Estimating Transition Probabilities in Markov Chain-Based Deterioration Models for Management of Wastewater Systems

Hyeon-Shik Baik[1]; Hyung Seok (David) Jeong[2]; and Dulcy M. Abraham[3]

Abstract: Accurate prediction of the current and future conditions of wastewater systems using available assessment data is crucial for developing appropriate proactive maintenance and rehabilitation strategies for an aging wastewater collection and conveyance system. This paper proposes a method to estimate the transition probabilities of different condition states in Markov chain-based deterioration models for wastewater systems using an ordered probit model. The proposed model is applied and evaluated using the condition data of sewer pipes managed by the City of San Diego's Metropolitan Wastewater Department. The developed model presents some advantages in estimating transition probabilities over the approaches developed in the past, including the nonlinear optimization-based approach, in terms of versatility in the implementation, precision of the estimated data, and appropriateness of the assumptions in the model. The paper concludes that the ordered probit model approach is a statistically sound and robust method; however, in order to gain greater accuracy in deterioration modeling, periodic assessment of the wastewater systems with more data types is desirable.

DOI: 10.1061/(ASCE)0733-9496(2006)132:1(15)

CE Database subject headings: Wastewater management; Deterioration; Markov chains; Probability; Maintenance; Rehabilitation.

Introduction

According to the American Society of Civil Engineers (ASCE 2003), the condition grade of 16,000 wastewater collection and conveyance systems in the United States, declined from C ("Mediocre") in 1988 to D ("Poor") in 2002. The report also indicated that, currently, there is a $12 billion annual shortfall in funding for the wastewater system needs, while federal funding has remained flat for a decade. As a result, managers of wastewater systems are facing severe challenges in the operation, maintenance, and rehabilitation of degrading systems under limited budgets and stringent environmental regulations.

For cost-effective management of wastewater systems, consensus opinion has increased in recommending that a proactive management system in which actions are planned and taken before any functional failure of the system occurs must be adopted. Traditional wastewater management systems tend to be reactive in general, relying on a "fix it if and when it fails" strategy (Fenner 2000). This reactive management actually incurs more costs ranging from two to ten times that of proactive strategies, not to mention the adverse publicity and health and safety problems (WEF/ASCE 1994; Hahn et al. 2002).

A proactive wastewater management system requires consideration of current and future performance data of the wastewater system components regarding the hydraulic, structural, and environmental conditions and also utilization of appropriate tools to prioritize proactive activities and optimize the budget allocation for these activities. Many condition assessment technologies for wastewater systems, such as closed-circuit television (CCTV), infrared thermography, ground penetrating radar, digital imaging technology, gyroscope technology, etc., are currently available. A comprehensive discussion of these technologies can be found in the literature (Wirahidikusumah et al. 1998; Makar 1999). Even though wastewater inspections clearly provide a better understanding of the current conditions of wastewater systems, the information obtained from wastewater inspections only provides a snapshot of the wastewater conditions at the time of the inspection. Prediction of the future condition of the assets or of the current condition of the assets for which only past condition data are available is crucial and necessary in order to develop efficient proactive management systems. With proper prediction models (i.e., deterioration models) of wastewater systems, operation and maintenance managers can prepare timely and cost-effective inspection, rehabilitation, and replacement schedules using optimization techniques.

Markov chain-based models have been extensively used for developing deterioration models for bridge and pavement systems (Madanat et al. 1995; Morcous et al. 2002) and have recently been applied to wastewater systems (Wirahadikusumah et al. 2001; Kleiner 2001; Micevski et al. 2002). One of the critical processes in the development of a Markov chain-based deterioration model is the estimation of transition probabilities,

[1]Assistant Director, Cost and Construction Management Division, Office of Subway Construction, Seoul Metropolitan Government, Seoul, Korea 100-813; formerly, Graduate Research Assistant, School of Civil Engineering, Purdue Univ., West Lafayette, IN 47907.

[2]Assistant Professor, School of Civil and Environmental Engineering, Oklahoma State Univ., Stillwater, OK 74078; formerly, Graduate Research Assistant, School of Civil Engineering, Purdue Univ., West Lafayette, IN 47907.

[3]Professor, School of Civil Engineering, Purdue Univ., West Lafayette, IN 47907.

Note. Discussion open until June 1, 2006. Separate discussions must be submitted for individual papers. To extend the closing date by one month, a written request must be filed with the ASCE Managing Editor. The manuscript for this paper was submitted for review and possible publication on October 28, 2004; approved on February 1, 2005. This paper is part of the ***Journal of Water Resources Planning and Management***, Vol. 132, No. 1, January 1, 2006. ©ASCE, ISSN 0733-9496/2006/1-15–24/$25.00.

which provide information about the probabilities of condition changes and can be used for prediction of the time of condition changes in a system.

In this study, an ordered probit model-based methodology is developed to estimate the transition probabilities in a Markov chain-based deterioration model for a wastewater system. The proposed model is applied and evaluated using the condition data of sewer pipes managed by the City of San Diego's Metropolitan Wastewater Department. The advantages of this model as compared to the use of the nonlinear optimization-based approach, which was applied in Wirahadikusumah et al. (2001) for estimating transition probabilities, are described. The limitations of the developed model are also discussed, along with recommendations and guidelines in order to improve the modeling processes and obtain more accurate results for the management of wastewater systems.

Deterioration Models for Wastewater Systems

During the last decade, various efforts have been made to develop a quantitative deterioration model for wastewater systems. A cohort survival model was developed in Norway by Rostum et al. (1999). Using the Herz distribution (Herz 1998), cohorts are defined as a set of elements installed in the same year with a particular failure probability. Condition assessment data using CCTV were used in this study, and five different conditional states based on a Norwegian standard procedure (NORVAR 1998) were employed for developing the deterioration model. In this model, sewer systems were grouped based on the construction period and other features such as pipe material, pipe dimensions, and soil conditions. Because the state survival functions allow the asset managers to forecast the residual lifetime of the pipe and the transition year between different states (e.g., from acceptable states to a critical state), proactive inspection strategies can be developed for the sewer network and the value of the sewer system can be estimated.

Using a logistic regression model, Ariaratnam et al. (2001) developed a model to predict the likelihood that a particular sewer segment was in a deficient state. In this study, sewer network data in Edmonton, Canada, were utilized to develop the model. Three explanatory variables (age, diameter, and waste type) were identified to be statistically significant factors in sewer deterioration. Both the material type and the depth of cover were determined to be statistically insignificant. Because the outcome variable in the logistic regression model is binary and assumes a Bernoulli distribution, the model is only useful to predict whether or not the pipe is in a deficient state, which includes two worst conditions among five different condition ratings. Thus, the model is not applicable to the prediction of a condition rating of a sewer segment.

Wirahadikusumah et al. (2001) developed a Markov chain-based deterioration model for large combined sewers in Indianapolis, Indiana. An exponential model was employed in the regression analysis for establishing the relationship between the overall structural grade and sewer age. It was assumed that the condition of a sewer does not drop by more than one state in a 1 year transition. In order to predict the transition probability among the five different structural grades, the nonlinear optimization-based approach used for pavements and bridges (Jiang et al. 1988; Jiang and Sinha 1989) was adopted. This study developed different deterioration models for different combinations of factors such as pipe material, groundwater level, backfill material, and depth of cover. The writers conceded that the lack of at least three consecutive data sets in different time periods prohibited verification of the key assumption in the Markov-chain model, which states that the conditional probability of any future event is independent of the past event and depends only upon the present state. In addition, the criticism regarding the transition probability values based on regression analysis (Madanat et al. 1995) was not overcome in the developed model.

Another Markov chain-based deterioration model for water and wastewater systems, such as water transmission pipes and trunk sewers, was proposed by Kleiner (2001). A single-state transition among different condition states was assumed, and the transition time was modeled as a random variable with a Weibull probability distribution. A Monte Carlo simulation was performed to generate data for the calculation of the duration in each state. However, the model developed in this study remains a theoretical framework only, because no actual data was employed to validate the model. In addition, due to the lack of actual data, age was the only factor considered to be affecting the deterioration of the system. The impact of the various factors, such as pipe material, pipe diameter, soil type, etc., on the deterioration process was not analyzed.

Micevski et al. (2002) also presented a Markov chain-based deterioration model for storm water pipes in Australia. Multiple state transitions among four different states during a 1 year transition period were assumed and the transition probabilities were estimated using a Metropolis-Hastings algorithm. Using a split sample analysis, the developed Markov model was proved to be statistically consistent with the data at the 5% significance level. Furthermore, the study reported that pipes in four categories, including pipe diameter, pipe construction material, soil type, and adjacency to the coastline, have different deterioration rates, so different Markov models for these categories are required.

Markov Chain Processes

The Markov chain is a discrete-time stochastic process (Winston 1994), where the conditional probability of any future event depends only on the present state and is independent of the past states (Ross 2000). The Markovian property can be expressed as follows for all states $i_0, i_1, \ldots, i_{t-1}, i_t, i_{t+1}$ and all $t \geqslant 0$:

$$\begin{aligned} P(X_{t+1} = i_{t+1} | X_t = i_t, X_{t-1} = i_{t-1}, \ldots, X_1 = i_1, X_0 = i_0) \\ = P(X_{t+1} = i_{t+1} | X_t = i_t) \end{aligned} \tag{1}$$

Because the Markov chain assumes that the conditional probability does not change over time, for all States i and j and all $t, P(X_{t+1}=j|X_t=i)$ is independent of t, as expressed in Eq. (2):

$$P(X_{t+1} = j | X_t = i) = p_{ij} \tag{2}$$

where p_{ij}=transition probability that, given the system is in State i at time t, it will be in a State j at time $(t+1)$.

The transition probabilities are commonly expressed as an $m \times m$ matrix called the transition probability matrix (or transition matrix), $\mathbf{P}$. The transition probability matrix $\mathbf{P}$ and its characteristics are given as follows:

$$\mathbf{P} = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1m} \\ p_{21} & p_{22} & \cdots & p_{2m} \\ \vdots & \vdots & \vdots & \vdots \\ p_{m1} & p_{m2} & \cdots & p_{mm} \end{bmatrix} \tag{3}$$

Table 1. Samples of Standard Defect Codes and Point Values for San Diego MWWD

Code and severity	Observation	Description	Maintenance points	Structural points
D-S	Deformation, small	Less than 15% of inside diameter	0	50
D-M	Deformation, medium	Between 15 and 30% of inside diameter	0	100
D-L	Deformation, large	>30%	0	150
X-N	Collapsed pipe	Use if a section of pipe wall has fallen in and structural integrity of pipe has been compromised	0	700
DEG-S	Debris–grease, small	Slight indication 1/4 to 1/2 in. thick	50	0
DEG-M	Debris: grease, medium	1/2 to 2 in. thick	75	0
DEG-L	Debris–grease, large	Greater than 2 in. thick	150	0
CRA-S	Roots around lateral, small	Small roots from around outside of lateral	20	50
CRA-M	Roots around lateral, medium	Medium roots from around outside of lateral	50	50
CRA-L	Roots around lateral, large	Heavy roots from around outside of lateral	75	100

$$\sum_{j=1}^{m} p_{ij} = 1 \quad \text{for } i = 1,2,\ldots,m \tag{4}$$

Based on the Chapman-Kolmogorov equation, the probability of the system moving from State i to State j after n periods (n transitions), that is, the n-step transition probability matrix, $\mathbf{P}^{(n)}$, can be obtained by multiplying the matrix $\mathbf{P}$ by itself n times (Ross 2000). Thus:

$$\mathbf{P}^{(n)} = \mathbf{P}^{n} \tag{5}$$

Let the initial state vector, $\mathbf{Q}^{(0)}$, be the probability that the Markov chain is in State i at time 0. Then, the state vector, $\mathbf{Q}^{(n)}$, which is the probability that the chain is in State j after n transitions, can be expressed (Winston 1994) as

$$\mathbf{Q}^{(n)} = \mathbf{Q}^{(0)}\mathbf{P}^{(n)} \tag{6}$$

where $\mathbf{Q}^{(0)}=[q_1,q_2,\ldots,q_m]$; and q_i=probability of being in State i at Time 0.

In the application of the Markov chain processes to the development of a deterioration model for wastewater systems, once a reliable transition probability matrix for the system is identified, the expected condition of the system in the future or the expected years that the system will be in a degraded condition can be easily obtained. These prediction values can be used for cost-effective operation, maintenance, and rehabilitation decision making. The next section discusses the development of an order probit model approach to estimate the transition probabilities for sewer pipes in the City of San Diego. The nonlinear optimization technique-based approach used by Wirahadikusumah et al. (2001) is also applied to the data set.

Estimation of Transition Probabilities

The estimation of transition probabilities in a Markov chain-based deterioration model requires data from the condition assessments of existing systems. The first step in a condition assessment of wastewater systems is determination of the current status of the structural and hydraulic condition of the systems. An assessment of the structural conditions of sewer pipes establishes the severity of the defects. The adequacy of the capacity of the existing wastewater systems is evaluated through assessments of hydraulic conditions. Structural conditions are investigated through internal inspections, whereas hydraulic conditions are analyzed through hydraulic modeling. Infiltration and inflow are also investigated to identify the causes for structural failures and hydraulic surcharges. Using the results of a structural condition assessment, transition probabilities for a Markov chain-based deterioration model can be estimated. In this study, condition assessment data sets obtained from the City of San Diego were used.

Data Source

The City of San Diego's Metropolitan Wastewater Department (MWWD) manages approximately 3,000 mi (4,800 km) of sanitary sewer lines. As a part of a 10 year capital program of pipe replacement and rehabilitation, MWWD is conducting an inspection program to evaluate the conditions of the sewer pipes. The data used in this paper was obtained from the inspections of phase 1A, in which approximately 90 km (55 mi) of sewer pipes were inspected during the latter half of 2001 (San Diego MWWD 2002). Most of the sewer pipes are vitrified clay (VC) and polyvinyl chloride (PVC) pipes with sizes ranging from 150 to 250 mm (6 to 10 in.) in diameter. About 90% of the pipes were installed before 1965, and 94% of the pipes were 100 mm (8 in.) or smaller in diameter.

The condition rating system used by MWWD for the inspection of these sewers consists of 108 criteria, which can be categorized into seven subgroups. Each subgroup contains rating criteria describing the characteristics as well as the severity of the defects. For each criterion, maintenance and/or structural points are assigned to evaluate the conditions of the sewer pipes. Samples of the criteria and assigned points are presented in Table 1.

The condition ratings are computed using the assigned maintenance and structural points from the inspection. The equation used for the rating is

$$\text{Score} = \frac{\sum SP \times SW + \sum \times MP \times MW}{LS} \tag{7}$$

where SP=structural points; SW=structural weight; MP=maintenance points; MW=maintenance weight; and LS=length of segment (ft).

The structural weight and maintenance weight used for the condition rating are 1 and 0, respectively. Once the score for each sewer segment is calculated, the condition of the pipe is represented using a letter grade from A to E. For instance, scores in the range of 0 to 2.5 are classified as Grade A, while scores that are greater than 6.0 are classified as Grade E. In this grading system, Grade A indicates that the pipe is in the best condition,

Table 2. Score Ranges for Grades (Condition Rating)

Condition rating	Grade	Score range
1	A	0–2.5
2	B	0–2.5
3	C	2.5–4.0
4	D	4.0–6.0
5	E	Above 6.0

whereas Grade E denotes the worst condition. The ranges used to categorize the conditions of the pipes are given in Table 2.

In this grading system, Grade A and Grade B pipes have the same range. The difference between the two grades is the existence of major defects that can cause relatively severe damage to the pipes, e.g., a broken pipe, a hole in the pipe, deformation, a broken joint, etc. If there is at least one major defect in the sewer segment, the segment is rated as Grade B even though the score is less than 2.5. In this study, these condition ratings are converted to a number from 1 (Grade A) to 5 (Grade E).

Ordered Probit Model Approach

Extensive efforts have been made in the development of deterioration models using econometric methods in pavements, bridge decks, and bridge expansion joints, e.g., Ramaswamy and Ben-Akiva (1990), Ben-Akiva et al. (1991), Ben-Akiva and Ramaswamy (1993), Ben-Akiva and Gopinath (1995), Madanat and Wan Ibrahim (1995), Madanat et al. (1995), Madanat et al. (1997), Bulusu and Sinha (1997), Prozzi and Madanat (2000), Mauch and Madanat (2001), Mishalani and Madanat (2002), and Lee and Chang (2003). These models include the use of the simultaneous equation model, the latent variable model, the Poisson regression model, the negative binomial model, the random-effects binary probit model, the duration model, and the ordered probit model.

In this paper, the concepts of an ordered probit model, along with an incremental model, are applied for the estimation of transition probabilities for wastewater systems, as this model was successfully used for bridge deterioration models in order to account for the drawbacks of the nonlinear optimization-based approach (Madanat et al. 1995). The concept of the incremental model was introduced by Madanat et al. (1995), in which the increments, i.e., the changes in condition ratings, during a transition period are calculated and used as the discrete outcomes in the ordered probit model. The probability estimated for a specific discrete outcome (increment) can be interpreted as the transition probability of the Markov chain model.

If the condition of a sewer segment changes from Condition State i to j during a transition period, the increment for this transition is $(j-i)$. By estimating probabilities for increments for every condition state, the values of each row of the transition probability matrix can be obtained. Because the transition matrix is estimated for each transition, the transition matrix based on this approach is nonstationary, or time dependent.

In the ordered probit model, the unobserved (latent) variable, z_{ik}, is used as the basis for the ranking of discrete data. In this study, the actual deterioration of a wastewater system is the latent variable, assumed to be continuous and varying between 0 and $+\infty$. Let k and i denote a specific sewer segment and its condition state, respectively. Then, the latent deterioration variable, z_{ik}, can be specified as a linear function as shown in Eq. (8) (Washington et al. 2003):

$$z_{ik} = \boldsymbol{\beta}_i \mathbf{X}_k + \varepsilon_{ik} \tag{8}$$

where $\boldsymbol{\beta}_i$=vector of estimable parameters for Condition State i; $\mathbf{X}_k$=vector of variables determining the discrete ordering for segment k; and ε_{ik}=random disturbance term. Using measurement equations that map the continuous latent variable (deterioration), z_{ik}, to a discrete indicator variable (condition increments), y_{ik}, the relationship between the latent variable and the indicator variable can be defined (Washington et al. 2003) as

$$y_{ik} = j - i;\ \text{if}\ \mu_{i(j-i)} \leqslant z_{ik} \leqslant \mu_{i(j-i+1)};\ \text{for}\ (j-i) = 0,\ldots,I-1 \tag{9}$$

where $(j-i)$=change in condition state of segment k after one transition; μ=thresholds, $\mu_{i0}=0$ and $\mu_{i(j-i+1)}=\infty$; and I=highest number for condition rating.

Eq. (9) indicates that, if the latent deterioration, z_{ik}, falls between the two thresholds, the change of condition rating (increment) becomes y_{ik}. By substituting Eq. (8) for Eq. (10), the ordered probit model can be expressed as

$$y_{ik} = j - i;\ \text{if}\ \mu_{i(j-i)} - \boldsymbol{\beta}_i\mathbf{X}_k \leqslant \varepsilon_{ik} \leqslant \mu_{i(j-i+1)} - \boldsymbol{\beta}_i\mathbf{X}_k;\quad \text{for}\ (j-i) = 0,\ldots,I-1 \tag{10}$$

For an ordered probit model, the disturbance term, ε_{ik}, is assumed to be normally distributed with mean=0 and variance=1. Therefore, the probability that the condition changes, y_{ik}, are equal to $(j-i)$ can be expressed using cumulative normal distribution, $\Phi(.)$, as shown in Eq. (11). This probability is the transition probability from condition i to j:

$$P(y_{ik} = j - i) = \Phi(\mu_{i(j-i+1)} - \boldsymbol{\beta}_i\mathbf{X}_k) - \Phi(\mu_{i(j-i)} - \boldsymbol{\beta}_i\mathbf{X}_k);\quad \text{for}\ (j-i) = 0,\ldots,I-1 \tag{11}$$

The likelihood function for the maximum likelihood estimation (MLE) over the population of K_i, which is the total number of sewer segments that are in State i, can be expressed as

$$L(y|\boldsymbol{\beta},\mu) = \prod_{k=1}^{K_i} \prod_{(j-i)=0}^{I-1} [\Phi(\mu_{i(j-i+1)} - \boldsymbol{\beta}_i\mathbf{X}_k) - \Phi(\mu_{i(j-i)} - \boldsymbol{\beta}_i\mathbf{X}_k)]^{\delta_{ik}} \tag{12}$$

where $\delta_{ik}=1$ if the observed increment of condition rating for segment k is $(j-i)$, and 0 otherwise.

The log-likelihood function for the ordered probit model is

$$LL = \sum_{k=1}^{K_i} \sum_{(j-i)=0}^{I-1} \delta_{ik} \log[\Phi(\mu_{i(j-i+1)} - \boldsymbol{\beta}_i\mathbf{X}_k) - \Phi(\mu_{i(j-i)} - \boldsymbol{\beta}_i\mathbf{X}_k)] \tag{13}$$

By maximizing the log-likelihood function given in Eq. (13), the model parameters, β, and thresholds, μ, can be jointly estimated.

The first step in the development of transition matrices for the Markov chain is the estimation of probabilities for increments in condition changes for each condition state. For instance, for condition ratings ranging from 1 (best) to 5 (worst) in this study, the possible number of increments for Condition State 2 is four (2→2, 2→3, 2→4, and 2→5), assuming no preservation or improvement activities are performed to upgrade the condition of the pipe. Because the total number of condition states is five, four incremental deterioration models are required, as the last row in the transition matrix is regarded as an absorbing state [i.e., the probability that the Condition State 5 will be 5 in Time t is always 1 (100%)].

Table 3. Variables Used for Ordered Probit Modeling

Name of variable	Description of variable
Length	Length of pipe segments between manholes in feet
Size	Diameter of pipe segments in inches
Type of material	Vitrified clay or PVC
Age	Age at year 2001 from installation year
Slope	Slope of pipe segments between manholes
	Slope=(elevation of upstream invert—elevation of downstream invert)/length

Based on the parameters for the ordered probit model estimated from the maximum log-likelihood function given in Eq. (13), the transition probabilities for each segment of wastewater collection systems can be computed as

$$\hat{P}(y_{ik}=0|\mathbf{X}_k,i)=\Phi(\hat{\mu}_{i1}-\hat{\boldsymbol{\beta}}_i\mathbf{X}_k)$$

$$\hat{P}(y_{ik}=1|\mathbf{X}_k,i)=\Phi(\hat{\mu}_{i2}-\hat{\boldsymbol{\beta}}_i\mathbf{X}_k)-\Phi(\hat{\mu}_{i1}-\hat{\boldsymbol{\beta}}_i\mathbf{X}_k)$$

$$\hat{P}(y_{ik}=2|\mathbf{X}_k,i)=\Phi(\hat{\mu}_{i3}-\hat{\boldsymbol{\beta}}_i\mathbf{X}_k)-\Phi(\hat{\mu}_{i2}-\hat{\boldsymbol{\beta}}_i\mathbf{X}_k) \tag{14}$$

$$\cdots$$

$$\hat{P}(y_{ik}=I-1|\mathbf{X}_k,i)=1-\Phi(\hat{\mu}_{i(I-1)}-\hat{\boldsymbol{\beta}}_i\mathbf{X}_k)$$

where $\hat{P}(y_{ik}|\mathbf{X}_k,i)$=transition probability from Condition State i to j for a segment with attribute vector $\mathbf{X}_k$.

The overall fit of the ordered probit model can be measured by the ρ^2 statistic (a measure of overall statistical fit, similar to R^2 for simple regression models). This statistic is sometimes referred to as the likelihood ratio index and is a measure of the likelihood improvement attributable to estimated coefficients [see Washington et al. (2003) for additional detail and examples], as shown in Eq. (15). The closer the ρ^2 statistic is to 1, the better the estimated model fits the observation datas:

$$\rho^2 = 1-\frac{LL(\boldsymbol{\beta})}{LL(0)} \tag{15}$$

where $LL(\boldsymbol{\beta})$=log-likelihood at convergence with parameter vector $\boldsymbol{\beta}$; and $LL(0)$=initial log-likelihood with all parameters set to 0.

Application of Ordered Probit Model Approach for San Diego Data Set

Using the given data set with 545 data points obtained from San Diego MWWD, Statistical Software Tools (SST) (Dubin and Rivers 1987) was used to find the maximum likelihood estimator of the model parameters, $\boldsymbol{\beta}$, and thresholds, μ, given in Eq. (11). One year was used as a transition period. A total of five variables were used in the modeling process, as shown in Table 3. However, depending on the availability of data, other variables, such as depth of installation, source of sewer (industrial and residential), soils surrounding pipes, groundwater level, traffic volume above pipe segments, and frequencies of overflow, can be included in the analysis. This information, however, was not available for this study.

The results of the estimates of the parameters and the thresholds for the ordered probit model for Condition States 1, 2, 3, and 4 are presented in Table 4. The estimation results show that the type of material (vitrified clay or polyvinyl chloride) is not a significant variable for all condition states in the analysis. This result is consistent with the findings by Ariaratnam et al. (2001), in which four different materials (concrete, clay tile, reinforced concrete, and polyvinyl chloride) of sewer pipes in Edmonton, Canada, were evaluated. However, it contradicts the findings by Micevski et al. (2002), in which two different materials (concrete and vitrified clay) for storm water pipes in Australia were statistically significant factors for different deterioration models. [Note that the study conducted by Miceviski et al. (2002) is based on data of stormwater pipes in Australia that do not transport sewage. Direct comparison with the results obtained from this study, which used data of sewer pipes in San Diego, may need care.]

Table 4. Estimation Results for Ordered Probit Model

	Condition State 1[a]		Condition State 2[b]		Condition State 3[c]		Condition State 4[d]	
Name of variable	Parameter estimate	t-statistic[e]	Parameter estimate	t-statistic	Parameter estimate	t-statistic	Parameter estimate	t-statistic
Constant	−2.39	−5.01	−1.64	−2.86	−0.72	−1.30	-	-
Length	-0.61×10^{-3}	−1.31	-2.11×10^{-3}	−3.74	-2.06×10^{-3}	−2.98	-2.11×10^{-3}	−2.49
Size	0.07	1.60	0.09	1.81	-	-	0.05	1.80
Age	0.07	14.04	0.04	5.57	0.03	2.73	-	-
Slope	1.72	2.51	2.56	3.13	1.73	1.88	3.93	2.59
Threshold 1	0.99	21.27	0.70	15.74	0.77	10.56	N/A	N/A
Threshold 2	1.61	39.52	1.29	22.49	N/A	N/A	N/A	N/A
Threshold 3	2.16	40.12	N/A	N/A	N/A	N/A	N/A	N/A

[a]Number of observations=545; $LL(0)=-1065.87$; $LL(\beta)=-767.24$; $\rho^2=0.280$.

[b]Number of observations=442; $LL(0)=-698.45$; $LL(\beta)=-576.82$; $\rho^2=0.174$.

[c]Number of observations=305; $LL(0)=-354.55$; $LL(\beta)=-323.35$; $\rho^2=0.088$.

[d]Number of observations=194; $LL(0)=-134.47$; $LL(\beta)=-124.52$; $\rho^2=0.074$.

[e]The use of t-statistics is a reliable approximation to evaluate the statistical significance in practice (Washington et al. 2003). Alternatively, the likelihood ratio test or Akaike information criterion can be used to identify the best model.

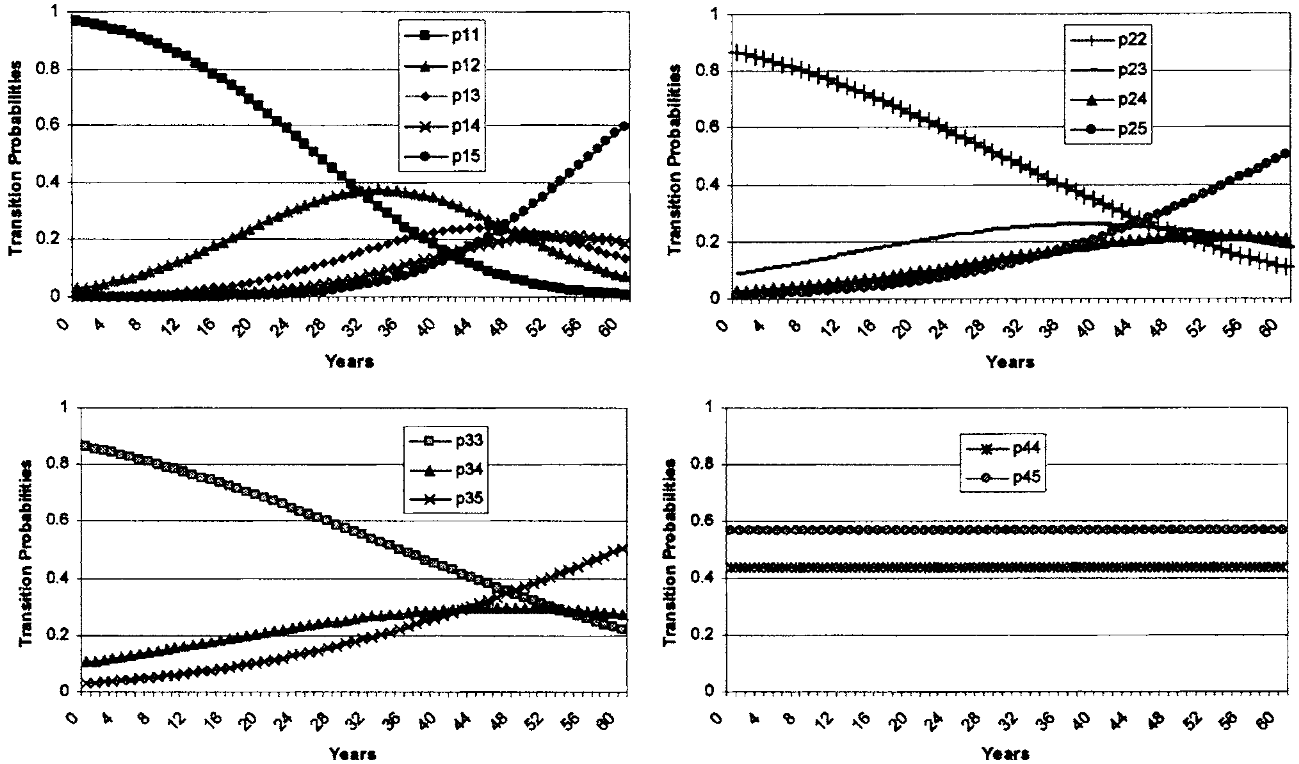

Fig. 1. Transition probabilities for 200 mm (8 in.) sewer pipes (0–60 years) (ordered probit model approach)

Length and Size

The estimation results indicate that longer sewer runs are less likely to deteriorate than shorter ones. This can be attributed to the fact that longer runs have fewer bends in which less debris is accumulated, creating fewer blockages or damage to the pipe from standing sewage. For Condition States 1, 2, and 4, larger pipes are more likely to have higher rates of deterioration, because they have more surface area exposed to sewage and surrounding soils, possibly causing more damage. The findings of this study indicate that the size of the pipe segments is not significant for the transitions of pipes in Condition State 3. Conversely, Ariaratnam et al. (2001) and Micevski et al. (2002) reported that, based on the statistical analysis of their specific data sets, the greater the diameter, the lower the likelihood that the pipes would deteriorate.

Age and Slope

For Condition States 1, 2, and 3, older pipes are more likely to deteriorate at a faster rate, which is consistent with the general perception of the deterioration of facilities; i.e., the deterioration rate is lower during the early years of useful life and higher during the later years. But, for Condition State 4, age is not a significant variable for the transitions of pipes, which implies that the deterioration of the pipes in Condition State 4 is significantly affected by the length, size, and slope of the pipe segments, rather than by age. This causes the transition probability for a pipe in Condition State 4 to remain stationary throughout the entire useful life. For all condition states, the steeper the slope, the higher the probability that pipe segments deteriorate. Low stability due to slopes and faster flow rates can be attributed to higher deterioration rates.

The transition probabilities (p_{ij}), i.e., the probabilities for the changes in condition rating (increments), for each condition state for each pipe segment can be estimated using the obtained $\boldsymbol{\beta}$ and μ incorporated with standard normal distribution, as shown in Eq. (14). Fig. 1 shows the transition probabilities of the 200 mm (8 in.) sewer pipe, which is 70 m (210 ft) long with a slope of 0.05 when the ordered probit model is used.

Nonlinear Optimization-Based Approach

The nonlinear optimization-based approach estimates transition probabilities by minimizing the absolute distance between the condition data points (or average condition ratings from a regression curve) and the expected value obtained from the Markov chain model. Details of the procedures for estimation of transition probabilities for wastewater systems based on the nonlinear optimization-based approach can be found in earlier studies (Butt et al. 1987; Jiang and Sinha 1989; Wirahadikusumah et al. 2001).

Because this approach allows only the age (t) of the pipe as an explanatory variable to predict the condition rating, pipes with different characteristics that are expected to have a different deterioration process must be categorized before the analysis. Two hundred millimeter (8 in.) VC pipes (316 data points) were used for this study's analysis. To establish the relationship between the condition ratings and the ages of the sewer pipes, regression analyses must be performed for the data set. An exponential model provided a good fit for the data set in this analysis. The average condition rating at age t (number of years after installation), $Y(t)$, for 200 mm (8 in.) VC pipes can be expressed as follows from the regression analysis:

$$Y(t) = e^{(-0.949+0.044t)} \tag{16}$$

The regression model given in Eq. (16) forms the basis for the estimation of transition probabilities when using the nonlinear optimization-based approach. The regression function is shifted to cross condition rating 1 at age 0. The transitions of the

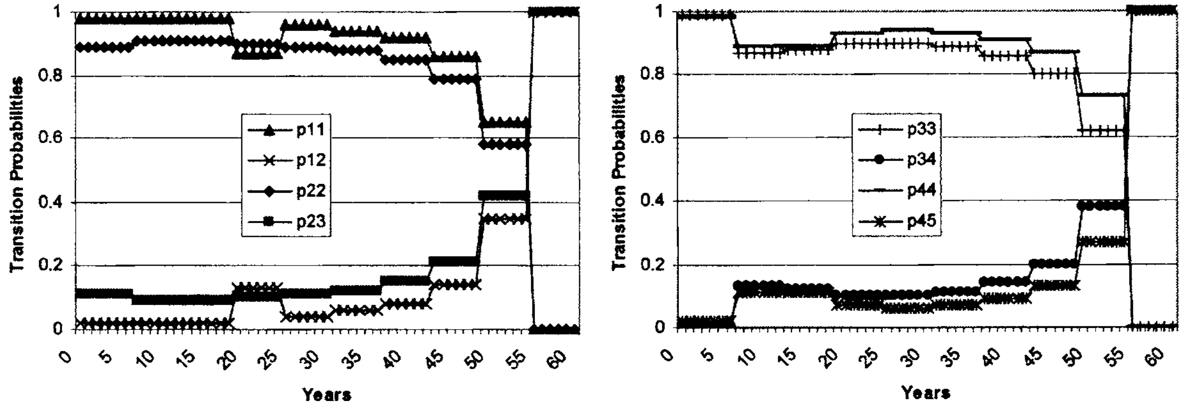

Fig. 2. Transition probabilities for 200 mm (8 in.) VC pipes (0–60 years) (nonlinear optimization based approach)

Markov chain model from State i to State j is represented by a 5×5 transition probability matrix, because the conditions of wastewater systems in the City of San Diego are divided into five different states. One year was used as a transition period. To meet the homogeneity assumption of the Markov chain model, a "zoning" concept was used wherein a 6 year term was used for a zone. Therefore, it was assumed that the values of the transition probabilities would not change over 6 years. To determine the entire deterioration pattern, it was assumed that no improvement activities were performed over the life of the systems. Hence, the transition probabilities have null values where i is greater than j. It was also assumed that the condition levels of the wastewater system do not drop more than one level in a transition (1 year). Thus, the transition probabilities where j is greater than $(i+1)$ will be 0. The transition probability matrix $\mathbf{P}$ can be expressed as shown in Eq. (17). The sum of probabilities in each row must be 1:

$$\mathbf{P}=\begin{bmatrix} p_{11} & p_{12} & 0 & 0 & 0 \\ 0 & p_{22} & p_{23} & 0 & 0 \\ 0 & 0 & p_{33} & p_{34} & 0 \\ 0 & 0 & 0 & p_{44} & p_{45} \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \tag{17}$$

By minimizing the objective values in Eq. (18) for each zone, the transition probabilities (p_{ij}) for 200 mm (8 in.) VC pipes can be estimated:

$$\text{Minimize}\sum_{t=t_s}^{t_e}\sum_{n=1}^{N}|Y(t)-E(n,\mathbf{P})|$$

$$\text{Subject to } 0\leqslant p_{ij}\leqslant 1; i,j=1,2,\ldots,5 \tag{18}$$

$$\sum_{j=1}^{5}p_{ij}=1 \qquad \text{for } i=1,2,\ldots,5$$

where t=age of the sewer pipe; t_s=starting age for each zone; t_e=ending age for each zone; n=number of transition periods (stages); N=total number of transition periods in each zone (N=6 years in this study); $Y(t)$=average condition rating at Age t, which is estimated from Eq. (16); and $E(n,\mathbf{P})$=expected value of the condition ratings of a sewer pipe segment for n transitions estimated based on the Markov chain model. The expected condition rating, $E(n,\mathbf{P})$ can be calculated by multiplying the state vector of Stage n as shown in Eq. (6) and the condition rating vector, $\mathbf{S}=[1,2,3,4,5]$, as shown in Eq. (19):

$$E(n,\mathbf{P})=\mathbf{Q}^{(n)}\mathbf{S}^T=\mathbf{Q}^{(0)}\mathbf{P}^{(n)}\mathbf{S}^T \tag{19}$$

where $\mathbf{Q}^{(n)}$=condition vector at Stage n; $\mathbf{Q}^{(0)}$=initial condition vector at Stage 0; $\mathbf{P}^{(n)}$=probability matrix after n transitions; and $\mathbf{S}^T$=transpose of the condition rating vector $\mathbf{S}$. The estimated transition probabilities using the nonlinear optimization approach are plotted in Fig. 2.

Discussion

Correct estimation of transition probabilities in a Markov chain-based deterioration model is a key ingredient for successful and cost-effective proactive management of wastewater systems. The proposed ordered probit model approach presents some advantages in estimating transition probabilities over the approaches developed in the past, including the nonlinear optimization-based approach. This study also identified some limitations of the model in applying the ordered probit model to current wastewater system inspection data. This section discusses the proposed model in terms of versatility in the implementation of the model, the precision of the estimated data, and the appropriateness of assumptions in the model, as compared with the nonlinear optimization-based approach. The limitations of the developed model then follow.

Versatility in Implementation of Model

The ordered probit model approach is more versatile in dealing with deterioration factors. This approach requires only four models (as illustrated in Table 4) to establish a mathematical relationship between the transition probabilities and the explanatory variables. In this model, the explanatory variables can include both categorical variables, such as size and material, and continuous variables, such as age, slope, and length.

In the use of nonlinear optimization-based approach, only the time variable (age of the pipe) can be considered. Thus, if some factors are expected to significantly impact the deterioration process, the entire data set must be categorized accordingly before the analysis and tested as to whether the deterioration models are statistically significantly different after each deterioration model is established. Suppose that three different categories (for instance, material type, pipe diameter,

Table 5. Inspection Schedule for Two Different Sewer Pipes

Sewer pipe	Material	Diameter [mm(in.)]	Length [m(ft)]	Slope	Inspection year A-OPM	Inspection year B-NLO
1	Vitrified clay	200 (8)	198 (650)	0.01	35	49
2	Vitrified clay	200 (8)	229 (750)	0.001	40	49

and soil type), having three entries each, are determined to be statistically significant factors affecting the deterioration process. Then, 27 combinations of different models (3^3) in total are required to describe the deterioration process of the sewer systems.

In addition, due to this categorization, the nonlinear optimization-based approach is inherently limited to categorical variables, and no continuous variables except age can be included in the model. Suppose that a maintenance manager of a wastewater system is planning to establish inspection schedules for two sewer pipes that are different in terms of continuous variables but share the same characteristics based on aspects of categorical variables, as given in Table 5. Assume that the inspection year for the pipes is determined as the expected year that the sewer pipe moves from Condition State 3 into state 4. Then, applying Eq. (19) for both the ordered probit model approach (A-OPM) and the nonlinear optimization-based approach (B-NLO), the inspection schedule can be estimated as shown in Table 5. The ordered probit model approach portrays the impact of continuous variables (length and slope in this case) on the deterioration process, so the planned inspection years for the pipes are different. However, the nonlinear optimization-based approach results in the same inspection year for both pipes, due to the lack of a mechanism involving continuous variables.

Precision of Estimated Data

Figs. 1 and 2 show that the number of transition probability curves, when the ordered probit model approach is used, is greater than that of the nonlinear optimization-based approach. The ordered probit model approach allows multiple transitions (i.e., changes of condition state from 1 to 2, 3, 4, or 5) during a transition period, while only single transitions (i.e., change of condition state from 1 to 2 only) are assumed in the nonlinear optimization-based approach. This single transition assumption requires only eight transition probabilities to be estimated for a zone, as shown in Eq. (17). However, it is possible for a condition state to move more than one state down during a transition period, even though the probability might be very low. Micevski et al. (2002) also points out that the Weibull and Herz models for estimating transition probabilities in Markov chain deterioration models are also based on this single transition assumption, which is not appropriate for representing the true nature of the deterioration process. The ordered probit model approach can provide information about these transitions.

Appropriateness of Assumptions in Model

The ordered probit model generates relatively smooth transition probability curves, as seen in Fig. 1. It allows determination of the annual transition probabilities. However, the nonlinear optimization approach produces step-wise function curves for transition probabilities to years (i.e., the age of the pipe), as shown in Fig. 2. A 6 year term for a zone used in the nonlinear optimization approach in which the transition probabilities in a zone are assumed to be constant prevents the actual representation of the inherent nonstationary nature of deterioration. Moreover, as Madanat et al. (1995) pointed out, the nonlinear optimization-based approach uses the continuous values obtained from the linear regression model, which is not appropriate, because the dependent variable, in this case the condition ratings, is discrete and ordinal. The assumptions of zero error mean and a constant variable are not satisfied and the ordinal scale of the independent variable is not reflected when using the regression model.

Limitations of Model

Even though the developed method using the ordered probit model approach is theoretically and statistically sound, the outputs of the deterioration modeling are not fully satisfactory for analyzing the deterioration of the wastewater systems in the City of San Diego. As shown in Table 4, the measurement of goodness-of-fit for the ordered probit model, the ρ^2 statistics for Condition States 1 and 2, are relatively acceptable. However, the ρ^2 statistics for Condition States 3 and 4 are smaller than expected to account for the effectiveness of the models. These drawbacks may be due to the lack of integrity in the data set. Only cross-sectional data (one time inspection data) were available for the estimation of Markovian transition probabilities in this study. The condition data were collected only during the year 2001. Thus, the increments computed from the condition data do not completely reflect the condition changes of the individual pipe segments. In bridge deterioration modeling studies where an ordered probit model was successfully employed, panel data that contained information about the condition of bridges rated every 2 years for 8 years were utilized (Madanat et al. 1995).

Another reason for the low goodness-of-fit can be attributed to the measurement errors included in the data set. Because the condition assessment of wastewater systems is typically based on closed-circuit TV (CCTV) inspection and the subjective rating of inspectors, these data have a high possibility for errors. To obtain satisfactory results for the deterioration modeling of a wastewater system, it is important to reduce and correct measurement errors in condition ratings. Possible avenues for future research could be the identification of factors for reducing errors in order to facilitate more accurate condition assessment, the development of a multimedia educational tool (web-based or CD-ROM movie clips for various defects) to reduce subjectivity in the condition rating, and the development of mathematical models to correct measurement errors. When sufficient data, such as panel data from periodic condition assessments, are available and further research is performed to reduce the measurement errors in condition assessment, the econometric model based-approach, including the ordered probit model, can be more efficient for the development of deterioration models for wastewater systems.

Conclusions

This paper proposed a method to estimate the transition probabilities for a Markov chain-based deterioration model for wastewater systems using the ordered probit model. The condition assessment data set used to evaluate the developed methodology was obtained from the City of San Diego. The ordered probit model approach provides a theoretically and statistically more robust model as compared to the nonlinear optimization-based approach for the estimation of transition probabilities. However, for developing accurate models using the ordered probit model, it is necessary to have panel data that spans over multiple time periods. In order to predict more accurate and detailed deterioration patterns of wastewater systems, factors such as the depth of the installation, the soil condition, the groundwater level, the frequency of sewage overflows, etc., should be collected and evaluated. In current inspection practices, such information is not readily available for wastewater systems.

A standardized condition rating system is required to generate a more robust deterioration model and to evaluate the deterioration processes of wastewater systems among different municipalities. By employing a standardized condition rating system, current management practices and future investment planning that ensures better performance of the facilities and provision of stable services to the residents of the community can be evaluated. Currently, each municipality uses a different rating system for its wastewater systems. The use of different condition rating systems prevents objective comparison of the effects of maintenance on wastewater systems and information sharing regarding condition assessment among municipalities. Standardized condition rating systems for wastewater systems in England (Water Research Center 1986), the Pavement Condition Index (PCI) (Carnahan et al. 1987), and concrete bridge deck condition ratings (FHwA 1979) can be good benchmarks for wastewater systems.

Acknowledgments

This research was funded in part by a grant from the Purdue Research Foundation and in part by grant CMS-0201364 from the National Science Foundation. Their support is gratefully acknowledged. Any opinions, findings, conclusions, or recommendations expressed in this study are those of the writers and do not necessarily reflect the views of the National Science Foundation or the Purdue Research Foundation. The writers thank Mr. Dean Gipson of the Metropolitan Wastewater Department (MWWD), City of San Diego, for his assistance in providing data used in this study. The writers thank Professor Fred Mannering for his advice in developing and calibrating the ordered probit model.

References

ASCE. (2003). "The 2003 progress report for America's Infrastructure." ⟨http://www.asce.org/reportcard/⟩ (Sept. 9, 2003).

Ariaratnam, S. T., El-Assaly, A., and Yang, Y. (2001). "Assessment of infrastructure inspection needs using logistic models." *J. Infrastruct. Syst.*, 7(4), 160–165.

Ben-Akiva, M., and Gopinath, D. (1995). "Modeling infrastructure performance and user costs." *J. Infrastruct. Syst.*, 1(1), 33–43.

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

Ben-Akiva, M., Humplick, F., Madanat, S., and Ramaswamy, R. (1991). "Latent performance approach to infrastructure management." *Transportation Research Record 1311*, Transportation Research Board, Washington, D.C., 188–195.

Bulusu, S., and Sinha, K. C. (1997). "Comparison of methodologies to predict bridge deterioration." *Transportation Research Record 1597*, Transportation Research Board, Washington, D.C., 34–42.

Butt, A. A., Feighan, K. J., and Shahin, M. Y. (1987). "Pavement performance prediction model using the Markov process." *Transportation Research Record 1123*, Transportation Research Board, Washington, D.C., 12–19.

Carnahan, J. V., Davis, W. J., Shahin, M. Y., Keane, P. L., and Wu, M. I. (1987). "Optimal maintenance decisions for pavement management." *J. Transp. Eng.*, 113(5), 554–572.

Dubin, J. A., and Rivers, R. D. (1987). *Statistical software tools: Version 1.1*, Dubin-Rivers Associates, Riverside, Calif.

Federal Highway Administration (FHwA). (1979). *Recording and coding guide for structure inventory and appraisal of the nation's bridges*, FHwA, United States Department of Transportation, Washington, D.C.

Fenner, R. A. (2000). "Approaches to sewer maintenance: A review." *Urban Water*, 2, 343–356.

Hahn, M. A., Palmer, R. N., Merrill, M. S., and Lukas, A. B. (2002). "Expert system for prioritizing the inspection of sewers: Knowledge base formulation and evaluation." *Urban Water*, 128(2), 121–129.

Herz, R. K. (1998). "Exploring rehabilitation needs for drinking water distribution networks." *Aqua*, 47(6), 275–283.

Jiang, Y., Saito, M., and Sinha, K. C. (1988). "Bridge performance prediction model using the Markov chain." *Transportation Research Record 1180*, Transportation Research Board, Washington, D.C., 25–32.

Jiang, Y., and Sinha, K. C. (1989). "Bridge service life prediction model using the Markov chain." *Transportation Research Record 1223*, Transportation Research Board, Washington, D.C., 24–30.

Kleiner, Y. (2001). "Scheduling inspection and renewal of large infrastructure assets." *J. Infrastruct. Syst.*, 7(4), 136–143.

Lee, Y., and Chang, L. (2003). "Econometric model for predicting deterioration of bridge deck expansion joints." *Proc., 9th Int. Bridge Management Conf.*, Transportation Research Board, Washington, D.C, 255–265.

Madanat, S. M., and Wan Ibrahim, W. H. (1995). "Poisson regression models of infrastructure transition probabilities." *J. Transp. Eng.*, 121(3), 267–272.

Madanat, S. M., Karlaftis, M. G., and McCarthy, P. S. (1997). "Probabilistic infrastructure deterioration models with panel data." *J. Infrastruct. Syst.*, 3(1), 4–9.

Madanat, S. M., Mishalani, R., and Wan Ibrahim, W. H. (1995). "Estimation of infrastructure transition probability from condition rating data." *J. Infrastruct. Syst.*, 1(2), 120–125.

Makar, J. M. (1999). "Diagnostic techniques for sewer systems." *J. Infrastruct. Syst.*, 5(2), 69–78.

Mauch, M., and Madanat, S. (2001). "Semiparametric hazard rate models of reinforced concrete bridge deck deterioration." *J. Infrastruct. Syst.*, 7(2), 49–57.

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), 139–148.

Morcous, G., Rivard, H., and Hann, A. M. (2002). "Modeling bridge deterioration using case-based reasoning." *J. Infrastruct. Syst.*, 8(3), 86–95.

NORVAR. (1998). "Data flow classification of sewers." *Prosjektrapport 76/1997*, Norsk VA-Verksforening, Hamar, Norway.

Prozzi, J. A., and Madanat, S. M. (2000). "Using duration models to analyze experimental pavement failure data." *Transportation Research Record 1699*, Transportation Research Record, Washington, D.C., 87–94.

Ramaswamy, R., and Ben-Akiva, M. (1990). "Estimation of highway pavement deterioration from in-service pavement data." *Transportation Research Record 1272*, Transportation Research Board, Washington, D.C., 96–106.

Ross, S. M. (2000). *Introduction to probability models*, 7th Ed., Academic, San Diego.

Rostum, J., Baur, R., Saegrov, S., Horold, S., and Schilling, W. (1999). "Predictive service-life models for urban water infrastructure management." *Proc. 8th Int. Conf. on Urban Storm Drainage*, 594–601.

San Diego Metropolitan Wastewater Department (San Diego MWWD). (2002). *CCTV Inspection Program Condition Assessment Rep., Phase 1A*, San Diego.

Water Research Center (WRC). (1986). *Sewage rehabilitation manual*, 2nd Ed., Water Authorities Association, Wiltshire, U.K.

Washington, S. P., Karlaftis, M. G., and Mannering, F. L. (2003). *Statistical and econometric methods for transportation data analysis*, CRC, Boca Raton, Fla.

Winston, W. L. (1994). *Operations research applications and algorithms*, 3rd Ed., Duxbury, Belmont, Calif.

Wirahadikusumah, R., Abraham, D., and Iseley, T. (2001). "Challenging issues in modeling deterioration of combined sewers." *J. Infrastruct. Syst.*, 7(2), 77–84.

Wirahadikusumah, R., Abraham, D., Iseley, T., and Prasanth, R. K. (1998). "Assessment technologies for sewer system rehabilitation." *Autom. Constr.*, 7(4), 259–270.

Water Environment Federation and ASCE (WEF/ASCE). (1994). "Existing sewer evaluation and rehabilitation." 2nd Ed., *Manual and Rep. No. 62*, Joint Task Force of the Water Environment Federation and ASCE, New York.