_{i,t-1}$.

The second nonlinearity considered in the model specification was the condition-dependent maintenance effectiveness (x − OVR). As explained in the introduction, maintenance effectiveness depends on the pavement condition and can be formulated as a

J. Infrastruct. Syst. 2013.19:85-98.

function of historical pavement condition to predict the current condition. Since the exact formulation of the effectiveness of the overlay is unclear, the following four specifications, which are frequently used to describe nonlinearity, were assumed to estimate empirical pavement prediction models. Again, satisfactory specifications can be found among these arbitrarily selected specifications, which will be shown later. All of these specifications indicate that maintenance effectiveness is a nonlinear function of pavement condition, and each specification assumes a different nonlinearity of the relationship:

- Power specification: $h^{(4)} x_{i,t-1}^{h^{(5)}} \mathrm{OVR}_{i,t-1}$
- Logarithmic specification: $h^{(4)} \log(h^{(5)} x_{i,t-1}) \mathrm{OVR}_{i,t-1}$
- Exponential specification: $h^{(4)} \exp(h^{(5)} x_{i,t-1}) \mathrm{OVR}_{i,t-1}$
- Polynomial specification: $(h^{(4)} + h^{(5)} x_{i,t-1} + h^{(6)} (x_{i,t-1})^2) \mathrm{OVR}_{i,t-1}$.

To understand the effect of all candidate nonlinear specifications, combinations of the four types of traffic impact-structural design and four types of condition-maintenance effectiveness relationships were estimated. The cells under SN − TRF specifications without constraints in Table 1 list Akaike's information criterion (AIC) of the 16 (4×4) combinations. The AIC is a measure for selecting statistical models, and a smaller value is better. The measure includes the number of parameters as a penalty in addition to the data fit. However, because the estimation results presented below indicate that higher-order polynomials are insignificant at the 95% level in all of these models, the first-order polynomial specifications are used for both nonlinearities. As a result, the number of parameters is identical for all 16 models in this example. Because its value depends on the scale of the variables, it is a relative measure and does not provide information on the absolute quality of the models. More details about AIC can be found in Harvey (1990).

It is clear from the table that the specifications chosen for condition-maintenance effectiveness (x − OVR) and structural design-traffic impact (SN − TRF) are relevant, and the results are more sensitive to the specifications of x − OVR than those of SN − TRF. First-order polynomial specifications provide a better model fit than others for the condition and maintenance effectiveness relationship. Logarithmic is second-best, exponential is third, and the power specification is the worst. For the specification of structural design-traffic impact, the order of specifications based on the AIC is first-order polynomial, exponential, logarithm, and power. Table 2 lists the parameter estimates and their t-statistics of polynomial condition-maintenance effectiveness (x − OVR) specifications. All parameters have intuitive signs and are statistically significant at the 95% level except for $h^{(2)}$ parameters. For example, $g^{(1)} < 1$ indicates that the condition deteriorates over time. Note that the inspection interval is relatively small and the deterioration between two inspections is minimal, which explains why the parameter values for $g^{(1)}$ are very close to 1.

The results of the two nonlinear factors can be interpreted as follows. First, a negative value for $h^{(1)}$ indicates that the traffic causes damage to the pavement. All the parameters related to traffic impact and structural design ($h^{(1)}$ and $h^{(2)}$) except for one are insignificant at the 95% level. The reason for these insignificant parameters is the autoregressive parameter ($g^{(1)}$). The decreasing trend of the PSI was captured by both the autoregressive parameter and the traffic impact. Modeling the complex relationship of traffic impact and structural design is clearly more difficult than simple autoregression, which makes the parameters insignificant. Note that the insignificant parameters and the use of the first-order polynomial specification do not imply that the structural design has no effect on the traffic impact or that their relationship is linear. A nonlinear relationship exists between the structural design and traffic impact, e.g., Highway Research Board (1962), AASHTO (1993), Archilla and Madanat (2000, 2001), Prozzi and Madanat (2003, 2004), and Small and Winston (1988). The polynomial specification was eventually selected for the traffic impact-structural design relationship because those parameters were more significant than the others; $h^{(2)}$ is significant at approximately the 85% level. Fig. 2 shows the relationship between the structural number and the corresponding traffic impact. As expected, pavement with high SN (i.e., stronger) experiences small traffic impact, and vice versa. This specification is more meaningful than including SN as an additive variable.

Second, Fig. 1 shows the relationship between the condition before the overlay and the maintenance effectiveness of the overlay and compares the observations and the predictions of the maintenance effectiveness dependent on the condition before the overlay. The figure clearly indicates that the effectiveness of the overlay increases as the condition before the overlay decreases, which again

Table 1. AIC for Specification Combinations (Smaller Values Are Better)

Specification	Power SN − TRF without constraints	Logarithm SN − TRF without constraints	Exponential SN − TRF without constraints	Polynomial SN − TRF without constraints	Polynomial SN − TRF with constraints
Power x − OVR	−7.824	−7.831	−7.832	−7.841	−10.916
Exponential x − OVR	−9.459	−9.466	−9.466	−9.475	−10.915
Logarithm x − OVR	−9.509	−9.513	−9.515	−9.523	−10.915
Polynomial x − OVR	−10.873	−10.875	−10.878	−10.885	−10.915

Table 2. Estimation Results of Nonlinear Model without Constraints

SN − TRF	Polynomial		Polynomial		Polynomial		Polynomial	
x − OVR	Polynomial		Exponential		Logarithm		Power	
Parameter	Estimate	t-stat	Estimate	t-stat	Estimate	t-stat	Estimate	t-stat
σ_ω	−0.181	−68.2	−0.181	−68.0	−0.180	−68.3	−0.179	−68.1
σ_ξ	0.108	41.1	0.109	41.7	0.110	42.6	0.112	43.9
$g^{(1)}$	0.991	1692.2	0.991	1691.5	0.991	1697.2	0.991	1703.5
$h^{(1)}$	−0.036	−2.6	−0.040	−2.9	−0.038	−2.8	−0.040	−3.0
$h^{(2)}$	**0.004**	**1.3**	**0.005**	**1.6**	**0.004**	**1.4**	**0.005**	**1.6**
$h^{(4)}$	3.407	49.3	4.199	27.8	−1.703	−21.4	2.674	38.1
$h^{(5)}$	−0.997	−27.1	−0.519	−25.0	0.215	21.8	−0.864	−19.0
AIC	−10.885		−9.475		−9.523		−7.841	

Note: Entries in bold-face type indicate that a parameter is insignificant at the 95% confidence level.

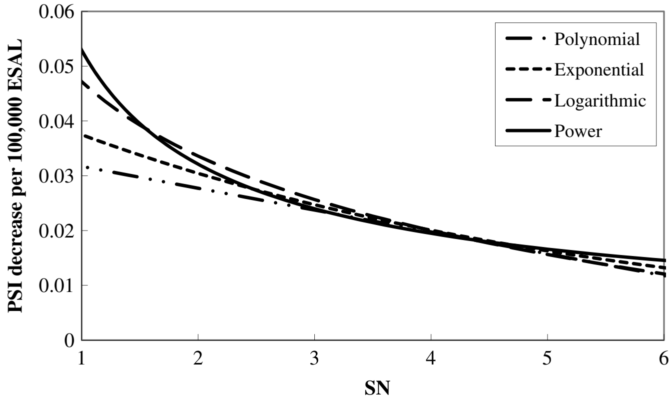

Fig. 2. Relationship between structural number and traffic impact

is logical. It is clear that all the specifications capture fairly the trend within the range of observations (PSI ≤ 2.2). The polynomial specification has the best AIC. However, the prediction that the maintenance effectiveness would be negative for a PSI greater than approximately 3.4 seems unreasonable. Compared to the polynomial specification, the exponential specification has positive values over the whole range. On the other hand, the predictions of the logarithmic and power specifications are too large for poor conditions, and the condition after the overlay could exceed the theoretical maximum value of PSI $= 5.0$, which is inconsistent with expectations. The polynomial specification was chosen to represent the condition-maintenance effectiveness ($x - \mathrm{OVR}$) relationship due to its AIC.

As an example, Eq. (17) shows the first-order autoregressive specification with polynomial traffic impact-structural design and condition-maintenance effectiveness. Eqs. (18) and (19) show the estimated model. Note that the error terms are assumed to be additive. The major advantage of additive error is the simplified calculation in the estimation procedure, which is discussed in the Appendix. Because the main focus of this paper is to study the nonlinear relationship between the condition and explanatory variables, the assumption that the error terms are independent of other factors is natural. Moreover, the assumption of additive error terms is reasonable and widely accepted in engineering. Thus, the formulation was adopted to simplify the estimation in the empirical example. Even so, the methodology described above still applies if multiplicative errors are assumed:

$$x_{i,t} = g^{(1)}x_{i,t-1} + (h^{(1)} + h^{(2)}\mathrm{SN}_{i,t})\mathrm{TRF}_{i,t-1} + (h^{(4)} + h^{(5)}x_{i,t-1})\mathrm{OVR}_{i,t-1} + \omega_{i,t-1}, \omega_{i,t} \sim N(0, \sigma_\omega^2) \quad (17)$$

$$x_{i,t} = 0.991x_{i,t-1} + (-0.036 + 0.004\mathrm{SN}_{i,t})\mathrm{TRF}_{i,t-1} + (3.407 - 0.997x_{i,t-1})\mathrm{OVR}_{i,t-1} + \omega_{i,t-1}, \omega_{i,t} \sim N(0, -0.181^2) \quad (18)$$

$$z_{i,t} = x_{i,t} + \xi_{i,t}, \xi_{i,t} \sim N(0, 0.108^2) \quad (19)$$

Nonlinear Models with Physical Constraints

The models presented previously are estimated mainly based on the observed data, and no physical constraints are used in the model specifications. In the following model estimation, the physical conditions of maintenance effectiveness of pavement overlay are considered to study the condition-maintenance effectiveness ($x - \mathrm{OVR}$) relationship closely. As previously mentioned, the Highway Research Board (1962) observed that the condition of newly constructed pavements was PSI $= 4.2$ on average. Therefore, improving the condition to a PSI greater than 4.2 is unreasonable. In addition, it was reported that the average PSI value immediately after overlay application was only 3.4 due to cold weather and short pavement sections. These observations provide valuable insights regarding the results from the previous section. It is somewhat surprising that the polynomial model captures this trend perfectly without any constraints. Its horizontal axis intercept is 3.407, which is in agreement with the value reported by the Highway Research Board (1962). The function can be interpreted as the application of overlay bringing the condition of the pavement to approximately PSI $= 3.4$. According to these observations, the predictions of the other models for pavement with PSI > 2.2 are clearly overestimated.

The predictions from most of the foregoing specifications are not consistent with expectation due to the fact that the overlay was only observed in a certain range of PSI values. More treatments would be necessary to obtain more reasonable models. It is not uncommon that the sample is too small or the variability of one or more variables is limited for estimating a statistical model, and constrained estimation is a valid choice in such situations (Ben-Akiva and Lerman 1985). Imposing external constraints is justified if these constraints are based on trustworthy, independent data sources or prior knowledge. If the constrained parameters are set to correct values, the statistical efficiency of the model relative to other unconstrained parameters can be improved (Ben-Akiva and Lerman 1985). This technique has been adopted in studies in various fields (Hauser and Shugan 1980; Srinivasan et al. 1983; Carey and Revelli 1986; Swait and Ben-Akiva 1986). N.D. Lea International (1995) summarized the studies of overlay effectiveness in multiple countries and estimated equations for the relationship between IRI (International Roughness Index) before overlay and IRI change due to overlay. The equations were developed for the World Bank's Highway Design and Maintenance Standards Model-4 (HDM-4), which has been widely used by consultants and government departments to study the impacts of maintenance investments in transportation infrastructure. To formulate the constraints for the constrained estimation, two key values are extracted from the equations for the overlay with a thickness of 80 mm (3.1 in., which is close to the overlay thickness applied in the test).

First, N.D. Lea International (1995) reported that the highest PSI achievable due to overlay is 3.4 (IRI $= 2$) for pavements with extremely poor premaintenance conditions. This can be used to set the maintenance effectiveness to 3.4 when PSI $= 0$ and formulate the first constraint. [The equation used to convert IRI to PSI is IRI $= 5.5\ln(5.0/\mathrm{PSI})$, which was estimated by Paterson (1986).] As stated in the "Data Source" section, the average PSI value of the pavement immediately after overlay application was only 3.4 for premaintenance $0 \leq \mathrm{PSI} \leq 2.2$ in the AASHO Road Test. Therefore, the two sources are consistent in the maintenance effectiveness for the worst premaintenance conditions.

Second, N.D. Lea International (1995) also found that the overlay had no effect for pavement with PSI ≥ 3.4 (IRI ≤ 2). The information can be used to formulate a second constraint that the effectiveness is zero for pavements with PSI $= 3.4$. Again, the constraint can be compared with the AASHO Road Test to gain more confidence. It was reported in the Highway Research Board (1962) that the applications of overlay were difficult due to the cold weather and short pavement sections. Thus, it is unlikely that the overlay brings pavements back to the average PSI value for brand new pavements, 4.2. It would be more plausible to use the observation from the range of $0 \leq \mathrm{PSI} \leq 2.2$ to extrapolate the range of PSI > 2.2. That is, the PSI after overlay for pavements

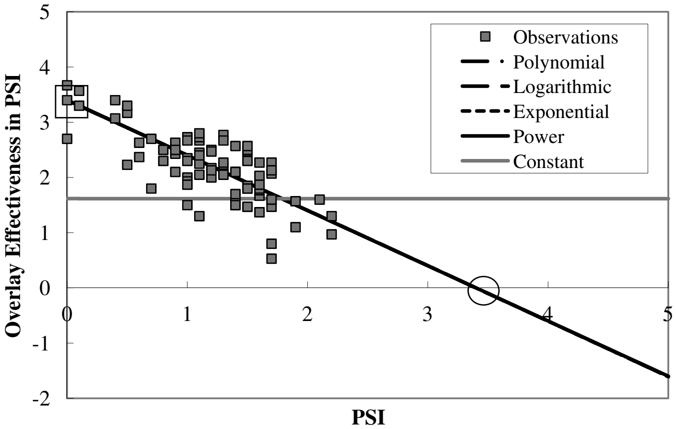

Fig. 3. Relationship between pavement condition and maintenance effectiveness (nonlinear models with Constraints 1 and 2) (after imposing the two constraints, the predictions generated by the different specifications are almost identical, and thus the lines overlap in the figure)

with good condition should be close to 3.4, which is consistent with the second constraint determined previously. In summary, the constrained estimation must satisfy that the following conditions:

- Constraint 1: effectiveness is 3.4 for $x_{i,t-1} = 0$
- Constraint 2: effectiveness is 0 for $x_{i,t-1} = 3.4$.

By imposing these two constraints, the foregoing specifications for the condition-maintenance effectiveness ($x - \mathrm{OVR}$) relationship can be rewritten as follows. Because the exponential and power functions do not intersect the horizontal axis and the logarithm function does not intersect the vertical axis, an additional constant must be added to each of these specifications to satisfy the constraints. Also, one of the parameters of the polynomial function is fixed due to Constraint 1. Therefore, the two constraints will eliminate one parameter from each of the model specifications as well as the additional or constrained constant.

- Power specification: $3.4(1 - x_{i,t-1}^{h^{(5)}}/3.4^{h^{(5)}})\mathrm{OVR}_{i,t-1}$
- Logarithmic specification: $3.4\log(1 + h^{(5)}(x_{i,t-1} - 3.4))/\log(1 - h^{(5)}3.4)\mathrm{OVR}_{i,t-1}$
- Exponential specification: $3.4(\exp(h^{(5)}x_{i,t-1}) - \exp(h^{(5)}3.4))/1 - \exp(h^{(5)}3.4)\mathrm{OVR}_{i,t-1}$
- Polynomial specification: $(3.4 + h^{(5)}x_{i,t-1} - 3.4 + h^{(5)}3.4/3.4^2(x_{i,t-1})^2)\mathrm{OVR}_{i,t-1}$.

Fig. 3, which parallels Fig. 1, shows the relationship between the current condition and the maintenance effectiveness for the two constraints. Constraint 1 forces all specifications to pass the location of the square, and Constraint 2 requires the functions to pass the location of the circle. Clearly, the different specifications are almost identical after imposing the two constraints, and, more importantly, they all produce very reasonable results. The AICs and estimation results of nonlinear models with polynomial $\mathrm{SN} - \mathrm{TRF}$ and different Constraints 1 and 2 for $x - \mathrm{OVR}$ are listed in Tables 1 and 3. The AICs of the models with the constraints are clearly superior to those without constraints. The results also show that the methodology is flexible for imposing constraints, and these constraints could be beneficial for capturing the underlying pavement mechanism and producing more reasonable models. Except for the parameters related to condition-maintenance effectiveness, the estimates of the parameters are close to those of the models without considering the constraints. Note that $h^{(5)}$ parameters are insignificant at the 95% level for the exponential and logarithmic $x - \mathrm{OVR}$ specifications because they are very close to zero, which fits the definition of insignificance and is consistent with the linear forms shown in the figure. The linear forms for the specifications are also caused by $h^{(5)}$ close to -1 for the polynomial $x - \mathrm{OVR}$ and $h^{(5)}$ close to 1 for the power specification. The model of polynomial $x - \mathrm{OVR}$ and $\mathrm{SN} - \mathrm{TRF}$ specifications is selected as the preferred model and listed in Eqs. (20) and (21) as an example. More calculation details of the model can be found in the appendix:

$$x_{i,t} = 0.991x_{i,t-1} + (-0.035 + 0.004\mathrm{SN}_{i,t})\mathrm{TRF}_{i,t-1} + (3.400 - 0.998x_{i,t-1} - 0.0005(x_{i,t-1})^2)\mathrm{OVR}_{i,t-1} + \omega_{i,t-1}, \omega_{i,t} \sim N(0, -0.181^2) \quad (20)$$

$$z_{i,t} = x_{i,t} + \xi_{i,t}, \xi_{i,t}^{\sim} N(0, 0.108^2) \quad (21)$$

Finally, the serial dependence and the normality of the preferred model were tested to see if the model was satisfactory. The test for serial independence of residuals is critical because it is the fundamental assumption of dynamic models. Although multivariate versions of tests should be used to test a multivariate model, they are not applicable because the total number of facilities (188) is much larger than the number of time periods (56). This structure of the data implies that the sample covariance matrices required in these tests are singular. Note that because this structure of data is rarely seen in the literature on dynamic modeling, a formal procedure that takes advantage of the data in the example is not available to the best of the author's knowledge. To address this problem, a Q-test for univariate time series data was conducted for individual facilities (Harvey 1990). The results show that most of the facilities (158 out of 188, or 84%) have serially independent residuals at the 95% significance level, although the interpretation of this percentage is unclear. Similarly, the Lilliefors test was conducted for individual facilities to check the normality of the residuals (Lilliefors 1967). Unlike the serial dependence, the assumption of normality is not necessary for modeling dynamic systems, but it is required by the Kalman filters and the associated model estimation procedure. The results show that only 55% (104 out of 188) of the facilities had normally distributed residuals at the 95% significance level.

Table 3. Estimation Results of Nonlinear Models with Constraints 1 and 2

$\mathrm{SN} - \mathrm{TRF}$	Polynomial		Polynomial		Polynomial		Polynomial	
$x - \mathrm{OVR}$	Polynomial		Exponential		Logarithm		Power	
Parameter	Estimate	t-stat	Estimate	t-stat	Estimate	t-stat	Estimate	t-stat
σ_ω	−0.181	−68.2	−0.181	−68.2	−0.181	−68.2	−0.181	−68.2
σ_ξ	0.108	41.1	0.108	41.1	0.108	41.1	0.108	41.1
$g^{(1)}$	0.991	1691.8	0.991	1691.9	0.991	1691.8	0.991	1691.9
$h^{(1)}$	−0.035	−2.5	−0.035	−2.6	−0.035	−2.6	−0.034	−2.5
$h^{(2)}$	**0.004**	**1.2**	**0.004**	**1.3**	**0.004**	**1.3**	**0.004**	**1.2**
$h^{(5)}$	−0.998	−43.9	**0.002**	**0.2**	**−0.002**	**−0.2**	1.004	55.7
AIC	−10.915		−10.915		−10.915		−10.916	

Note: Entries in bold-face type indicate that a parameter is insignificant at the 95% confidence level.

Note that the potential deviation from the assumption of normal residuals does not invalidate the model specification. However, this raises the possibility that the relatively low percentage could be a signal for biased estimates. One of the possible reasons for the relatively low percentage is that the normality assumption does not hold for the approximated errors after the linear transformation in the extended Kalman filter (EKF) takes place, as explained in the Appendix. Therefore, alternative filters that do not violate normality could be used to replace the EKF in the future to improve the estimation results.

Applicability to In-Service Road Data Set

The main purpose of the numerical example is to demonstrate the methodology, which requires long time series data and detailed maintenance records. The AASHO Road Test data set provides satisfactory data and was thus used in this research. However, the type of data that is collected from such highly controlled experiments are not readily available for in-service pavement networks. This subsection identifies the issues of applying the proposed methodology to in-service pavement data sets and discusses possible extensions for addressing these issues. The first issue is the small number of observations for each facility in the data sets for in-service roads. The small number of observations is disadvantageous for estimating dynamic models. The general rule for the number of observations in time series analysis provided by Box et al. (1970) is 50, although the technique has been used for much shorter time series and the estimated models were still satisfactory, e.g., examples discussed in Lorek and McKeown (1978). Note that a sample size of 50 is suggested for a single time series. In the context of pavement management, panel data that includes time series from multiple facilities are often used. Therefore, although the observations for each facility might be short, a large sample size is still available given the large number of facilities. Therefore, this rule concerning the minimum number of observations is not applicable to the proposed methodology. To the author's knowledge, there are no universal rules for the minimum numbers of observations for multivariate nonlinear dynamic models in state-space specifications, which are considered in the paper. A straightforward method is to estimate the model with the available data and then test and diagnose the estimated model to determine whether it is satisfactory. If the model results are not satisfactory, more data could be collected to improve the results. However, shorter time series are only one of the possible reasons for unsatisfactory estimation results, so different model specifications or other estimation approaches should also be tested.

Another important consideration in applying the proposed methodology to in-service pavements is the facility inspection interval. Long or irregular inspection intervals are often observed in practice, and special attention is required when the inspection intervals are short or uneven. The issue related to long inspection periods is discussed first. Note that long inspection intervals do not always cause problems as long as the number of observations for each facility is sufficient. However, they do influence the information that can be extracted from the data. For example, if the facilities are inspected annually, estimating the seasonal cycles of facility deterioration within a year based on the data is impossible. Next, when the inspection intervals are not equal, the problem of missing values will appear because the length for each time period is equal in the model formulation. Estimating dynamic models when values are missing is addressed by Durbin and Koopman (2001) for the standard Kalman filter and by Chu and Durango-Cohen (2007) for the square root filter, a Kalman filter variant. The basic concept is to remove the facilities whose condition measurements are not observed from the Kalman filter calculation and the log-likelihood function in the estimation procedure. This is done by changing the dimensions of dynamic models in every time period according to the data availability. Extending the proposed methodology to consider missing values using the same method seems straightforward and could be pursued in the future. However, the treatment will be more appropriate when the missing values are distributed randomly over time and among the facilities. If not, significant errors may arise for facilities or time periods that have too many missing values, leading to highly questionable estimation results.

The final main issue in the in-service pavement data set is the quality of the data. Clearly, the quality of data collected in the field would be poorer than that of data collected in a controlled experiment such as the AASHO Road Test. The benefit of the proposed methodology is that measurement errors of the facility condition can be considered directly. However, measurement errors might also exist in the exogenous variables such as the pavement strength and maintenance history. Thus, considering measurement errors for the explanatory variables would be useful and should be considered in the future. Another problem that occurs frequently in the databases for in-service roads is that maintenance history records are incomplete. That is, while maintenance activities took place, their types, amounts, and timings were not properly recorded. In this case, the techniques of *intervention analysis* in time series analysis could be used to address the problem partially. For example, Tsay (1988) proposes an approach to identify the level shift due to an external event in a data series. Identifying the timing, type, or amount of maintenance based on the history of facility condition using the techniques of intervention analysis seems promising and worth further study to be able to apply the proposed methodology to actual road networks more adequately.

Conclusion and Future Research

A methodology for estimating nonlinear dynamic performance models for transportation infrastructure management has been proposed. The dynamic models combine panel data, performance prediction, maintenance effectiveness, and variable interactions, none of which has been previously studied. It is known that maintenance effectiveness data depend on the facility condition. The most significant feature of the nonlinear specification is the condition-dependent maintenance effectiveness, which is necessary to predict the facility condition. As a result, a state-of-the-art optimization framework—optimal control—can be supported correctly with models estimated using empirical maintenance data.

It is clear that random sampled data would be ideal for the methodology. However, the data used in the numerical example suffer from the problem of selectivity bias. Although the methodology is not proposed specifically for the self-selected samples, physical constraints based on external sources can be imposed to address this data limitation. The constraints that extrapolate the overlay effectiveness for pavements in the unobserved range of PSI values are used in the estimation procedure. As a result, more reasonable models are generated. The constraints considered in the empirical examples include the effectiveness of maintenance action for the worst possible pavements and the best pavement condition that can be generated by the maintenance action. The estimation results show that formulating the condition-dependent maintenance effectiveness as a polynomial specification was the best approach based on the AIC. Another relationship, the traffic impact dependence on the structural design, was also shown to be best formulated as a polynomial specification. However, its effect was not significant due to the existence of autoregressive variables.

J. Infrastruct. Syst. 2013.19:85-98.

Finally, it is recognized that although the constraints stem from a trustworthy, independent data source, the constrained estimation approach adopted in the example depends heavily on engineer judgment. Since the complete samples were never observed, the correct values of the parameters are always uncertain. Thus, it should be expected that the estimated effectiveness function would still be biased to some degree for pavement in good condition. In the future, more treatments for correcting the bias should be considered to address this problem completely.

Appendix. Estimation Procedure

This Appendix provides an overview of the procedure used to estimate the parameters of the nonlinear dynamic models using panel data. In particular, data preparation, log-likelihood function, extended Kalman filter (EKF), and the optimization routine are described. The actual data and the preferred model [Eqs. (20) and (21)] in the numerical example are used to illustrate the procedure in more detail.

Data Preparation

The first step of the procedure is preparation of the data. As an illustrative example, the observation vectors for the data set used in the numerical example should be arranged as shown below. Due to space limitations, only a very small part of the actual data is displayed in these examples:

$$\mathbf{Z_1} = \begin{bmatrix} z_{1,1} \\ \vdots \\ z_{188,1} \end{bmatrix} = \begin{bmatrix} 3.8 \\ \vdots \\ 4.2 \end{bmatrix}, \qquad \mathbf{Z_2} = \begin{bmatrix} z_{1,2} \\ \vdots \\ z_{188,2} \end{bmatrix} = \begin{bmatrix} 3.7 \\ \vdots \\ 4.1 \end{bmatrix}, \ldots,$$

$$\mathbf{Z_{56}} = \begin{bmatrix} z_{1,56} \\ \vdots \\ z_{188,56} \end{bmatrix} = \begin{bmatrix} 4.0 \\ \vdots \\ 3.3 \end{bmatrix}$$

where the first element in each vector is the PSI measurements for Facility 1 at the corresponding time, and so on.

Similarly, the vectors of explanatory variables for the data set used in the numerical example should be arranged as follows:

$$\mathbf{A_1} = \begin{bmatrix} A_{1,1}^{(1)} \\ A_{1,1}^{(2)} \\ A_{1,1}^{(3)} \\ \vdots \\ A_{188,1}^{(1)} \\ A_{188,1}^{(2)} \\ A_{188,1}^{(3)} \end{bmatrix} = \begin{bmatrix} 3.04 \\ 0.00 \\ 0.00 \\ \vdots \\ 5.66 \\ 0.02 \\ 0.00 \end{bmatrix}, \qquad \mathbf{A_2} = \begin{bmatrix} A_{1,2}^{(1)} \\ A_{1,2}^{(2)} \\ A_{1,2}^{(3)} \\ \vdots \\ A_{188,2}^{(1)} \\ A_{188,2}^{(2)} \\ A_{188,2}^{(3)} \end{bmatrix} = \begin{bmatrix} 3.04 \\ 0.00 \\ 0.00 \\ \vdots \\ 5.66 \\ 0.24 \\ 0.00 \end{bmatrix}, \ldots,$$

$$\mathbf{A_{56}} = \begin{bmatrix} A_{1,56}^{(1)} \\ A_{1,56}^{(2)} \\ A_{1,56}^{(3)} \\ \vdots \\ A_{188,56}^{(1)} \\ A_{188,56}^{(2)} \\ A_{188,56}^{(3)} \end{bmatrix} = \begin{bmatrix} 3.04 \\ 0.00 \\ 0.00 \\ \vdots \\ 5.66 \\ 0.02 \\ 0.00 \end{bmatrix}$$

where the first three elements in each vector are SN, TRF, and OVR for Facility 1 at the corresponding time, and so on. The SN does not change over time, so its values are fixed in these vectors. Pavement 1 was not subjected to traffic, so its TRF values are all zeros; on the other hand, Pavement 188 was subjected to traffic application, so most of its TRF values are nonzero. The OVR variables are also zero for the two selected pavements because no overlay was applied to them.

The formulation of the preferred model in the numerical example is explicitly shown in Eqs. (22) and (23), which is equivalent to Eqs. (7) and (8). The error terms and their covariance matrices for the preferred model should be arranged as in Eqs. (24) and (25):

$$\mathbf{X_t} = \begin{bmatrix} x_{1,t} \\ \vdots \\ x_{188,t} \end{bmatrix}$$
$$= \begin{bmatrix} g^{(1)}x_{1,t-1} + (h^{(1)} + h^{(2)}\mathrm{SN}_{1,t-1} + h^{(3)}(\mathrm{SN}_{1,t-1})^2)\mathrm{TRF}_{1,t-1} + \left(3.4 + h^{(5)}x_{1,t-1} - \frac{3.4+h^{(5)}3.4}{(3.4)^2}(x_{1,t-1})^2\right)\mathrm{OVR}_{1,t-1} + \omega_{1,t-1} \\ \vdots \\ g^{(1)}x_{188,t-1} + (h^{(1)} + h^{(2)}\mathrm{SN}_{188,t-1} + h^{(3)}(\mathrm{SN}_{188,t-1})^2)\mathrm{TRF}_{188,t-1} + \left(3.4 + h^{(5)}x_{188,t-1} - \frac{3.4+h^{(5)}3.4}{(3.4)^2}(x_{188,t-1})^2\right)\mathrm{OVR}_{188,t-1} + \omega_{188,t-1} \end{bmatrix} \tag{22}$$

$$\mathbf{Z_t} = \begin{bmatrix} z_{1,t} \\ \vdots \\ z_{188,t} \end{bmatrix} = \begin{bmatrix} x_{1,t} + \xi_{1,t} \\ \vdots \\ x_{188,t} + \xi_{188,t} \end{bmatrix} \tag{23}$$

$$\mathbf{\Omega}_t = [\omega_{1,t} \quad \cdots \quad \omega_{188,t}]', \qquad \Sigma_{\Omega_t} = \begin{bmatrix} \Sigma_{\omega_t} & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \Sigma_{\omega_t} \end{bmatrix} \tag{24}$$

J. Infrastruct. Syst. 2013.19:85-98.

$$\Xi_t = [\xi_{1,t} \quad \cdots \quad \xi_{188,t}]', \qquad \Sigma_{\Xi_t} = \begin{bmatrix} \Sigma_{\xi_t} & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \Sigma_{\xi_t} \end{bmatrix} \tag{25}$$

Maximum Likelihood Estimation

Maximum likelihood estimation (MLE) was used because of the asymptotic properties of the resulting estimates. The observations in a time series analysis are assumed to follow a joint distribution, $P(\mathbf{X_1}, \mathbf{X_2}, \ldots, \mathbf{X_{T-1}}, \mathbf{X_T})$. The assumption required for MLE to estimate dynamic models is that Ξ_t and $\mathbf{\Omega}_t$ follow multivariate normal distributions. It follows that the conditional distribution of $\mathbf{X_t}$ given data up to $t-1$ is also normal. Based on this assumption, the log-likelihood function of the nonlinear state-space models can be expressed as Eq. (26). A detailed description of the derivation of the loglikelihood function was provided by Harvey (1990):

$$\log \mathcal{L} = -\frac{N(T-d)}{2}\log 2\pi - \frac{1}{2}\sum_{t=1+d}^{T}\log\ \det(\mathbf{B_t}) - \frac{1}{2}\sum_{t=1+d}^{T}\mathbf{V_t'}\mathbf{B_t^{-1}}\mathbf{V_t} \tag{26}$$

where $\mathbf{V_t}$ = one step prediction error at time t; $\mathbf{B_t}$ = prediction error variance at time t; N = total number of facilities; T = number of time periods; and d = cardinality of state vector of individual facilities.

The evaluation of the preceding function depends on the values of $\mathbf{V_t}$ and $\mathbf{B_t}$. These values can be generated by the Kalman filter described in the next subsection.

Extended Kalman Filter

The Kalman filter is an algorithm that calculates the optimal state estimates of state-space models. For nonlinear state-space models, the EKF is used most widely for its simplicity and computational efficiency (Julier and Uhlmann 2004). The EKF in Eqs. (27)–(36) modifies the linear Kalman filter and transforms the nonlinear function into a linear function around the values of the state vector being evaluated. More specifically, it replaces g_t with the Jacobian matrix of $\mathbf{F}$, $\mathcal{F}_t$ and Λ_t with the Jacobian matrix of $\mathbf{H}$, $\mathcal{H}_t$ as shown in Eqs. (32) and (29). Thus, the element in the ith row and jth column of $\mathcal{F}_t$ is the partial derivative of the ith element in $\mathbf{F}$ with respect to the jth element in $\mathbf{X_t}$. The element in the ith row and jth column of $\mathcal{H}_t$ is the partial derivative of the ith element in $\mathbf{H}$ with respect to the jth element in $\mathbf{X_t}$. The covariance matrices of $\mathbf{\Omega}$ and Ξ are also transformed using $\mathcal{L}_t$ and $\mathcal{M}_t$ as shown in Eqs. (33) and (30). Similarly, the element in the ith row and jth column of $\mathcal{L}_t$ is the partial derivative of the ith element in $\mathbf{F}$ with respect to the jth element in Ω_t, and the element in the ith row and jth column of $\mathcal{M}_t$ is the partial derivative of the ith element in $\mathbf{H}$ with respect to the jth element in Ξ. It is clear that if the errors are independent of the other variables, then the extension can be simplified because $\mathcal{L}_t = I_{Nd\times Nd}$ and $\mathcal{M}_t = I_{Nk\times Nk}$. All the preceding terms are evaluated by the values of $\mathbf{X_t}$ and thus are variable at different times, which explains the subscript t. Note that despite its simplicity and efficiency, the EKF is only an approximation. Linear approximation transforms the distribution of the error terms, and the assumption of normality may no longer hold for the approximated errors after the transformation (Tanizaki and Mariano 1994). Linear transformation could also lead to larger errors when the models are highly nonlinear (Simon 2006).

To illustrate the application of the algorithm, the corresponding equations for the preferred model in the numerical example are displayed in parentheses where applicable. Because the examples for some of the equations are simply the combinations of other examples, they are not included. Eq. (29) shows that $\mathcal{H}_t$ would be simply the identity matrix given Eq. (23), and Eq. (33) indicates that $\mathcal{L}_t$ would be the identity matrix given Eq. (22). The reason is that the error terms are additive in the preferred model, i.e., they do not have nonlinear relationship with other variables. Similarly, $\mathcal{H}_t$ would also be an identity matrix in Eq. (30) because the PSI measurement has a linear relationship with its true value, as shown in Eq. (23). As a result, no approximation is required for these variables, and the identity matrices have no effect on the transformation of the algorithm. On the other hand, the facility condition has a nonlinear relationship with other variables, as shown in Eq. (22). Therefore, $\mathcal{F}_t$ in Eq. (32) is used to transform the nonlinear relationship into a linear relationship around $\mathbf{X_t}$ in Eqs. (34) and (36).

Given a set of values for the parameters, including g_t, h_t, Λ_t, and $\Sigma_{\mathbf{\Omega_t}}$ and Σ_{Ξ_t}, the Kalman filter can be executed and the log-likelihood function evaluated for the given parameter values. Note that initializing the Kalman filter requires the initial values of $\mathbf{X_1}$ and $\mathbf{P_1}$. If these values are known, they should be used. If they are unknown, then $\mathbf{X_1} = [0 \quad \cdots \quad 0]'$ and $\mathbf{P_1} = [10^6 \quad \cdots \quad 10^6]$ can be used, where 10^6 is an arbitrarily large value. These indicate that the initial values of $\mathbf{X_1}$ and $\mathbf{P_1}$ are unknown and will be estimated by the first few observations in the time series (Harvey 1990).

Extended Kalman Filter Algorithm

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

$$\hat{Z}_t = \mathbf{H}(\Lambda_t, \mathbf{X_t}, 0)(= \mathbf{X_t}) \tag{27}$$

$$\mathbf{V_t} = \mathbf{Z_t} - \mathbf{\hat{Z}_t} \tag{28}$$

$$\mathcal{H}_t = \left.\frac{\partial \mathbf{H}(\Lambda_t, \mathbf{X_t}, \Xi_\mathbf{t})}{\partial \mathbf{X_t}}\right|_{\mathbf{X_t}} = \left(\begin{bmatrix} \left.\frac{\partial z_{1,t}}{\partial x_{1,t}}\right|_{\mathbf{X_t}} & \cdots & \left.\frac{\partial z_{1,t}}{\partial x_{188,t}}\right|_{\mathbf{X_t}} \\ \vdots & & \vdots \\ \left.\frac{\partial z_{188,t}}{\partial x_{1,t}}\right|_{\mathbf{X_t}} & \cdots & \left.\frac{\partial z_{188,t}}{\partial x_{188,t}}\right|_{\mathbf{X_t}} \end{bmatrix} = \begin{bmatrix} 1 & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & 1 \end{bmatrix} \right) \tag{29}$$

$$\mathcal{M}_t = \left.\frac{\partial \mathbf{H}(\Lambda_t, \mathbf{X_t}, \Xi_t)}{\partial \Xi_t}\right|_{\mathbf{X_t}} = \left(\begin{bmatrix} \left.\frac{\partial z_{1,t}}{\partial \xi_{1,t}}\right|_{\mathbf{X_t}} & \cdots & \left.\frac{\partial z_{1,t}}{\partial \xi_{188,t}}\right|_{\mathbf{X_t}} \\ \vdots & & \vdots \\ \left.\frac{\partial z_{188,t}}{\partial \xi_{1,t}}\right|_{\mathbf{X_t}} & \cdots & \left.\frac{\partial z_{188,t}}{\partial \xi_{188,t}}\right|_{\mathbf{X_t}} \end{bmatrix} = \begin{bmatrix} 1 & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & 1 \end{bmatrix} \right) \tag{30}$$

$$\mathbf{B_t} = \mathcal{H}_t\mathbf{P_t}H_t' + \mathcal{M}_t\Sigma_\Xi\mathcal{M}_t' \tag{31}$$

J. Infrastruct. Syst. 2013.19:85-98.

$$\mathcal{F}_t = \frac{\partial \mathbf{F}(g_t, \mathbf{X_t}, h_t, \mathbf{A_t}, \mathbf{\Omega}_t)}{\partial \mathbf{X_t}}\bigg|_{\mathbf{X_t}} = \left(\begin{bmatrix} \frac{\partial x_{1,t+1}}{\partial x_{1,t}}\Big|_{\mathbf{X_t}} & \cdots & \frac{\partial x_{1,t+1}}{\partial x_{188,t}}\Big|_{\mathbf{X_t}} \\ \vdots & & \vdots \\ \frac{\partial x_{188,t+1}}{\partial x_{1,t}}\Big|_{\mathbf{X_t}} & \cdots & \frac{\partial x_{188,t+1}}{\partial x_{188,t}}\Big|_{\mathbf{X_t}} \end{bmatrix} \right.$$

$$\left. = \begin{bmatrix} g^{(1)} + h^{(5)} - \frac{3.4 + h^{(5)}3.4}{(3.4)^2}(2x_{1,t})\mathrm{OVR}_{1,t} & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & g^{(1)} + h^{(5)} - \frac{3.4 + h^{(5)}3.4}{(3.4)^2}(2x_{188,t})\mathrm{OVR}_{188,t} \end{bmatrix} \right) \tag{32}$$

$$\mathcal{L}_t = \frac{\partial \mathbf{F}(g_t, \mathbf{X_t}, h_t, \mathbf{A_t}, \mathbf{\Omega}_t)}{\partial \mathbf{\Omega}_t}\bigg|_{\mathbf{X_t}} = \left(\begin{bmatrix} \frac{\partial x_{1,t+1}}{\partial \omega_{1,t}}\Big|_{\mathbf{X_t}} & \cdots & \frac{\partial x_{1,t+1}}{\partial \omega_{188,t}}\Big|_{\mathbf{X_t}} \\ \vdots & & \vdots \\ \frac{\partial x_{188,t+1}}{\partial \omega_{1,t}}\Big|_{\mathbf{X_t}} & \cdots & \frac{\partial x_{188,t+1}}{\partial \omega_{188,t}}\Big|_{\mathbf{X_t}} \end{bmatrix} = \begin{bmatrix} 1 & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & 1 \end{bmatrix} \right) \tag{33}$$

$$\mathbf{K_t} = \mathcal{F}_t \mathbf{P_t} \mathcal{H}_t' \mathbf{B_t}^{-1} \tag{34}$$

$$\mathbf{X_{t+1}} = \mathbf{F}(g_t, \mathbf{X_t}, h_t, \mathbf{A_t}, 0) + \mathbf{K_t} \mathbf{V_t} \tag{35}$$

$$\mathbf{P_{t+1}} = \mathcal{F}_t \mathbf{P_t} (\mathcal{F}_t - \mathbf{K_t} \mathcal{H}_t)' + \mathcal{L}_t \Sigma_{\mathbf{\Omega}} \mathcal{L}_t' \tag{36}$$

where $\mathbf{P_t}$ = covariance matrix of state estimates at time t.

Optimization Routine

The estimation problem consists in identifying parameters that maximize the log-likelihood function, i.e., Eq. (26). The estimation problem was solved with the nonlinear unconstrained optimization routine in MATLAB. The output of the estimation procedure is the parameters to be estimated, including the transition parameter g, the parameters for the explanatory variable h, the measurement parameter Λ, and the covariance matrices $\Sigma_{\mathbf{\Omega}}$ and $\Sigma_{\mathbf{\Xi}}$. Note that, as discussed in the numerical example, the models are assumed to be time-homogeneous to reduce the number of parameters and thus the parameters are not indexed by time. Finally, the complete estimation procedure is summarized as follows:

1. Collect the observations $\mathbf{Z_t}$ and $\mathbf{A_t}$, $t = 1, \ldots, T$, and formulate the dynamic model in the state-space form of Eqs. (7) and (8).
2. The maximum likelihood estimators of g, h, Λ, Σ_Ω, and $\Sigma_{\mathbf{\Xi}}$ are given by arg max arg max $\log \mathcal{L}$, where $\log \mathcal{L}$ is evaluated in the following steps:
   a. For a set of parameters (g, h, Λ, $\Sigma_{\mathbf{\Omega}}$, $\Sigma_{\mathbf{\Xi}}$) and the given observations ($\mathbf{Z_t}$), apply the EKF algorithm [Eqs. (27)–(36)] to the dynamic model to obtain $\mathbf{V_t}$ and $\mathbf{B_t}$.
   b. Calculate $\log \mathcal{L}$ in Eq. (26) with $\mathbf{V_t}$, $\mathbf{B_t}$.

Acknowledgments

This work was supported by the National Science Council of Taiwan through research grant NSC 97-2218-E-008-013.

References

AASHTO. (1993). *AASHTO guide for design of pavement structures 1993*, Washington, DC.

Al-Mansour, A. I., Sinha, K. C., and Kuczek, T. (1994). "Effects of routine maintenance on flexible pavement condition." *J. Transp. Eng.*, 120(1), 65–73.

Archilla, A. (2006). "Repeated measurement data analysis in pavement deterioration modeling." *J. Infrastruct. Syst.*, 12(3), 163–173.

Archilla, A. R., and Madanat, S. (2000). "Development of a pavement rutting model from experimental data." *J. Transp. Eng.*, 126(4), 291–299.

Archilla, A. R., and Madanat, S. M. (2001). "Estimation of rutting models by combining data from different sources." *J. Transp. Eng.*, 127(5), 379–389.

Ben-Akiva, M., and Lerman, S. (1985). *Discrete choice analysis: Theory and application to travel demand*, The MIT Press, Cambridge, MA.

Ben-Akiva, M., and Ramaswamy, R. (1993). "An approach for predicting latent infrastructure facility deterioration." *Transp. Sci.*, 27(2), 174–193.

Box, G., Jenkins, G., and Reinsel, G. (1970). *Time series analysis: Forecasting and control*, Holden-Day, San Francisco.

Carey, M., and Revelli, R. (1986). "Constrained estimation of direct demand functions and trip matrices." *Transp. Sci.*, 20(3), 143–152.

Chu, C.-Y., and Durango-Cohen, P. L. (2007). "Estimation of infrastructure performance models using state-space specifications of time series models." *Transp. Res. Part C*, 15(1), 17–32.

Chu, C.-Y., and Durango-Cohen, P. L. (2008a). "Estimation of dynamic performance models for transportation infrastructure using panel data." *Transp. Res. Part B*, 42(1), 57–81.

Chu, C.-Y., and Durango-Cohen, P. L. (2008b). "Incorporating maintenance effectiveness in the estimation of dynamic infrastructure performance models." *Comput.-Aided Civ. Infrastr. Eng.*, 23(3), 174–188.

Durango-Cohen, P. L. (2007). "A time series analysis framework for transportation infrastructure management." *Transp. Res. Part B*, 41(5), 493–505.

Durbin, J., and Koopman, S. J. (2001). *Time series analysis by state space methods*, Oxford Univ. Press, New York.

Harvey, A. C. (1990). *Forecasting, structural time series models and the Kalman filter*, Cambridge Univ. Press, New York.

Hauser, J., and Shugan, S. (1980). "Intensity measures of consumer preference." *Oper. Res.*, 28(2), 278–320.

Highway Research Board. (1962). "The AASHO road test". *Special Rep. 61A-E*, National Academy of Science, National Research Council, Washington, DC.

J. Infrastruct. Syst. 2013.19:85-98.

Julier, S., and Uhlmann, J. (2004). "Unscented filtering and nonlinear estimation." *Proc. IEEE*, 92(3), 401–422.

Labi, S. (2004). "Effectiveness of highway pavement seal coating treatments." *J. Transp. Eng.*, 130(1), 14–23.

Labi, S., and Sinha, K. C. (2003). "Measures of short-term effectiveness of highway pavement maintenance." *J. Transp. Eng.*, 129(6), 673–683.

Lilliefors, H. (1967). "On the Kolmogorov-Smirnov test for normality with mean and variance unknown." *J. Am. Stat. Assoc.*, 62(318), 399–402.

Livneh, M. (1998). "Simplified formulation of virgin and overlaid pavement deterioration models." *J. Infrastruct. Syst.*, 4(2), 63–70.

Lorek, K., and McKeown, J. (1978). "The effect on predictive ability of reducing the number of observations on a time-series analysis of quarterly earnings data." *J. Account. Res.*, 16(1), 204–214.

Lytton, R. L. (1987). "Concepts of pavement performance prediction and modeling." *Proc., 2nd North American Conf. on Managing Pavements*, Vol. 2, Ontario Ministry of Transportation and Communications, Toronto, Canada, 2.3–2.19.

Madanat, S., and Ben-Akiva, M. (1994). "Optimal inspection and repair policies for infrastructure facilities." *Transp. Sci.*, 28(1), 55–62.

Madanat, S., and Mishalani, R. (1998). "Selectivity bias in modeling highway pavement maintenance effectiveness." *J. Infrastruct. Syst.*, 4(3), 134–137.

N.D. Lea International. (1995). *Modelling road deterioration and maintenance effects in HDM-4*, Asian Development Bank, Manila, Philippines.

Ouyang, Y., and Madanat, S. (2006). "An analytical solution for the finite-horizon pavement resurfacing planning problem." *Transp. Res. Part B*, 40(9), 767–778.

Paterson, W. D. O. (1986). "International roughness index: Relationship to other measures of roughness and riding quality." *Transportation Research Record 1084*, Transportation Research Board, Washington, DC, 49–59.

Paterson, W. D. O. (1987). *Road deterioration and maintenance effects: Models for planning and management*, Johns Hopkins Univ. Press, Baltimore.

Paterson, W. D. O. (1990). "Quantifying the effectiveness of pavement maintenance and rehabilitation." *Proc., 6th Road Engineering Association of Asia and Australia Conf.*, Road Engineering Association of Asia and Australasia (REAAA), Kuala Lumpur, Malaysia.

Prozzi, J. A., and Hong, F. (2010). "Average treatment effect for modeling maintenance work." *J. Infrastruct. Syst.*, 16, 122–128.

Prozzi, J. A., and Madanat, S. M. (2003). "Incremental nonlinear model for predicting pavement serviceability." *J. Transp. Eng.*, 129(6), 635–641.

Prozzi, J. A., and Madanat, S. M. (2004). "Development of pavement performance models by combining experimental and field data." *J. Infrastruct. Syst.*, 10(1), 9–22.

Simon, D. (2006). *Optimal state estimation: Kalman, H_∞ and nonlinear approaches*, Wiley-Interscience, Hoboken, NJ.

Small, K. A., and Winston, C. (1988). "Optimal highway durability." *Am. Econ. Rev.*, 78(3), 560–569.

Srinivasan, V., Jain, A., and Malhotra, N. (1983). "Improving predictive power of conjoint analysis by constrained parameter estimation." *J. Market. Res.*, 20(4), 433–438.

Swait, J., and Ben-Akiva, M. (1986). "Constraints on individual travel behavior in a Brazilian City." *Transportation Research Record 1085*, Transportation Research Board, Washington, DC, 75–85.

Tanizaki, H., and Mariano, R. (1994). "Prediction, filtering and smoothing in non-linear and non-normal cases using Monte Carlo integration." *J. Appl. Econometr.*, 9(2), 163–179.

Tsay, R. (1988). "Outliers, level shifts, and variance changes in time series." *J. Forecast.*, 7(1), 1–20.

Watanabe, T., Harral, C. G., Paterson, W. D. O., Dhareshwar, A. M., Bhandara, A., and Tsunokawa, K. (1987). *The highway design and maintenance standards model*, Johns Hopkins Univ. Press, Baltimore.

J. Infrastruct. Syst. 2013.19:85-98.