ASSESSMENT OF INFRASTRUCTURE INSPECTION NEEDS USING LOGISTIC MODELS

By Samuel T. Ariaratnam,1 Ashraf El-Assaly,2 Associate Members, ASCE, and Yuqing Yang3

ABSTRACT: Use of various deterioration models in the area of infrastructure management has provided decision makers with a vehicle for predicting future deterioration. This paper presents a methodology for predicting the likelihood that a particular infrastructure system is in a deficient state, using logistic regression models, a special case of linear regression. What distinguishes these two models is that the outcome variable in the logistic regression model is binary or dichotomous and assumes a Bernoulli distribution. The methodology is illustrated in a case study involving the evaluation of the local sewer system of Edmonton, Alta. Canada. Variables of age, diameter, material, waste type, and average depth of cover are modeled, using historical data, as factors contributing to deterioration of the sewer network. The outcome of this model does not produce a prediction of condition rating but rather uses historical inspection records to provide decision makers with a means of evaluating sewer sections for the planning of future scheduled inspection, based on the deficiency probability.

INTRODUCTION

The sewer system is often referred to as the “lifeline” for a municipality and is generally considered as the most cost-intensive infrastructure system in the municipal arena. These systems deteriorate because of a number of factors including excessive usage, aging, characteristics of the surrounding soil, and lack of maintenance. Unfortunately, there has been a serious neglect of sewer systems according to the Water Research Center (WRC) (1993). Because of their lack of visibility, rehabilitation of this “invisible” infrastructure is frequently neglected until a catastrophic failure occurs, resulting in difficult and costly rehabilitation [Water Environment Foundation-ASCE (WEF-ASCE) 1994]. Subsequently, diagnosis and repair of the sewer system is often conducted on a “reactionary” basis, which frequently results in public impact, environmental concerns, and higher construction costs due to short-term “quick-fix” solutions. Additionally, the methods and materials used to repair the deficient sewer primarily address immediate needs, rather than long-term needs, ultimately resulting in prohibitively high costs to the municipality. Even if a sewer system is in service, it is possible that some deficiency already exists, which can worsen with time and result in total failure of the system. If the decision maker has prior knowledge of the location of the sewer to be inspected based on historical likelihood of failure, inspection and assessment may be performed in an economical manner. The choice of rehabilitation approach and materials can then be based on long-term needs rather than a quick-fix solution. Additionally, measures can be taken to minimize the inconvenience to the public and other effects from the rehabilitation, resulting in overall cost savings.

“Proactive” action for sewer rehabilitation is gaining more attention because it allows the decision maker to plan and schedule the inspection and rehabilitation of critical sections prior to the occurrence of emergency response scenarios (Ariaratnam et al. 1998). There are two main avenues to improving sewer rehabilitation planning: (1) collection and storage of adequate inspection information regarding the current condition of the sewer system; and (2) ability to predict sewer deficiency prior to failure to facilitate timely sewer inspection and repair. Because of limited budgets, resource allocation for any maintenance and rehabilitation must be prioritized (Li and Haines 1992). Therefore, an objective method to identify the most urgent need is necessary, as suggested by Quimpo and Wu (1997). The purpose of this paper is not to suggest a course of corrective action but rather to demonstrate the use of a logistic regression modeling application to provide decision makers with a means to prioritize sewer sections for scheduled inspection, given budget constraints.

CONDITION ASSESSMENT

As infrastructure systems age, methods for assessing current and predicting future conditions must be implemented into the planning and programming of inspection and rehabilitation actions. Categorization of infrastructure condition is generally accomplished through a form of condition assessment rating. A review of the literature reveals these ratings as providing the basis for assessing most infrastructure including pavements (Shahin 1994; Hudson et al. 1997), bridges (O’Connor and Hyman 1989; Aktan et al. 1996; Hearnd et al. 1997; Sirsak and Basu 1998), buildings (O’Hara et al. 1997; Uzarski and Burley 1997), railroad tracks (Uzarski 1993), and dams (Andersen and Torrey 1995; Greimann et al. 1997; McKay et al. 1999). Many previous researchers have translated these condition ratings into models for predicting the deterioration state of various infrastructure (Kulkarni 1984; Butt et al. 1987; Feighan et al. 1988; Jiang et al. 1998; Ariaratnam 1994; Lu and Madanat 1994; Ndooh and Ashford 1994; Ben-Akiva and Gopinath 1995; Madanat et al. 1995, 1997; Chouinard et al. 1996; Razaqpur et al. 1996; Livneh 1997; Abraham et al. 1998). These prediction models include the use of straight-line extrapolation, regression models, Markovian models, nonlinear regression, probit models, artificial neural networks, and simulation.

Unfortunately, the condition of sewers in most cities is generally not fully documented. Malik et al. (1997) discovered in a survey of the current state of the practice that only 14% of the cities surveyed revealed having any kind of sewer condition data contained in their infrastructure information systems. This lack of data presents a challenge for municipal engineers in planning for future inspection and maintenance actions. The goal for owners of this infrastructure is to maintain their assets in such a condition and state that failure and the need for emergency repair are minimized. This is particularly relevant when assessing underground sewer utilities in which there is a direct effect on the public sector. Literature reveals the WRC
contains the only documented detailed procedures for establishing the condition of sewer pipes. Currently, there is an initiative under way to train municipalities in inspection procedures to adopt the WRC condition rating system as the national standard.

LOGISTIC REGRESSION MODEL DEVELOPMENT

Background

Regression methods have become an integral component of any data analysis concerned with describing the relationship between a response variable and one or more explanatory variables. Prior to engaging in a study of logistic regression modeling, it is important to understand that the goal of an analysis using this method is the same as that of any model-building technique used in statistics, that is, to find the best fitting and most parsimonious model. A logistic model describes the relationship between an outcome (i.e., dependent or response) and a set of prediction (i.e., independent or explanatory) variables, often referred to as covariates.

What distinguishes these models is that the outcome variable in logistic regression is binary or dichotomous and assumes a Bernoulli distribution. The methods employed in an analysis using logistic regression follow the same general principles used in linear regression. Therefore, the techniques used in linear regression analysis motivate our approach to logistic regression.

To illustrate a multivariable problem, consider a dichotomous deficiency outcome with 0 representing “not deficient” and 1 representing “deficient.” The dichotomous deficiency outcome might be, for example, sewer defects due to cracks, with subjects being classified as either 0 (without cracks) or 1 (with cracks).

To better explain the concept of logistic regression, the logistic function that describes the mathematics behind this regression should be defined. The function f(z) is shown by the following equation:

\[ f(z) = \frac{1}{1 + e^{-z}} \quad (1) \]

The logistic function f(z) ranges between 0 and 1 for the value of z varying from \(-\infty \) to \(+\infty \), as illustrated in Fig. 1.

To obtain the logistic model from the logistic function, the parameter \( z \) is written as the linear sum of the explanatory variables as follows:

\[ z = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k \quad (2) \]

where the \( X_1, \ldots, X_k \) terms are defined as the independent variables of interest; and \( \beta_1, \ldots, \beta_k = \) constant terms representing unknown parameters.

There are two primary reasons for choosing the logistic distribution. The first, from a mathematical point of view, it is an extremely flexible and easily used function. Second, logistic distribution lends itself to a meaningful interpretation. To simplify the notation, consider the term \( \pi(X) = E(Y|X) \) to represent the conditional mean of \( Y \) (i.e., status of deficiency) given \( X \) when the logistic distribution is used. The specific form of the logistic regression model used is as follows:

\[ \pi(X) = \frac{e^{\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots}}{1 + e^{\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots}} \quad (3) \]

where \( \pi \) represents the response probability; \( \beta_0 = \) constant; and \( \beta_1, \beta_2, \ldots = \) regression coefficients for the variable \( X \).

Definition of Logistic Model

Categorical response variables generally have only two categories that refer to binary response variables. The observation for each subject might be classified as a “success” or a “failure” and could be represented by either 1 or 0. When evaluating infrastructure systems, the category of interest is deficiency. For pavements and bridges, this would represent the minimum acceptable level of service as dictated by condition ratings. Therefore, the binary response is the “infrastructure status” categorized as either being deficient (\( Y = 1 \)) or non-deficient (\( Y = 0 \)).

Distribution Assumption

Binary data follow a Bernoulli probability mass function rather than a normal distribution for the response because the response variable is binary and corresponds with the conditions listed below (Agresti 1990). For observations on a categorical variable with two categories, the binomial distribution applies to the sum of the outcomes when the following three conditions hold true:

- For a fixed number of observation \( n \), each falls into one of two categories;
- The probability of falling in each category, \( \pi \), for the first category and \( 1 - \pi \) for the second category, is the same for every observation;
- The outcomes of successive observations are independent; that is, the category that occurs for one observation does not depend on the outcomes of other observations.

The Bernoulli distribution for binary random variables specifies probabilities \( P[Y = 1] = \pi \) and \( P[Y = 0] = 1 - \pi \) for the two outcomes, for which \( \pi = E(Y) \). When \( Y \) has a Bernoulli distribution with parameter \( \pi \), the probability mass function is

\[ f(Y, \pi) = \pi^Y (1 - \pi)^{1 - Y} \quad (4) \]

where \( Y = 0 \) or 1.

EDMONTON LOCAL SEWER NETWORK ANALYSIS

Sewerage systems form one of the most capital-intensive infrastructure systems. When the sewer system deteriorates, waste from excessive infiltration and inflow enters the system, resulting in decreased capacity of the sewer system and increased hydraulic loading. Reducing infiltration can result in lower capital, operating, and maintenance costs to the municipality (Wirahadikusumah et al. 1998). Edmonton owns and maintains an underground sewer network of approximately 4,400 km, worth in excess of $40 billion. The local sewer network currently undergoes approximately 220 km of closed-circuit television (CCTV) inspection per year in an attempt to detect and repair deficient sections prior to collapse. Given budgetary constraints, the development of a methodology to predict the likelihood of a given section of sewer being in a “deficient” state would provide Edmonton with guidance in the planning of these inspections, rather than employing the current practice of random inspections.

After discussions with personnel at the Edmonton Drainage Branch, five factors or explanatory variables (age, diameter, material, waste type, and average depth of cover) that contribute directly to deterioration of the local sewer network were selected for model development. Pipe age, diameter, and average depth of cover are clustered in ranges and as such are defined as interval covariates in the logistic model to take advantage of distinct data categories classifications. Material and waste type are qualitative values and, therefore, are defined as categorical covariates in the logistic model. Categorical variables in any logistic model are the same as dummy variables, which are described in a later section. The response variable of the logistic model is given in binary format. For
binary response variables, the logistic model describes how the proportion of deficient pipes within the overall network depends on the chosen five explanatory variables. For example, let \( \pi \) denote the probability that a randomly selected subject has a deficient response

\[
\log \left( \frac{\pi}{1 - \pi} \right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5
\]

(5)

where \( \beta_0 \) = interception; \( X_1 \) = age; \( X_2 \) = diameter; \( X_3 \) = average depth of cover; \( X_4 \) = waste type; and \( \beta_1, \beta_2, \beta_3, \beta_4, \text{ and } \beta_5 \) represent logistic regression coefficients for the individual variables. The ratio \( \pi/(1 - \pi) \) represents the odds ratio. The odds ratios for a binary response variable is defined as the probability of success \( \pi \) divided by the probability of failure, \( 1 - \pi \).

Data Acquisition

The data acquisition phase included gathering both physical and condition information regarding the local sewer network governed by Edmonton. These records were an aggregated data set obtained from historical inspection records stored in a computerized database system called DRAINage Asset INventory System (DRAINS). The types of data required to achieve the objectives of the research were analyzed and insignificant information in the collected data files were eliminated. The insignificant data included those deemed to have no relation to the attributes of the local sewer network being modeled. For example, data for sewers with diameters <150 mm (i.e., services) and greater than or equal to 1,200 mm (i.e., trunks) were eliminated as they are not contained in the local sewer system. A sample size of 784 CCTV inspection records was obtained from a population size of 8,180 records to serve as a representative sample of the entire local sewer network. These sample records represent sewer sections randomly chosen for inspection from five neighborhoods that were selected based on discussions with personnel at the Drainage Branch. The 784 CCTV records observed satisfy a 95% confidence level with a ±10% margin of error in the sample size. Edmonton personnel separated the records into 26 groups of sewers based on age, diameter, material, waste type, and average depth of cover. The aggregate data were accessed in the development of the logistic model.

One problem in data analysis using statistical measures is how to define the pipe status, that is, how to determine the severity of pipe deterioration in order to classify the pipe status into deficient and nondeficient states. The assessment of pipe deficiency from the CCTV inspections were quantified by Edmonton personnel into five condition levels, from 1 (best) to 5 (worst), with pipe status dependent on the structural condition of the sewer pipe (Standard 1996). Using the rating system adopted by Edmonton (Table 1), sewers with structural condition ratings of 4 or 5 were considered candidates for rehabilitation and were classified as deficient. Subsequently, nondeficient sewers were classified as those with ratings of 1–3. The original ratings of 1–5 could have been used as the dependent variable in the model; however, within Edmonton, all sewers rated between 4 and 5 are grouped together and flagged for rehabilitation whereas any rating from 1 to 3 is labeled as acceptable.

Variable Definitions

Binary Response Variable

The term binary response refers to any variable having only two possible outcomes. In sewer systems, as previously mentioned, pipe status is a binary response variable and may be classified as either deficient or nondeficient.

<table>
<thead>
<tr>
<th>Defect description</th>
<th>Code</th>
<th>Rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cracking light</td>
<td>CL</td>
<td>1</td>
</tr>
<tr>
<td>Corrosion light</td>
<td>HM</td>
<td>1</td>
</tr>
<tr>
<td>Sag light</td>
<td>SL</td>
<td>1</td>
</tr>
<tr>
<td>Open joint light</td>
<td>OL</td>
<td>1</td>
</tr>
<tr>
<td>Sag moderate</td>
<td>SM</td>
<td>2</td>
</tr>
<tr>
<td>Joint displacement light</td>
<td>JL</td>
<td>2</td>
</tr>
<tr>
<td>Open joint moderate</td>
<td>OM</td>
<td>2</td>
</tr>
<tr>
<td>Deformed pipe light</td>
<td>DL</td>
<td>3</td>
</tr>
<tr>
<td>Fracture light</td>
<td>FL</td>
<td>3</td>
</tr>
<tr>
<td>Crack moderate</td>
<td>CM</td>
<td>3</td>
</tr>
<tr>
<td>Corrosion moderate</td>
<td>HM</td>
<td>3</td>
</tr>
<tr>
<td>Deformed pipe moderate</td>
<td>DM</td>
<td>4</td>
</tr>
<tr>
<td>Fracture moderate</td>
<td>FM</td>
<td>4</td>
</tr>
<tr>
<td>Joint displacement moderate</td>
<td>JM</td>
<td>4</td>
</tr>
<tr>
<td>Crack severe</td>
<td>CS</td>
<td>4</td>
</tr>
<tr>
<td>Fracture severe</td>
<td>FS</td>
<td>4</td>
</tr>
<tr>
<td>Corrosion severe</td>
<td>HS</td>
<td>4</td>
</tr>
<tr>
<td>Open joint severe</td>
<td>OS</td>
<td>4</td>
</tr>
<tr>
<td>Sag severe</td>
<td>SS</td>
<td>4</td>
</tr>
<tr>
<td>Collapse of pipe</td>
<td>DX</td>
<td>5</td>
</tr>
<tr>
<td>Broken pipe</td>
<td>FX</td>
<td>5</td>
</tr>
<tr>
<td>Deformed pipe severe</td>
<td>DS</td>
<td>5</td>
</tr>
<tr>
<td>Joint displacement severe</td>
<td>JS</td>
<td>5</td>
</tr>
</tbody>
</table>

Covariates

A regression model can simultaneously handle both quantitative and qualitative explanatory variables. In this case, the model combines elements of standard regression analysis, for which the predictors are quantitative, and analysis of variance, for which the predictors are qualitative. In the logistic regression model, the response variable is a binary variable; all explanatory variables are considered as covariate. For sewer systems, all explanatory variables consist of two types of variables: quantitative and qualitative. The quantitative variables such as pipe age, diameter, and average depth of cover are covariate, and the qualitative variables such as material and waste type are considered categorical covariates.

Dummy Variables

Dummy variables are artificial explanatory variables in a regression model, which represent the categories of the qualitative variable. Each variable assumes one of two values, 0 or 1, indicating whether an observation falls in a particular group. For example, there are three categories when considering waste type: sanitary, storm, and combined. Dummy variable coding is used because it assumes no distance between groups.

Interval Variables

An interval variable is one that has numerical distances between any two levels of a scale, referred to as an interval scale. Pipe age, diameter, and average depth of cover are considered to be interval variables within the interval scale. For example, the interval between 15 and 20 years of age is the same as the interval between 20 and 25 years. Analogously, the interval is also similar for the difference between 200 and 250 mm of pipe diameter and between 250 and 300 mm.

MODEL SELECTION

Linear Regression Variable Selection Method

The linear regression variable selection method enables the user to specify the manner in which independent variables are entered into the analysis. A backward stepwise elimination procedure begins with the complete regression model, one that includes all possible independent variables, and removes se-
lected variables from the model one at a time, based on specific criteria.

For categorical covariates, the reference category must be determined so that all other categories can be compared to this reference. For example, when considering waste type there are three categories: sanitary, storm, and combined. By specifying combined as the reference category, the effect of sanitary and storm waste types to pipe deficiency is represented by (7) that the logistic regression model only containing main effects may be excluded from the final model. We can rewrite (5) so that cover and material as nonsignificant to pipe deficiency and 26 groups of sewer pipes. The results show average depth of Finlay 1997). Tables 2 and 3 present the results of the main hypothesis is true and when it need not be true (Agresti and

### Analysis of Main Effects

Model selection consists of two steps: (1) examination of the significance for each parameter by performing a Wald Test; and (2) determination of parameters to include in the logistic regression model using a likelihood-ratio test.

#### Step 1

The Wald statistic is the square of the z-test. The hypothesis \( H_0: \beta = 0 \) states that \( X_i \) has no effect on the probability \( \pi \) that \( Y = 1 \). It has a chi-square distribution with \( df = 1 \) and the same P-value as the z-test statistic for a two-sided alternative \( H_0: \beta \neq 0 \).

#### Step 2

The likelihood-ratio test compares two models by testing whether the extra parameters in the complete model equal zero. The test refers a key ingredient of maximum likelihood inference, the likelihood function, denoted by \( L \). The equation for the likelihood-ratio test statistic is denoted by (6)

\[
-2 \log \left( \frac{L}{L_0} \right) = (-2 \log L) - (-2 \log L_0)
\]

Eq. (6) compares the values of \( -2 \log L \) when the null hypothesis is true and when it need not be true (Agresti and Finlay 1997). Tables 2 and 3 present the results of the main effects tests in the logistic regression model analysis using the 26 groups of sewer pipes. The results show average depth of cover and material as nonsignificant to pipe deficiency and may be excluded from the final model. We can rewrite (5) so that the logistic regression model only containing main effects is represented by (7)

\[
\log \left( \frac{\pi}{1 - \pi} \right) = -0.0839 + 0.0259X_1
\]

\[-0.0039X_2 + 1.2191Z_1 + 0.3085Z_2\]

#### RESULTS ANALYSIS

An interpretation of the logistic regression coefficient \( \beta \) is as an effect on the odds. Specifically, applying antilogs to both sides of the logistic regression equation, take the simple example

\[
\log \left( \frac{\pi}{1 - \pi} \right) = \alpha + \beta X
\]

which translates into the final model (8) expressed in terms of deficiency probability \( \pi \)

\[
\pi = \frac{e^{\alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3}}{1 + e^{\alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3}}
\]

where \( X_i = \text{age (years)}; X_2 = \text{diameter (mm)}; Z_i \) and \( Z_2 \) = dummy variables for waste type; and \( \beta_1 - \beta_4 \) = coefficients.

### Interaction Evaluation

Cross-product terms allow interaction among explanatory variables in their effects on the response. As previously mentioned, there are three explanatory variables in the current logistic model: age, diameter, and waste type. Subsequently, the possible cross-product terms are age \( \times \) waste type, age \( \times \) diameter, and diameter \( \times \) waste type. Partial correlation analysis performed on the data sets conclude that no interaction exists between age and waste type. Therefore, we should examine the effect of the other two cross products to the pipe status. To compare the effect of these two cross products, we need to compare the simple model (i.e., containing only the main effects)

\[
\log(\pi) = \alpha + \beta_1 \text{age} + \beta_2 \text{diameter} + \beta_3 \text{waste type}
\]

with the full model (containing the main effects with the cross products)

\[
\log(\pi) = \alpha + \beta_1 \text{age} + \beta_2 \text{diameter} + \beta_3 \text{waste type} + \gamma_1 \text{age} \times \text{waste type} + \gamma_2 \text{diameter} \times \text{waste type}
\]

where \( \alpha = \text{constant; and } \beta_1, \beta_2, \beta_3, \gamma_1, \gamma_2 = \text{regression coefficients.} \)

The difference in likelihood ratio for the two models is calculated, which is an approximate chi-square statistic with degree of freedom \( df \) given by the extra parameters in the full model. Further analysis was conducted for the two models to obtain the logistic regression analysis output, which indicates the difference between (10) and (11) to be 7.806 with \( df = 4 \) for the cross-product. Because the \( P \)-value = 0.1 > 0.05, we can accept the hypothesis that the effects of the cross-products are nonsignificant. Therefore, the final logistic regression model is defined as (8).

### Effect of Age on Pipe Deficiency

The logistic regression coefficient \( \beta \) for age is 0.0259 with \( \exp(\beta) = 1.0262 \). This implies that, with each yearly increase of age, the estimated odds of pipe deficiency increases by
2.62%. Subsequently, when age increases by 10 years, the estimated odds of pipe deficiency increases by \((1.0262)^{10} = 1.295\) times. For example, the odds of pipe failure at age \(X_1 = 50\) years is 1.295 times the odds of pipes that are age \(X_1 = 40\). That is, the likelihood of a pipe being in a deficient over nondeficient state increases by 29.5% over this period.

### Effect of Diameter on Pipe Deficiency

The logistic regression coefficient \(\beta\) for diameter is \(-0.0039\) with \(\exp(\beta) = 0.9961\). Because the coefficient \(\beta\) is negative, when the pipe diameter increases by 1 mm, the estimated odds of pipe failure is multiplied by 0.9961, which is a decrease of 0.39%. When the diameter increases by 100 mm, the odds of pipe deficiency is \((0.9961)^{100} = 0.677\) times the original. That is, the likelihood of a pipe being in a deficient state increases by 32.3% when compared to pipes with diameters decreasing by 100 mm.

### Effect of Waste Type on Pipe Deficiency

Given that the effect of combined waste type to the odds of pipe deficiency is 1, the effect of sanitary and storm waste type sewer deterioration is 3.384 and 1.3613, respectively. Therefore, sanitary sewers are found to have the greatest effect on pipe deficiency, followed by storm and then combined sewers.

### Sensitivity Analysis of Model

A sensitivity analysis was performed to validate the proposed model. The logistic model was developed using two scenarios that included using 80 and 90% of the data set. These two models were subsequently compared to the original logistic model. For the purpose of demonstration, a 200-mm-diameter sanitary sewer was used in the validation. Eighty-six scenarios that included using 80 and 90% of the data set.

These two models were subsequently compared to the proposed model. The logistic model was developed using two parameters decreasing by 100 mm.

### Sensitivity Analysis Results (Example: 200-mm-Diameter Sanitary Sewer)

**TABLE 4.** Logistic Regression Model Outcomes

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Coefficient</th>
<th>Upper bound (95% confidence interval)</th>
<th>Lower bound (95% confidence interval)</th>
<th>Standard error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>0.0259</td>
<td>0.034</td>
<td>0.018</td>
<td>0.0042</td>
</tr>
<tr>
<td>Diameter</td>
<td>-0.0039</td>
<td>-0.003</td>
<td>-0.005</td>
<td>0.0006</td>
</tr>
<tr>
<td>Sanitary</td>
<td>1.2191</td>
<td>1.850</td>
<td>0.588</td>
<td>0.3221</td>
</tr>
<tr>
<td>Storm</td>
<td>0.3085</td>
<td>0.592</td>
<td>-0.531</td>
<td>0.2865</td>
</tr>
<tr>
<td>Constant</td>
<td>-0.0839</td>
<td>0.5043</td>
<td>-0.6721</td>
<td>0.3001</td>
</tr>
</tbody>
</table>

### CONCLUSIONS

Although current approaches to prediction of infrastructure condition give insight on the “state” of a system, the model presented in this paper provides information on the likelihood of an infrastructure system being in a deficient state. This knowledge is valuable in identifying candidate infrastructure sections for possible inspection, thus eliminating the randomness often associated with this activity. Subjectivity is reduced as a deficiency probability, based on historical inspection records, is provided rather than a single numerical condition rating. Naturally, one should be cautioned, as the model results will only be as good as the quality of data collected. Using a logistic approach provides a flexible and meaningful model through the use of both qualitative and quantitative variables to distinguish or classify infrastructure sections. The likelihood-ratio test on the historical data revealed that age, diameter, and waste type have a significant effect on the deterioration of the local sewer system owned by Edmonton. The methodology presented may be applied to any infrastructure system (i.e., pavements, bridges, and dams) given prior knowledge of factors contributing to deterioration, critical level of service or deficient state, and historical inspection records.

Future research should be performed to use the logistic model with a financial outlay model to facilitate network-level inspection budget allocation. This would entail collecting historical maintenance and repair cost records and analyzing these against classified infrastructure sections. Timely inspection will serve to detect deficiencies prior to failure and could reduce the occurrence of major rehabilitation, therefore resulting in cost savings and improved allocation of budget.

**TABLE 5.** Sensitivity Analysis Results

<table>
<thead>
<tr>
<th>Age (years)</th>
<th>Probability of Deficiency</th>
<th>80% data sample</th>
<th>90% data sample</th>
<th>100% data sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.5756</td>
<td>0.5940</td>
<td>0.5878</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.5877</td>
<td>0.6061</td>
<td>0.5999</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.5997</td>
<td>0.6180</td>
<td>0.6119</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.6115</td>
<td>0.6298</td>
<td>0.6238</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0.6232</td>
<td>0.6414</td>
<td>0.6355</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0.6348</td>
<td>0.6529</td>
<td>0.6470</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>0.6902</td>
<td>0.7074</td>
<td>0.7020</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>0.7406</td>
<td>0.7565</td>
<td>0.7517</td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>0.7853</td>
<td>0.7997</td>
<td>0.7956</td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>0.8242</td>
<td>0.8369</td>
<td>0.8334</td>
<td></td>
</tr>
<tr>
<td>60</td>
<td>0.8573</td>
<td>0.8684</td>
<td>0.8654</td>
<td></td>
</tr>
<tr>
<td>70</td>
<td>0.8850</td>
<td>0.8945</td>
<td>0.8921</td>
<td></td>
</tr>
<tr>
<td>80</td>
<td>0.9080</td>
<td>0.9160</td>
<td>0.9140</td>
<td></td>
</tr>
<tr>
<td>90</td>
<td>0.9267</td>
<td>0.9334</td>
<td>0.9318</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>0.9419</td>
<td>0.9474</td>
<td>0.9461</td>
<td></td>
</tr>
<tr>
<td>110</td>
<td>0.9540</td>
<td>0.9586</td>
<td>0.9575</td>
<td></td>
</tr>
<tr>
<td>120</td>
<td>0.9638</td>
<td>0.9675</td>
<td>0.9667</td>
<td></td>
</tr>
</tbody>
</table>

\[
T = \frac{12}{N(N + 1)} \sum_{i=1}^{r_j} \frac{r_j^2}{n} - 3(N + 1) \quad (13)
\]

5. Under \(H_0\), \(T\) is approximately the \(\chi^2\) distribution with \((t - 1)\) degrees of freedom.
6. Accept the \(H_0\) at \(\alpha\)-level (i.e., 5% for this test) if \(T \leq \chi^2_{\alpha-1}\).

The sensitivity analysis revealed that \(T = 0.391\) compared with the tabulated \(\chi^2_{0.05,2} = 5.991\). Therefore, the null hypothesis may be accepted, indicating that there is no significant difference between the three models. Thus, the proposed model is stable and may be deemed a good representation of the observed data.
REFERENCES