Comparison of Markov Chain and Semi-Markov Models for Crack Deterioration on Flexible Pavements

Omar Thomas, M.ASCE[1]; and John Sobanjo, M.ASCE[2]

Abstract: There is a growing demand to preserve transportation infrastructure utilizing limited funds, and the modeling of flexible pavement deterioration has become an integral component of any pavement preservation model. Markov chains have been used to model the performance of pavements in various pavement management systems (PMSs). The Markov property may be considered restrictive when modeling the deterioration of transportation assets, primarily because of the "memoryless" property and assumption of exponential distribution for sojourn times in the condition states. This paper outlines a semi-Markov model for modeling pavement deterioration in which the sojourn time in each condition state is assumed to follow a Weibull distribution and, thus, is more flexible than the traditional Markov chain model. The semi-Markov model does not possess the memoryless property if the sojourn time distribution is not exponential. Monte Carlo simulations are generated for the deterioration of flexible pavements over time based on both the traditional Markov chain model and the proposed semi-Markov model. The results of the work show that in some cases the semi-Markov model appears to be superior to the Markov chain model in modeling the actual deterioration patterns of the flexible pavements. **DOI: 10.1061/(ASCE)IS.1943-555X.0000112.**

CE Database subject headings: Pavement management; Markov process; Deterioration; Cracking; Maintenance; Model studies.

Author keywords: Pavement management; Markov chain; Semi-Markov; Pavement deterioration; Asset management; Highway maintenance.

Introduction

The preservation of flexible (asphalt) pavements involves developing models that are able to estimate the performance of pavement networks. A Markov chain is a stochastic model that has been used to model the performance of pavements in a number of pavement management systems (PMSs), such as the Arizona Department of Transportation (ADOT) Network Optimization System (NOS) (Wang 1992; Wang et al. 1994). Wang et al. (1994) outlined an approach to generating transition probabilities of a Markov chain model based on current pavement performance data. Nasseri et al. (2009) applied a Markov chain to investigate the crack histories of flexible pavements and to determine the cause of rapid deterioration of surface cracks. Markov chains have also been used to model the deterioration of other infrastructures, including bridge elements, storm-water pipes, and wastewater pipes (Golabi et al. 1993; Micevski et al. 2002; Baik et al. 2006; Mishalani and Madanat 2002). The advantage of stochastic models is that they are able to capture physical and inherent uncertainty, model uncertainty, and statistical uncertainty while predicting the future performance of infrastructure facilities (Lounis and Mirza 2001). It is expected that the use of semi-Markov processes in modeling the deterioration of transportation infrastructure will still maintain the advantages of stochastic models while addressing some of the limitations that arise from a purely Markov chain model.

The use of semi-Markov processes for modeling the crack performance of flexible pavements was mentioned by Yang et al. (2005, 2009) as a precursor to outlining how recurrent Markov chains can be used to model the crack performance of flexible pavements. Semi-Markov processes have been used in deterioration models for other assets such as bridge elements and transformers (Ng and Moses 1998; Sobanjo 2011; Black et al. 2005a, b). The main difference between semi-Markov processes and Markov chains is that a Markov chain assumes that the sojourn time in one state before transitioning to another follows an exponential distribution for continuous time, whereas for a semi-Markov process the sojourn time can have any continuous-time distribution (Ng and Moses 1998; Howard 1971). The sojourn time is the time from when a unit of pavement segment first enters a particular state to the time it enters another state.

In this paper the states are based on the condition level of the pavement and are henceforth referred to as condition states. Also, since the Weibull distribution is flexible in modeling time to deterioration or fatigue data (Ross 2002), it is assumed that the sojourn times in each condition state for the proposed semi-Markov model follow a Weibull distribution. Weibull distributions have been used by other researchers in modeling the deterioration of assets (Black et al. 2005a, b; Sobanjo 2011; Sobanjo and Thompson 2001).

In this study, Monte Carlo simulation was used to model the deterioration of flexible pavements using transition probabilities derived from both Markov chain and semi-Markov deterioration models. The results are compared against the actual pavement deterioration behavior obtained from historical pavement condition

[1]Lecturer, Gildart Haase School of Computer Sciences and Engineering, Fairleigh Dickinson Univ., Metropolitan Campus, 1000 River Rd., T-MU1-01, Teaneck, NJ 07666 (corresponding author). E-mail: o_thomas@fdu.edu

[2]Associate Professor, Dept. of Civil and Environmental Engineering, Florida A&M Univ.–Florida State Univ. (FAMU-FSU), College of Engineering, 2525 Pottsdamer St., Room 129, Tallahassee, FL 32310. E-mail: sobanjo@eng.fsu.edu

Note. This manuscript was submitted on October 13, 2011; approved on May 16, 2012; published online on May 19, 2012. Discussion period open until November 1, 2013; separate discussions must be submitted for individual papers. This paper is part of the ***Journal of Infrastructure Systems***, Vol. 19, No. 2, June 1, 2013. © ASCE, ISSN 1076-0342/2013/2-186-195/$25.00.

data. The transition probabilities for do-nothing actions derived from the semi-Markov deterioration model takes into consideration the probability of transitions of the embedded Markov chain and the sojourn time in a particular condition state before transitioning to another. The Markov chain approach, as outlined by Wang et al. (1994), was applied in this study to determine the general transition probabilities for do-nothing actions between condition states, where do-nothing action means that there was no intervention in the deterioration process. It is assumed that once an asset falls out of a condition state, that condition state is not visited again. In reality, state highway agencies carry out routine maintenance activities that do not reflect significant changes in conditions and may not be captured from the observation of the actual condition data.

A comparison was made between the Markov chain model, semi-Markov model, and actual pavement deterioration behavior in predicting a decrease in condition state over time. Based on the results of the simulation, the survival curve for pavement segments remaining in the best condition was also determined. The methodology outlined in this paper can be used in conjunction with additional engineering properties by categorizing pavement segments with similar engineering properties. The use of statistical tools, such as semi-Markov modeling, which takes into consideration the age of an infrastructure, can provide insights into the expected behavior of transportation structures at the network level.

Methodology

Data Description

The Florida Department of Transportation (FDOT) uses a pavement condition rating (PCR), on a 0–10 scale, as a measure of the overall condition of a segment of pavement. This PCR value is the minimum of the following three indices, which are also measured on 0–10 scales: (1) crack index; (2) rut index; and (3) ride index. For all the indices, 10 represents the best condition and 0 the worst. The FDOT now uses a multipurpose survey vehicle (MPSV) that scans images to evaluate highway conditions. The FDOT PMS, called the Florida Analysis System for Targets (FAST), is a statistical analysis system (SAS)-based software that is used to predict future pavement conditions. In FAST, regression equations based on the historical performance of pavements in a particular district are used to predict the performance of pavements within that same district (Dietrich 2010). The crack index (CRK) is the most critical indicator and typically governs the overall roadway condition. The CRK is obtained by visual inspections carried out by a survey crew and inspection vehicle, which collects data as it traverses a section of pavement (Yang 2004). It is known that the accurate prediction of crack condition is integral to any PMS (Yang et al. 2005, 2009; Yang 2004).

For the purpose of this study, the Markov chain and semi-Markov models were based on the analysis of the CRK values of flexible pavements extracted from the FDOT pavement condition database. The data were organized such that pavement conditions were tracked from the year they became new or overlaid, regardless of the thickness of the asphalt layer, and while the conditions in the subsequent years either remained the same or dropped to a lower value. For this study, a network-level analysis of pavement condition using CRKs as a measure of the condition was done. A total of 1,580 pavement segments were selected from the FDOT pavement condition database containing 8,597 pavement segments. These pavement segments were selected to ensure that all pavements were flexible (asphalt concrete) and that there were no missing data related to the following attributes for the years 1986–2005: beginning mile post, ending mile post, year overlaid, crack index (CRK), rut index (RUT), ride index (RIDE), average daily traffic (ADT), pavement condition rating (PCR), and truck percentage. From the 1,580 pavement segments selected, the ADT ranged from 500 to 183,500 and the truck percentage ranged from 0.68 to 59.28. The pavement segments in the database consist of interstates and state roads throughout Florida.

Training and Test Data

Of the 1,580 selected pavement segments, 50% were sampled randomly and used as training data to develop the Markov chain and semi-Markov models. The results from the models were compared against the actual behavior of the remaining 790 pavement segments, obtained from the test data. To reflect solely do-nothing actions, condition data that reflected improvements to pavement segments were filtered out, that is, for both the training and test data, the tracking was stopped at the point where the CRK either increased or at the year at which the study ended. For this study, seven condition states were formulated as follows:

- $9.5 \leq \text{CRK} \leq 10$: Condition State 10
- $8.5 \leq \text{CRK} < 9.5$: Condition State 9
- $7.5 \leq \text{CRK} < 8.5$: Condition State 8
- $6.5 \leq \text{CRK} < 7.5$: Condition State 7
- $5.5 \leq \text{CRK} < 6.5$: Condition State 6
- $4.5 \leq \text{CRK} < 5.5$: Condition State 5
- $\text{CRK} < 4.5$: Condition State 4

Markov Chain Model

A stochastic process, known as a Markov chain (Ross 1996), can be described as follows: if $X_n = i$ describes a process such that the process is in state i at time n, and the process in state i has a fixed probability $P_{i,j}$ of being in state j after a transition, then

$$P\{X_{n+1} = j | X_n = i, X_{n-1} = i_{n-1}, \ldots, X_0 = i_0\} = P_{i,j} \qquad (1)$$

for all states $i_0, i_1, \ldots, i_{n-1}, i, j$ and all $n \geq 0$.

The following equation, obtained from Wang et al. (1994), can be used to generate transition probabilities for the Markov chain model:

$$p_{i,j}(a_k) = \frac{m_{i,j}(a_k)}{m_i(a_k)} \qquad (2)$$

for $i, j = 10, 9, 8, 7, 6, 5$, and 4, where k = do-nothing action (in this case); $P_{i,j}(a_k)$ = transition probability from condition state i to j after action k is taken; $m_{i,j}(a_k)$ = total number of miles of pavement for which the condition state prior to action k was i and the condition state after the action k was j; $m_i(a_k)$ = total number of miles of pavement for which the condition state prior to action k was i.

For this study, a *MATLAB* software program was used to perform the calculations of Eq. (2) for do-nothing actions, using CRKs of 790 pavement segments from the training data. This resulted in the generation of a set of transition probabilities that is used to model the yearly transitions. Details of the simulation are outlined later in the paper.

Semi-Markov Model

To describe the semi-Markov process, consider a stochastic process with states 0, 1, 2, ..., which is such that whenever it enters state i, $i > 0$, then (1) it will enter the next state j with probability p_{ij}, i, $j > 0$ and (2) given that the next state is j, the sojourn time from i to j has distribution h_{ij}. As mentioned previously, a Weibull distribution was assumed for the sojourn time distribution. The probability

J. Infrastruct. Syst. 2013.19:186-195.

density function of the Weibull distribution is defined by (Billington and Allan 1983; Tobias and Trindade 1995):

$$h(t) = \frac{\beta}{\alpha}\left(\frac{t}{\alpha}\right)^{\beta-1} e^{-(\frac{t}{\alpha})^{\beta}} \tag{3}$$

where α and β = scale and shape parameters, respectively; and t = number of years each unit of a mile of pavement segment takes to sojourn from one condition state to another.

The method of maximum likelihood estimation (MLE) is used to estimate the Weibull distribution parameters. The pavement condition records are used to estimate the frequency distribution of the lengths of time a pavement segment spends in a particular condition state before transitioning to a lower condition state, from which the MLEs of the parameters are determined. The maximum likelihood is one of the most popular techniques in statistics for deriving estimators (Casella and Berger 2001), and the derivation of the equations used to estimate the scale and shape parameters of the Weibull distribution can be found in numerous statistical texts, including Birolini (2007).

To understand how a semi-Markov model can be developed by knowing the sojourn times in a particular condition state before transitioning, one may consider the semi-Markov kernel in the form shown in Eq. (4). Ibe (2009) defines the one-step transition probability $Q_{i,j}(t)$ of the semi-Markov process as

$$Q_{i,j}(t) = P[X_{n+1} = j, T_n \leq t | X_n = i] \qquad t \geq 0 \tag{4}$$

where $Q_{i,j}(t)$ = conditional probability that the process will be in state j next given that it is currently in state i and the waiting time in the current state i, T_n, is no more than t. It also follows that

$$Q_{i,j}(t) = p_{i,j} H_{i,j}(t) \tag{5}$$

where $p_{i,j}$ = transition probability of the embedded Markov chain, and the sojourn (holding) time cumulative probability,

$$H_{i,j}(t) = P[G_n \leq t | X_n = i, X_{n+1} = j] \tag{6}$$

where G_n = time the process spends in i before making a transition to j.

Semi-Markov Process

Howard (1971) presented the following formulation to determine the probability that a continuous-time semi-Markov process will be in state j at time n given that it entered state i at time zero:

$$\phi_{ij}(n) = \delta_{ij}{}^{>}w_i(n) + \sum_{k=1}^{N} p_{ik} \int_0^n h_{ik}(m)\phi_{kj}(n-m)$$
$$i, j = 1, 2, \ldots, N; \quad n = 0, 1, 2, \ldots \tag{7}$$

$$\delta_{ij} = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases} \tag{8}$$

where $\phi_{ij}(n)$ = probability that a continuous-time semi-Markov process will be in state j at time n given that it entered state i at time n = 0 and is called the interval transition probability from state i to state j in interval $(0, n)$; ${}^{>}w_i(n)$ = probability that the process will leave its starting state i at a time greater than n.

The second term in Eq. (7) describes the probability of the sequence of events, in which the process makes an initial transition from state i to some state k at some time m and thereafter proceeds from state k to state j in the remaining time n–m. To consider all possible scenarios, the probability is summed over all states k to which the first transition could have been made and over all times of the initial transition, m, between 1 and n. p_{ik} = probability of transitioning from i to k; and $h_{ik}(m)$ = probability distribution of the sojourn time from i to k at time m. The matrix formulation of Eq. (7) is

$$\Phi(n) = {}^{>}W(n) + \int_0^n [P\square H(m)]\Phi(n-m) \qquad n = 0, 1, 2, \ldots \tag{9}$$

The box notation ($\square$) means that the ijth element of Matrix P is multiplied by the ijth element of Matrix H. Also let

$$C(m) = P\square H(m) \tag{10}$$

and $C(m)$ is defined as the core matrix (Howard 1971). The elements of $C(m)$ are $c_{ij}(m) = p_{ij}h_{ij}(m)$, where p_{ij} is the transition probability of the embedded Markov chain; and $h_{ij}(m)$ = probability distribution of the sojourn time in state i before transitioning to j at time m.

It is also assumed that the sojourn time in condition state i before transitioning to j is the time from when the pavement segment first entered condition state i to the time when the pavement segment first entered condition state j. Instead of determining the interval transition probabilities for the interval (0, n), the conditional transition probabilities for each yearly interval $(m-1,\ m]$ for $m = 1, 2, \ldots, n$ are first determined and multiplied by each other to approximate the transition probabilities for the interval (0, n). It is assumed that only a single transition can occur in a year. In other words, let

$$\Phi^{0,n}(n) = \Phi^{0,1} \cdot \Phi^{1,2} \cdot \ldots \cdot \Phi^{n-1,n} \tag{11}$$

where $\Phi^{m-1,m}$ is a 1-year single transition probability matrix from time $m-1$ to m (i.e., mth interval), $m = 1, 2, \ldots, n$. If it is assumed that condition states can have a maximum drop of two condition states, then the interval transition probability for the first year results in

$$\Phi^{0,m}(m) = \begin{pmatrix} 1-\sum_{j=4}^{9} p_{10,j}H_{10,j}(m) & p_{10,9}H_{10,9}(m) & p_{10,8}H_{10,8}(m) & 0 & 0 & 0 & 0 \\ 0 & 1-\sum_{j=4}^{8} p_{9,j}H_{9,j}(m) & p_{9,8}H_{9,8}(m) & \cdots & 0 & 0 & 0 \\ 0 & 0 & 1-\sum_{j=4}^{7} p_{8,j}H_{8,j}(m) & \cdots & \cdots & 0 & 0 \\ 0 & 0 & 0 & \ddots & \cdots & \cdots & 0 \\ 0 & 0 & 0 & 0 & \ddots & \cdots & p_{6,4}H_{6,4}(m) \\ 0 & 0 & 0 & 0 & 0 & \ddots & p_{5,4}H_{5,4}(m) \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \tag{12}$$

J. Infrastruct. Syst. 2013.19:186-195.

where $m = 1$. Based on the observation of the actual condition data, very few pavement segments have drops more than two levels in the condition state. As a result, it would be too difficult to establish sojourn time distributions in condition states that experienced a drop in excess of two levels, and therefore this was not done in this model. To determine the transition probability matrix for the subsequent intervals, a different formulation is used in which it is assumed that the sojourn time is left truncated at the start of each interval (Cleves et al. 2001; Castillo et al. 2005). Therefore, for interval $(m-1, m)$ the cumulative distribution of the sojourn time in the interval can be described as

$$H_{i,j_{T|T>n-1}}(t) = \frac{H_{i,j_T}(t) - H_{i,j_T}(m-1)}{1 - H_{i,j_T}(m-1)} \qquad m-1 < t \leq m \quad (13)$$

At $t = m$ the cumulative distribution of the sojourn time becomes

$$H_{i,j_{T|T>n-1}}(m) = \frac{H_{i,j_T}(m) - H_{i,j_T}(m-1)}{1 - H_{i,j_T}(m-1)} \quad (14)$$

Therefore, the transition probability for interval $(m-1, m)$ is given by

$$\Phi^{m-1,m}(m) = \begin{pmatrix} 1 - \sum_{j=4}^{9} p_{10,j} H_{10,j_{T|T>m-1}}(m) & p_{10,9} H_{10,9_{T|T>m-1}}(m) & p_{10,8} H_{10,8_{T|T>m-1}}(m) & 0 & 0 & 0 & 0 \\ 0 & \ddots & \cdots & \cdots & 0 & 0 & 0 \\ 0 & 0 & \ddots & \cdots & \cdots & 0 & 0 \\ 0 & 0 & 0 & \ddots & \cdots & \cdots & 0 \\ 0 & 0 & 0 & 0 & \ddots & \cdots & \cdots \\ 0 & 0 & 0 & 0 & 0 & \ddots & \cdots \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \quad (15)$$

Eq. (14) was used previously to determine the transition probabilities for one-step transitions, state-based deterioration models (Black et al. 2005a, b), where the transition probability of the embedded Markov chain was assumed to be 1. This therefore means that the probability obtained from Eq. (14) can be considered as describing the probability associated with the sojourn time to the end of the period, given that it survived up to the start of the period.

Sojourn Time of Pavements between Condition States

For the proposed semi-Markov model, the sojourn time distribution for each unit of a mile of pavement segment was analyzed to an accuracy of one decimal place. A *MATLAB* program was written to organize and analyze the data to estimate the parameters of the sojourn time distributions of pavement segments in each condition state based on the following steps:

1. The 790 pavement segments, used as training data, were imported with the following attributes: beginning mile post, ending mile post, year overlaid, and CRKs over the course of 20 years (1986–2005).
2. The length of each segment of a mile was determined.
3. The yearly decreases in CRKs for each segment subsequent to the year being newly constructed or overlaid were extracted, such that a series of decreasing CRKs for each segment could be used to model do-nothing actions on the pavement segments over time.
4. The pavement segments were assigned to condition states over time, as outlined earlier in this paper.
5. At some point in time, all the pavement segments tracked were either just becoming new or entered the current condition state from a higher condition state. If a unit of pavement segment exited the current condition state to a lower condition state at a known time, then the sojourn time of that unit of pavement segment in the current condition state was essentially known. However, if a unit of pavement segment was in a particular condition state and the tracking of the pavement ended because the condition state either increased or the study was terminated, the sojourn time of that unit of pavement segment was not precisely known and is therefore right-censored (Lee 1992).
6. Based on the complete and censored durations obtained from the data, the distribution of the sojourn time from condition state i to condition state j $[H_{i,j}(t)]$ was determined, where j is either $i-1$ or $i-2$, $i = 10, 9, 8, 7, 6, 5$ $\min(j) = 4$. The proportion of the length of pavement segments (miles) that left condition state i and went to condition state j to the total length of pavement segments (miles) that left condition state i and went to all condition states other than itself $(p_{i,j})$ was determined for each condition state i.

The distribution of the sojourn times was assumed to follow a Weibull distribution, as was the case in previous research in which semi-Markov deterioration models were developed (Black et al. 2005a, b; Sobanjo 2011; Sobanjo and Thompson 2011). The MLE of the scale (α) and shape (β) parameters of the Weibull distribution used to describe the sojourn time distributions was computed. Another *MATLAB* program was used to generate a series of transition probability matrices, based on the Weibull parameters α and β values) obtained for the sojourn times between condition states. The MLE of the α and β values serves as input for this program. The transition probabilities according to Eq. (15) are then determined and used to simulate the expected condition states and survival curves outlining the probability of remaining above particular threshold condition states over time.

J. Infrastruct. Syst. 2013.19:186-195.

Simulation Approach

Jiang and Sinha (1992) outlined a Monte Carlo simulation approach for predicting the condition of pavements based on transition probabilities from a Markov chain deterioration model. A modified approach was done in this study for both the semi-Markov and Markov chain deterioration models. The Monte Carlo method is one of the most commonly used simulation techniques in engineering modeling (Jiang and Sinha 1992). The *MATLAB* programs formulated were used to do the following simulations using transition probabilities derived from both the Markov chain and semi-Markov model:

- The average condition states over 20 years
- The survival curve of pavement segments remaining in the best condition also determined from the simulation results

A 20-year period simulation was performed using 10,000 trials. This number of trials was considered to be a significantly large number of trials for developing the model. The steps for one run of the simulation are outlined in the flow chart in Fig. 1. The first step of the simulation was to establish the yearly transition probability matrices, which were Π for the Markov chain model and Φ for the semi-Markov model. For the Markov chain case, the transition probability matrix was assumed to be independent of the interval, t, and thus remains the same for each interval. However, for the semi-Markov case, the transition probability matrix changes for each interval. Condition states i and j range from 10 (best condition) to 4 (worst condition). The next step was to create another matrix that had the cumulative probabilities of transition from condition state i to condition state j for each i, denoted

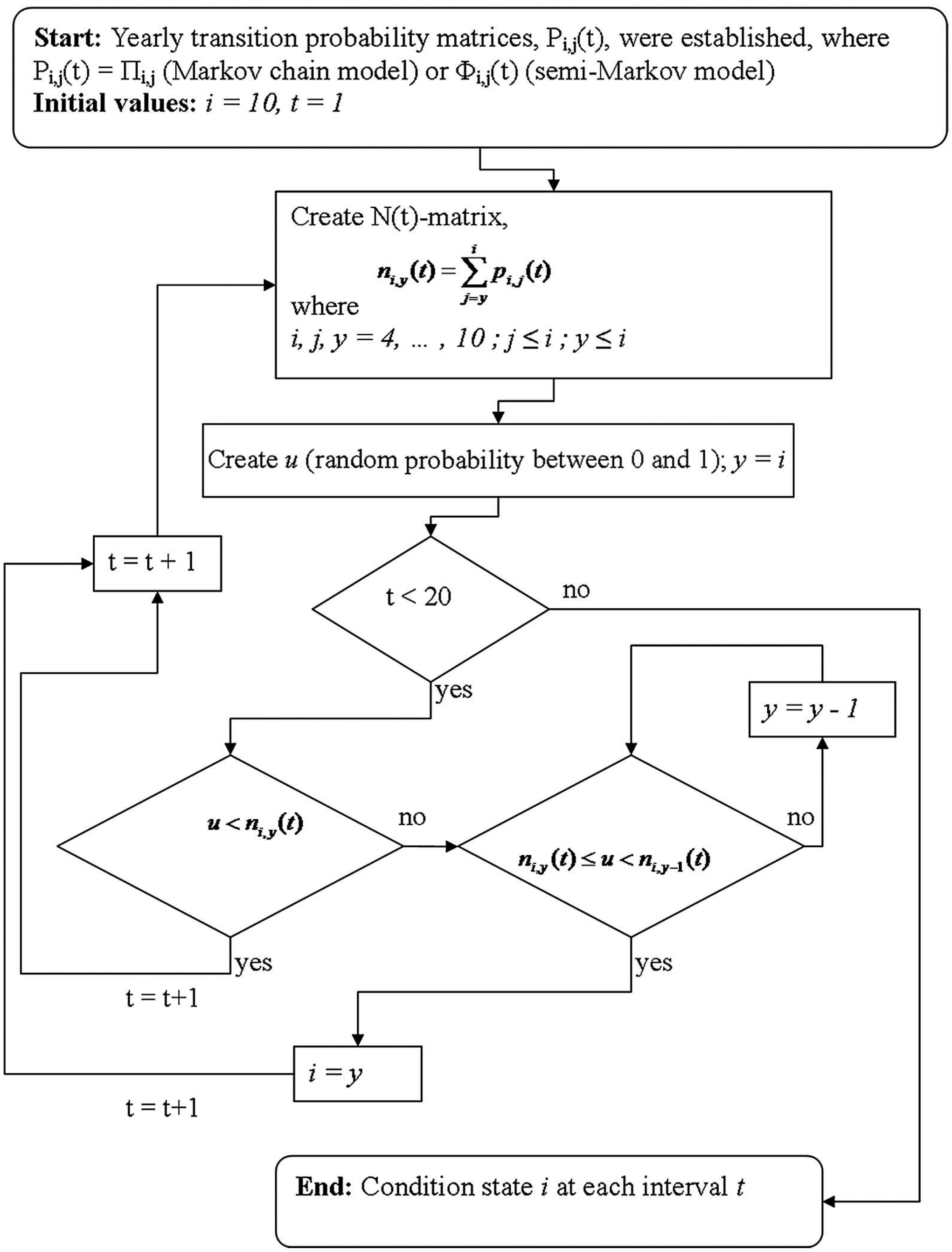

Fig. 1. Flowchart for a single run of simulation over 20 years

as the $N(t)$-matrix. The ith, jth term ($n_{i,j}$) in the $N_{i,j}$-matrix is the cumulative probability from condition state i of the respective $\pi_{i,j}$ terms in the $\Pi_{i,j}$ matrix for the Markov model. Likewise, $n_{i,j}$ in the $N_{i,j}(t)$-matrix is the cumulative probability from condition state i of the respective $\phi_{i,j}$ terms in the Φ matrix for the semi-Markov model. For the Markov chain case, $\Pi_{i,j}$ is shown later in Eq. (17) and the corresponding N-matrix is shown in Eq. (18).

The initial pavement condition state was set at $i = 10$. A random number, u, was then generated from a uniform distribution in the interval [0, 1]. If u was less than $n_{i,i}$, then the pavement remained in the same condition. If u fell between $n_{i,y}$ and $n_{i,y-1}$, where condition state $y = 5, \ldots, 10$, then the pavement transitioned to condition state $y - 1$.

Actual Pavement Deterioration Based on Test Data

To reflect the actual behavior of the pavement segments due to do-nothing actions, the test data were used to determine the average condition state over time and generate survival curves for the pavements in the network, which were estimated by computing the proportion of pavements that was greater than or equal to a particular condition state. The average condition state is synonymous with the weighted average pavement crack index over time (avgCRK). To determine the avgCRK for each year, the sum-product of CRK and length of pavement segment was computed for each year and divided by the total miles of the pavement segment for that year, outlined as follows:

$$\mathrm{avgCRK}_y = \frac{\sum_{i=1}^{n}(l_{i,y} * \mathrm{CRK}_{i,y})}{\sum_{i=1}^{n} l_{i,y}} \qquad i = \{1, 2, 3, 4, \ldots, n\}; \quad y = 1, 2, \ldots \tag{16}$$

where avgCRK_y = weighted average pavement crack index in year y; i = numerical indicator given to a unique pavement segment in a particular year for which the pavement condition due to a do-nothing action subsequent to becoming new is known; n = number of pavement segments in a particular year for which the pavement condition due to a do-nothing action subsequent to becoming new is known; y = corresponding number of years subsequent to segment's becoming new'; l_{iy} = length of a pavement segment i in year y; $\mathrm{CRK}_{i,y}$ = crack index of a pavement segment i in year y.

Discussion of Results

Results from both the Markov chain and semi-Markov models were compared against the actual test data on the CRKs. In the present study, 790 pavement segments from the training data were used to develop both the Markov chain and semi-Markov models. The terminal state, 4, was chosen due to the relatively small sample size of pavements with CRKs of 3.5 and less, particularly as it relates to the semi-Markov model. This limitation is explained subsequently in the paper. For comparative purposes, the same number of condition states used for the semi-Markov model was used for the Markov chain model.

Transition Probabilities Based on Markov Chain Model

The set of transition probabilities generated from the Markov chain model are represented as a transition probability matrix and is shown below in Eq. (17). To the left and top of the transition probability matrix in Eq. (17) are the condition states i and j, respectively, where Π_{ij} is the transition probability matrix used in the Markov chain model:

$$\Pi_{ij} = \begin{array}{c} \\ 10 \\ 9 \\ 8 \\ 7 \\ 6 \\ 5 \\ 4 \end{array} \begin{array}{c} \begin{array}{ccccccc} 10 & 9 & 8 & 7 & 6 & 5 & 4 \end{array} \\ \begin{pmatrix} 0.905 & 0.072 & 0.017 & 0.006 & 0 & 0 & 0 \\ & 0.737 & 0.157 & 0.090 & 0.016 & 0 & 0 \\ & & 0.660 & 0.274 & 0.042 & 0.014 & 0.010 \\ & & & 0.707 & 0.188 & 0.086 & 0.019 \\ & & & & 0.724 & 0.112 & 0.164 \\ & & & & & 0.582 & 0.418 \\ & & & & & & 1 \end{pmatrix} \end{array} \tag{17}$$

The corresponding N-matrix used in the simulation and described in Fig. 1 is shown in Eq. (18):

$$N_{ij} = \begin{array}{c} \\ 10 \\ 9 \\ 8 \\ 7 \\ 6 \\ 5 \\ 4 \end{array} \begin{array}{c} \begin{array}{ccccccc} 10 & 9 & 8 & 7 & 6 & 5 & 4 \end{array} \\ \begin{pmatrix} 0.905 & 0.977 & 0.994 & 1.0 & 1.0 & 1.0 & 1.0 \\ & 0.737 & 0.894 & 0.984 & 1.0 & 1.0 & 1.0 \\ & & 0.660 & 0.934 & 0.976 & 0.990 & 1.0 \\ & & & 0.707 & 0.895 & 0.981 & 1.0 \\ & & & & 0.724 & 0.836 & 1.0 \\ & & & & & 0.582 & 1.0 \\ & & & & & & 1.0 \end{pmatrix} \end{array} \tag{18}$$

Transition Probabilities Based on Semi-Markov Model

The transition probability of the embedded Markov chain of the semi-Markov process was estimated for each transition, and the results are shown in Table 1. The frequency distribution of the observed sojourn time for each unit mile before transition was determined. In addition, the sample sizes of the sojourn time for each unit of pavement segment in the higher condition states are generally more than that in the lower condition states. This means that for the semi-Markov deterioration model, there is likely to be an increase in the uncertainty in the predictions for the lower

J. Infrastruct. Syst. 2013.19:186-195.

Table 1. Transition Probabilities of Embedded Markov Chain of Semi-Markov Process ($p_{i,j}$) and Results of Maximum Likelihood Estimation

Transition			95% Confidence limit for $\hat{\alpha}$			95% Confidence limit for $\hat{\beta}$			
i to j	$p_{i,j}$	$\hat{\alpha}$	Lower	Upper	$\hat{\beta}$	Lower	Upper	Mean (years)	Standard deviation
10 to 9	0.707	9.432	9.332	9.533	2.128	2.094	2.163	8.35	4.13
9 to 8	0.752	4.887	4.777	4.999	1.579	1.539	1.620	4.39	2.84
8 to 7	0.645	3.496	3.394	3.602	1.345	1.304	1.387	3.21	2.41
7 to 6	0.468	5.039	4.811	5.278	1.257	1.208	1.308	4.69	3.75
6 to 5	0.214	6.304	5.754	6.906	1.523	1.412	1.641	5.68	3.80
5 to 4	1.000	3.164	3.030	3.304	2.062	1.940	2.193	2.80	1.43
10 to 8	0.293	13.126	12.904	13.351	3.182	3.088	3.278	11.75	4.05
9 to 7	0.248	6.103	5.845	6.372	1.249	1.204	1.295	5.69	4.58
8 to 6	0.355	9.672	8.744	10.697	1.465	1.350	1.591	8.76	6.08
7 to 5	0.532	9.103	8.355	9.918	1.236	1.165	1.312	8.50	6.91
6 to 4	0.786	5.417	5.092	5.763	1.693	1.591	1.802	4.84	2.94

Note: $p_{i,j}$ = transition probability from state i to j; $\hat{\alpha}$ = scale parameter estimate; $\hat{\beta}$ = shape parameter estimate.

condition states. A Weibull distribution was assumed for the sojourn times in a condition state before transitioning to the lower condition state. Table 1 summarizes these results.

The 95% confidence limits of the scale and shape parameters are also outlined in Table 1. Generally, the ranges of the confidence limits are not excessive, and the shape parameter, β, estimates are significantly greater than 1 because the value 1 does not fall within the confidence interval (CI). In Table 1 the means and standard deviations for each sojourn time (Weibull) distribution are also shown. The standard deviations for the sojourn time in Condition State 8 transitioning to 6, and that of Condition State 7 transitioning to 5 seems relatively high in comparison to the others. The high standard deviations are a function of the available data. It is believed that as more data for each transition become available, the standard deviations associated with those transitions will decrease.

The first five yearly transition probability matrices for the semi-Markov model are shown in Eqs. (19)–(23). For Eq. (19), $\phi_{10,10} = 0.991$ means that there is a 0.991 probability that the pavement segment will remain in Condition State 10. The other $\phi_{i,j}$ terms represent the probability of transition from i to j in a given year.

Year 1:

$$\Phi_{ij}(1) = \begin{array}{c} \\ 10 \\ 9 \\ 8 \\ 7 \\ 6 \\ 5 \\ 4 \end{array} \begin{array}{c} \begin{array}{ccccccc} 10 & 9 & 8 & 7 & 6 & 5 & 4 \end{array} \\ \begin{pmatrix} 0.991 & 0.008 & 0.001 & 0 & 0 & 0 & 0 \\ & 0.822 & 0.078 & 0.100 & 0 & 0 & 0 \\ & & 0.795 & 0.170 & 0.035 & 0 & 0 \\ & & & 0.814 & 0.123 & 0.063 & 0 \\ & & & & 0.886 & 0.059 & 0.056 \\ & & & & & 0.911 & 0.089 \\ & & & & & & 1 \end{pmatrix} \end{array} \tag{19}$$

Year 2:

$$\Phi_{ij}(2) = \begin{array}{c} \\ 10 \\ 9 \\ 8 \\ 7 \\ 6 \\ 5 \\ 4 \end{array} \begin{array}{c} \begin{array}{ccccccc} 10 & 9 & 8 & 7 & 6 & 5 & 4 \end{array} \\ \begin{pmatrix} 0.970 & 0.028 & 0.002 & 0 & 0 & 0 & 0 \\ & 0.716 & 0.150 & 0.134 & 0 & 0 & 0 \\ & & 0.690 & 0.249 & 0.061 & 0 & 0 \\ & & & 0.749 & 0.166 & 0.085 & 0 \\ & & & & 0.773 & 0.107 & 0.120 \\ & & & & & 0.744 & 0.256 \\ & & & & & & 1 \end{pmatrix} \end{array} \tag{20}$$

Year 3:

$$\Phi_{ij}(3) = \begin{array}{c} \\ 10 \\ 9 \\ 8 \\ 7 \\ 6 \\ 5 \\ 4 \end{array} \begin{array}{c} \begin{array}{ccccccc} 10 & 9 & 8 & 7 & 6 & 5 & 4 \end{array} \\ \begin{pmatrix} 0.944 & 0.049 & 0.007 & 0 & 0 & 0 & 0 \\ & 0.652 & 0.197 & 0.151 & 0 & 0 & 0 \\ & & 0.633 & 0.290 & 0.077 & 0 & 0 \\ & & & 0.717 & 0.188 & 0.095 & 0 \\ & & & & 0.695 & 0.138 & 0.167 \\ & & & & & 0.602 & 0.398 \\ & & & & & & 1 \end{pmatrix} \end{array} \tag{21}$$

Year 4:

$$\Phi_{ij}(4) = \begin{array}{c|ccccccc} & 10 & 9 & 8 & 7 & 6 & 5 & 4 \\ \hline 10 & 0.915 & 0.071 & 0.014 & 0 & 0 & 0 & 0 \\ 9 & & 0.603 & 0.234 & 0.163 & 0 & 0 & 0 \\ 8 & & & 0.591 & 0.319 & 0.090 & 0 & 0 \\ 7 & & & & 0.694 & 0.203 & 0.103 & 0 \\ 6 & & & & & 0.631 & 0.163 & 0.206 \\ 5 & & & & & & 0.484 & 0.516 \\ 4 & & & & & & & 1 \end{array} \qquad (22)$$

Year 5:

$$\Phi_{ij}(5) = \begin{array}{c|ccccccc} & 10 & 9 & 8 & 7 & 6 & 5 & 4 \\ \hline 10 & 0.883 & 0.093 & 0.024 & 0 & 0 & 0 & 0 \\ 9 & & 0.562 & 0.265 & 0.173 & 0 & 0 & 0 \\ 8 & & & 0.557 & 0.343 & 0.100 & 0 & 0 \\ 7 & & & & 0.676 & 0.215 & 0.109 & 0 \\ 6 & & & & & 0.577 & 0.183 & 0.240 \\ 5 & & & & & & 0.388 & 0.612 \\ 4 & & & & & & & 1 \end{array} \qquad (23)$$

Comparison of Markov Chain Model, Semi-Markov Model, and Actual Pavement Deterioration Behavior

Using the transition probabilities generated from both the Markov chain and semi-Markov models, Monte Carlo simulations were done for 10,000 trials. The overall expected (average) condition state over time was determined from the Markov chain and semi-Markov models. Both models were compared against the actual condition states over 20 years. The condition criteria outlined earlier in this paper were also used to determine the actual condition states based on the avgCRK values. The avgCRK for each year was computed by finding the sum product of the actual CRK values and length of pavement segment for each year and dividing by the total miles of the pavement segment for that year, as outlined previously in Eq. (16). Figs. 2 and 3 show a comparison of the actual condition states over time with the expected condition states from the Markov chain and the semi-Markov models, respectively. The actual and expected condition states of pavement segments can be treated as two groups of data, and the chi-squared significance test for comparing two proportions can be used to demonstrate statistically how close the prediction models are to the actual behavior of the flexible pavements and whether or not the models are good fits. The chi-squared test statistic is computed using the following formula (Montgomery and Runger 2006):

$$X_0^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} \qquad (24)$$

where O_i = observed value from prediction models; and E_i = expected value or actual value describing pavement behavior

To determine if the models are a good fit, consider the following hypotheses:

H_0: predicted values reflect actual pavement behavior

H_1: predicted values do not reflect actual pavement behavior

where $\alpha = 0.05$. The X_0^2 values of the results from the respective models in Figs. 2 and 3 are

- X_0^2 (semi-Markov model) = 4.9×10^3, $p = 0.00$
- X_0^2 (Markov chain model) = 2.8×10^3, $p = 0.00$

Therefore, both models are good fits based on the chi-squared test because the respective p-values are less than 0.05.

Figs. 2 and 3 also show the 95% CIs for the simulated values from the respective models. Fig. 2 indicates that the Markov chain model slightly underpredicts the actual pavement behavior at Years 2–6, as the actual condition states falls above the upper 95% CI bound. Although the semi-Markov model slightly overpredicts the actual pavement behavior at Years 7–11, it provides good predictions for the first 5 years because the actual condition states fall within the 95% CI bounds for the first 5 years (Fig. 3). Also,

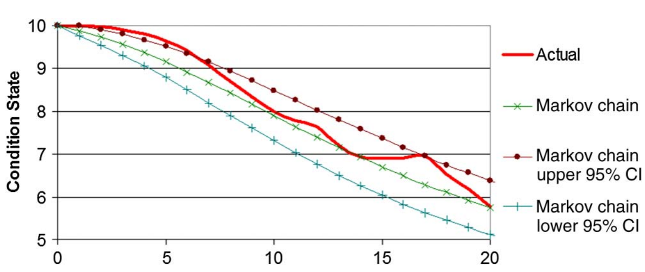

Fig. 2. Monte Carlo simulation results: 95% confidence interval bands on Markov chain model and actual condition

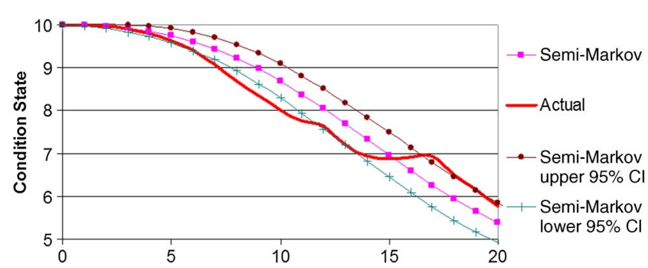

Fig. 3. Monte Carlo simulation results: 95% confidence interval bands on semi-Markov chain model and actual condition

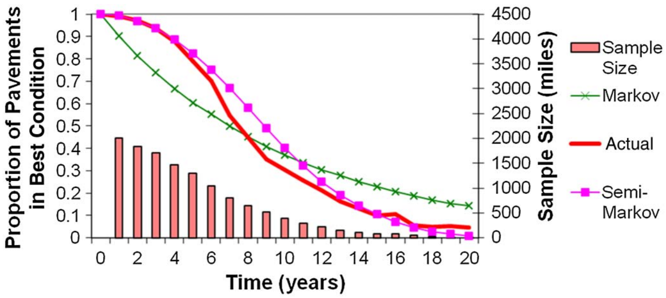

Fig. 4. Survivor curve: proportions of pavements in best condition state (State 10)

In Fig. 3, the 95% CI bands on the semi-Markov model are narrower when compared to those of the Markov chain model in Fig. 2, which suggests that there is less uncertainty based on the semi-Markov model. This is likely due to the fact that the semi-Markov model accounts for the change, with the age of the pavement, in sojourn times between states. The limitations of the semi-Markov model, particularly beyond Year 5, could be partially due to the sample size of the available data used to develop the model. However, the accuracy of the predictions in the earlier years is expected to be higher than that in the later years, where higher levels of uncertainty are typically expected. It should be noted that it is desirable to obtain good predictions for condition states as low as 6 because the decision threshold for pavements in Florida is typically at a crack index of 6.4.

In addition to estimating the expected condition states for the pavement segments, the survivor curve for pavements in the best condition (Condition State 10) was also produced and is shown in Fig. 4 (Wang et al. 1994). This curve shows the expected proportion of pavements above a predetermined threshold value, based on the actual pavement behavior of the test data and that due to the Markov chain and semi-Markov prediction models, respectively. The bar chart represents the sample size (in miles) from the test data that was used to generate the actual behavior curve (Fig. 4).

For the proportions of pavements in the best condition state (State 10), the test statistic, X_0^2, based on the results of the respective models in Fig. 4 are shown as follows:

- X_0^2 (semi-Markov model) = 2.8×10^3, $p = 0.00$
- X_0^2 (Markov chain model) = 8.6×10^3, $p = 0.00$

From the observations, the curve generated from the semi-Markov model closely matches the curve obtained from the test data. However, both models have a good fit because the respective p-values are less than 0.05.

Conclusion

This paper has outlined a feasible approach in which historical pavement conditions were used to develop a semi-Markov deterioration model that is applicable in modeling network-level performance. This methodology can be used on any group of pavement segments based on their engineering properties or external factors, such as environmental conditions and traffic loading. Once an adequate sample of pavement segments with particular characteristics were obtained, Monte Carlo simulations were used to generate deterioration curves based on the Markov chain and semi-Markov deterioration models. A comparison of results produced from the traditional Markov chain model, the approach based on the semi-Markov process, and the actual pavement condition behavior for the deterioration of flexible pavements was outlined.

Both the Markov chain and semi-Markov models were good fits for actual pavement deterioration behavior based on the chi-squared test. However, the Markov chain model slightly underpredicted the actual pavement behavior at Years 2–6, while the semi-Markov model slightly overpredicted the actual pavement behavior at years 7–11 at a significance level of $\alpha = 0.05$. The semi-Markov model appeared to be superior to the Markov chain model counterpart for predicting the onset of deterioration. It is expected that as the sample size of the data used to develop the model decreases over time, there is a higher degree of uncertainty in the predictions. Therefore, the fact that the semi-Markov model slightly overpredicted in Years 7–11 does not necessarily reflect the inadequacy of the model, but could have been a function of the size of the data for those years.

It can be concluded that if sufficient data are available, then the use of a semi-Markov process may be an improved alternative to the traditional Markov chain process in modeling the deterioration of flexible pavements. The deterioration model only forms a part of a PMS, and so future work will involve integrating the semi-Markov deterioration model with a cost-effectiveness model for maintenance, repair, and rehabilitation for various alternatives to obtain a PMS that can be used in the industry.

References

Baik, H.-S., Seok, H. J., and Abraham, D. M. (2006). "Estimating transition probabilities in Markov chain-based deterioration models for management of wastewater systems." *J. Water Resour. Plann. Manage.*, 132(1), 15–24.

Billington, R., and Allan, R. N. (1983). *Reliability evaluation of engineering systems: Concepts and techniques*, Pitman, London.

Birolini, A. (2007). *Reliability engineering: Theory and practice*, 5th Ed., Springer, Berlin.

Black, M., Brint, A. T., and Brailsford, J. R. (2005a). "A semi-Markov approach for modelling asset deterioration." *J. Oper. Res. Soc.*, 56, 1241–1249.

Black, M., Brint, A. T., and Brailsford, J. R. (2005b). "Comparing probabilistic methods for the asset management of distributed items." *J. Infrastruct. Syst.*, 11(2), 102–109.

Casella, G., and Berger, R. L. (2001). *Statistical inference*, 2nd Ed., Duxbury, Pacific Grove, CA.

Castillo, E., Hadi, A. S., Balakrishnan, N., and Sarabia, J. M. (2005). *Extreme value and related models with applications in engineering and science*, Wiley, Hoboken, NJ.

Cleves, M. A., Gould, W. W., and Gutierrez, R. G. (2001). *An introduction to survival analysis using stata, revised edition*, Stata Press, College Station, TX.

Dietrich, B. (2010). *New developments in pavement design and pavement management in Florida, Civil and Environmental Engineering Seminar Series*, College of Engineering, Florida A&M Univ.–Florida State Univ., Tallahassee, FL.

Golabi, K., Thompson, P. D., and Hyman, W. A. (1993). *Pontis Version 2.0 technical manual: A network optimization system for bridge improvements and maintenance*, *Rep. FHWA-SA-93-031*, Office of Technology, Applications, Federal Highway Administration, Washington, DC.

Howard, R. A. (1971). *Dynamic probabilistic systems. Volume II: Semi-Markov and decision processes*, Wiley, Toronto, ON, Canada.

Ibe, O. C. (2009). *Markov processes for stochastic modeling*, Elsevier, Burlington, MA.

Jiang, Y., and Sinha, K. C. (1992). "Comparison of methodologies to predict bridge deterioration." *Transportation Research Record 1597*, Transportation Research Board, Washington, DC, 34–42.

Lee, E. T. (1992). *Statistical methods for survival data analysis*, 2nd Ed., Wiley, Hoboken, NJ.

Lounis, Z., and Mirza, M. S. (2001). "Reliability-based service life prediction of deteriorating concrete structures." *Proc., 3rd Int. Conf. on Concrete Under Severe Conditions*.

J. Infrastruct. Syst. 2013.19:186-195.

MATLAB [Computer software]. MathWorks, Natick, MA.

Micevski, T., Kuczera, G., and Coombes, P. (2002). “Markov model for storm water pipe deterioration.” *J. Infrastruct. Syst.*, 8(2), 49–56.

Mishalani, R. G., and Madanat, S. M. (2002). “Computation of infrastructure transition probabilities using stochastic duration models.” *J. Infrastruct. Syst.*, 8(4), 113–120.

Montgomery, D. C., and Runger, G. C. (2006). *Applied statistics and probability for engineers*, 4th Ed., Wiley, Hoboken, NJ.

Nasseri, S., Gunaratne, M., Yang, J., and Nazef, A. (2009). “Application of improved crack prediction methodology in Florida’s highway network.” *Transportation Research Record 2093*, Transportation Research Board, Washington, DC, 67–75.

Ng, S.-K., and Moses, F. (1998). “Bridge deterioration modeling using semi-Markov theory.” *Proc., of ICOSSAR’97, The 7th Int. Conf. on Structure Safety and Reliability*, Vol. 1, Kyoto, 24–28 November 1997, 113–120.

Ross, S. M. (1996). *Stochastic processes*, 2nd Ed., Wiley, Hoboken, NJ.

Ross, S. M. (2002). *First course in probability*, 6th Ed., Prentice-Hall, Englewood Cliffs, NJ.

Sobanjo, J. O. (2011). “State transition probabilities in bridge deterioration based on weibull sojourn times.” *Struct. Infrastruct. Eng. Maint. Manage. Life-Cycle Des. Perform.*, 7(10), 747–764.

Sobanjo, J. O., and Thompson, P. D. (2001). *Enhancement of the FDOT’s project level and network level bridge management analysis tools*, Final Rep., *Contract BDK83 977-01*, Florida Dept. of Transportation, Tallahassee, FL.

Tobias, P. A., and Trindade, D. C. (1995). *Applied reliability*, 2nd Ed., Chapman and Hall/CRC Press, Boca Raton, FL.

Wang, C. P. (1992). “Pavement network optimization and analysis.” Ph.D. dissertation, Arizona State Univ., Tempe, AZ.

Wang, K. C. P., Zaniewski, J., and Way, G. (1994). “Probabilistic behavior of pavements.” *J. Transp. Eng.*, 120(3), 358–375.

Yang, J. (2004). “Road crack condition performance modeling using recurrent Markov chains and artificial neural networks.” Ph.D. dissertation, Univ. of South Florida, Tampa, FL.

Yang, J., Gunaratne, M., Lu, J. J., and Dietrich, B. (2005). “Use of recurrent Markov chains for modeling the crack performance of flexible pavements.” *J. Transp. Eng.*, 131(11), 861–872.

Yang, J., Lu, J. J., Gunaratne, M., and Dietrich, B. (2009). “Modeling crack deterioration of flexible pavements: Comparison of recurrent Markov chains and artificial neural networks.” *Transportation Research Record 1974*, Transportation Research Board, Washington, DC, 18–25.

J. Infrastruct. Syst. 2013.19:186-195.