Title: Identification of most significant factors for modeling deterioration of sewer pipes

Authors: Hind El-Housni[1], Maxim Ouellet[1], and Sophie Duchesne[1]*

[1]INRS-ETE, 490 rue de la Couronne, Québec, QC, G1K 9A9, Canada

*Corresponding author, tel.: 1-418-654-3776, fax: 1-418-654-2600
e-mail: sophie.duchesne@ete.inrs.ca

Word count: 9,435 (6,685 + 8 tables + 3 figures)

Abstract

Existing methods used to identify the important factors that can improve predicting structural deterioration of sewer pipes rarely take into account the interactions and correlations among them. Here we present a standardized method that combines use of the Cox model and likelihood ratio test, and overcomes these limitations of previously employed methods. This combined method is applied to the pipes of two Canadian sewer systems, and its results are compared to the results of two simpler methods for the identification of the factors that significantly influence sewer pipe deterioration. The three methods identified pipe age as the principal factor driving the structural deterioration of sewer pipes. However, slight differences between the methods for other potential influential factors (material, slope and diameter) showed that accounting for the interactions and correlations among factors, as is possible with the proposed method, is crucial to identifying the factors having a significant impact on pipe deterioration.

Keywords: covariates; Cox model; Kruskal-Wallis; likelihood ratio; structural state; survival analysis

Résumé

Les méthodes existantes permettant d’identifier les facteurs d’influence qui doivent être pris en compte dans la modélisation de la détérioration structurale des conduites d’égout prennent rarement en compte les interactions et/ou les corrélations entre ces facteurs. Une méthode standardisée, basée sur l’utilisation combinée du modèle de Cox et du test du rapport de vraisemblance, est proposée dans cet article. Cette méthode est appliquée aux conduites de deux réseaux d'égout canadiens et ses résultats sont comparés aux résultats de deux méthodes plus simples pour l’identification des principaux facteurs influents. Les trois méthodes identifient l’âge des conduites comme étant le principal facteur d’influence dans le processus de détérioration des conduites. Cependant, de légères différences entre les résultats de ces méthodes concernant certains facteurs potentiellement influents (matériau, pente et diamètre) démontrent que la prise en compte des interactions et des corrélations entre les facteurs, rendue possible avec la méthode proposée, est cruciale pour identifier les facteurs ayant un impact significatif.

Mots-clés: analyse de survie; covariables; état structural; Kruskal-Wallis; modèle de Cox; rapport de vraisemblance.

1. Introduction

Many mathematical models exist to predict the structural condition of sewer pipes over time, depending on several variables. These models can be classified into three groups (Ana and Bauwens, 2010): 1) physical models that are based on the physical mechanisms governing the deterioration of pipes (e.g. Konig, 2005); 2) artificial intelligence models (e.g. Tran et al. 2006; Kleiner et al. 2006); and 3) statistical models (e.g. Duchesne et al. 2013). The input data for each of the three model types are pipe condition ratings, which summarize the defects (nature, number, and severity) observed in sewer pipes during televisual inspection. Statistical models, however, remain the most commonly used method to predict the structural condition of sewer pipes (Duchesne et al. 2013).

The principal classes of statistical models that have been applied to the structural deterioration of sewer pipes are: 1) survival models, 2) Markovian models, 3) regression models, and 4) classification models. In survival models, the process of pipe deterioration is represented by the successive transition from one condition state to another (Ana and Bauwens, 2010; Baur et al., 2004). The period of time during which the pipes remain in a given structural state is considered a random variable, described by different distribution functions (e.g. Weibull, exponential, or Hertz), thus defining the process of sewer pipe deterioration over time (Horold and Baur 1999; Mailhot et al. 2000). The result of the survival model is the proportion of sewer lines in a given structural state according to age (Baur and Herz 2002; Duchesne et al. 2013; Ugarelli et al. 2013). Markovian models describe the discrete-time stochastic process whereby the transition probability to the following state class depends only on the current state (Baik et al. 2006; Ross 2000; Wirahadikusumah et al. 2001). This type of model gives the probability that a pipe moves from one condition state to another over a given time interval (Baik et al., 2006; Duchesne et al. 2013; Micevski et al. 2002; Ugarelli et al. 2013). This transition probability is constant over time, although in semi-Markovian models (also known as non-homogenous Markov models), the transition probability varies with time (Ana and Bauwens 2010). In regression models, the probability that a pipe is in a given state after a given period of time is evaluated according to multiple independent variables (Ana et al. 2009; Ana et al. 2008; Ariaratnam et al. 2001; Baik et al. 2006; Chughtai and Zayed 2008; Salman and Salem 2012; Younis and Knight 2010). Regression models can also be used to determine the transition probabilities of Markov models (Baik et al. 2006; Le Gat 2008). Finally, different classification models exist; an example of such a model based on a Random Forest Approach is given in Harvey and McBean (2014).

Variables included in statistical models designed to predict the structural integrity of sewer pipes may be specific to the pipes themselves, such as the age, time of installation, size, length, shape, material, network type, slope, burial depth, and hydraulic performance. Variables may also relate to the environment where the pipes are found, including the soil type, location, level of traffic in close proximity, type of pipe bedding, temperature, and freezes. Several studies have been conducted over the past fifteen years to determine which factors (i.e. independent variables) should be incorporated into models predicting the structural deterioration of sewer pipes. Table 1 provides a summary of these studies and their results.

- **<u>Insert Table 1 here</u>**

The factor that was most frequently identified as having an impact on the structural deterioration of sewer pipes was age, followed by pipe diameter, pipe length, pipe material, network type and pipe slope (Table 1). There was no consensus, however, concerning the method that should be applied to identify the factors that significantly affect sewer pipe deterioration. Also, the factors identified as influential varied greatly among studies because: 1) differences in how factors influence the structural deterioration of sewer pipes mainly depended on the networks where the studies were conducted, and 2) most studies were ad hoc and used different approaches and evaluated different combinations of factors for different networks. Additionally, the interactions among multiple factors were only evaluated in a few studies and correlations among factors were rarely taken into account (except in the case studies conducted by Ariaratnam et al. 2001, and Chughtai and Zayed 2008). In this context, the first objective of the work presented here is to propose a new, robust and standardized method to identify the most influential factors that should be retained in the models predicting the structural deterioration of sewer pipes. The method is based on the likelihood ratio test and the Cox model, used for the first time here to simulate the structural deterioration process of sewer pipes, which makes it possible to integrate several impact factors as uncorrelated explanatory variables. Also, given the large variation in the results found in the literature concerning the influential factors for sewer pipe deterioration, the second objective is to determine the factors that should be considered for modeling the structural deterioration of Canadian sewers, as a function of their main characteristics and the data that are usually available in Canadian municipalities. This information will help guide network managers to the most appropriate deterioration model for their needs. The third objective is to compare the results of the proposed method for the identification of the most influential factors with the results of two simpler methods, one of them often applied in previously published studies. To attain these objectives, analyses were performed using data provided by two Canadian municipalities, as described in the following sections.

2. Methodology

2.1 Case studies

The analyses were performed using data from two different Canadian sewer networks, hereafter referred to as Network A and Network B, located in the province of Quebec. The data provided by network managers consisted of 1) sewer pipe characteristics (installation date, material, diameter, length, and location), and 2) observed structural defects in the pipes that were inspected using a camera, along with the inspection date. The database of the inspected sewer pipes included 15 years of inspections for Network A (1998 to 2012), and 3 years of inspections for Network B (2003 to 2005). In both databases, a “pipe” is defined as a portion of the sewer network located between two manholes or adjacent street junctions, with a constant slope, diameter, and material.

For the analyses presented in this paper, only the results from inspections performed with a zoom camera were retained, as these were the most common type. For Network A, all the observed defects were categorized using the WRc (1994) system, on a scale of 1 to 5. For Network B, the defects were originally characterized using the CERIU (2004) system, thus CERIU grades were converted to WRc grades using the conversion table presented in Duchesne *et al.* (2011). The highest grade for a structural pipe defect was retained to quantify its overall structural state. Another state, state 0, was incorporated into our analyses for pipes without noted defects. Consequently, a pipe in state 0 would have no significant observable structural defects, while a pipe in state 5 would need immediate intervention. Because there were fewer pipes in the worst deterioration states, condition states 2 and 3, and condition states 4 and 5, were grouped together. In summary, the inspected pipes were classified into four distinct deterioration states: 1) state 0 (no structural defect observed); 2) state 1 (only minor structural defects observed); 3) state 2-3 (moderate deterioration); and 4) state 4-5 (one or more severe structural defects observed).

Table 2 presents the main characteristics of the pipes in Networks A and B. Since there is a high level of uncertainty in the installation dates of older pipes, “total network” refers to all the pipes installed in 1900 or later, whereas “inspected pipes” refers to all the pipes that were 70 years old or newer at the time of inspection. In Table 2, "other material" includes asbestos cement, non-reinforced concrete, corrugated steel, brick, cast iron, pipe reinforced with glass fiber, polyethylene, steel, and vitrified clay. Table 3 gives the proportions of the pipes in states 0, 1, 2-3, and 4-5 (corresponding respectively to very good, good, fair and poor structural state) in the different age ranges for Networks A and B.

- **Insert Table 2 here**

- **Insert Table 3 here**

2.2 Estimation of significant influential factors

The applied methodology is summarized in Figure 1. Details are provided in the sections below.

- **Insert Figure 1 here**

2.2.1 Proposition of the Cox method to identify the significant influential factors

For the identification of the influential factors that should be taken into account for predicting the deterioration state of sewer pipes, we propose use of a Cox model, and then to evaluate the statistical significance of the model coefficient related to each influential factor (or covariate) (Figure 1, method A). Only the factors related to the coefficients that have been identified as statistically significant should then be retained in the model.

2.2.1.1 Description of the proposed Cox model

The Cox model is based on survival analysis principles, which represents the time that a pipe remains in each deterioration state as a random variable. This model, widely used in medical science to predict the time before an event (death, recidivism of a disease or cure) occurs (Klein and Moeschberger 2003), has never been used, to our knowledge, for predicting the structural deterioration of sewer pipes (although it was used to predict breaks in water mains; e.g. Andreou *et al*., 1987a and 1987b).

The equations of the Cox model for the modeling of sewer pipes deterioration are developed below for the specific case of four possible structural condition states. However, they could be developed similarly for any number of structural condition states. When four different structural condition states are considered, three residence times t should be modeled. Indeed, once a pipe has entered the fourth and final state it will remain in this state until it is replaced or repaired. Consequently, the residence time in the final state does not have to be modeled. The probability density functions (pdf) of the three residence times, $f_j(t,X)$, are expressed in the proposed model by exponential functions, as suggested by Serpente (1994) and Duchesne *et al*. (2013). Consequently, the pdfs of residence times are expressed as follows (Equation 1):

$$f_j(t,X) = k_j e^{\beta_j X} e^{-k_j e^{\beta_j X} t} \tag{1}$$

with: t = residence time in structural state j; X = vector of explanatory variables; and k_j and β_j = model parameters corresponding to the structural state j (j = 0, 1 or 2). When the β_j coefficients associated with the model covariates X equal zero, the Cox model becomes equivalent to the model of Duchesne *et al.* (2013).

The probability that a pipe will remain longer than time t in any state j is expressed by the survival function $S_j(t,X)$ (Equation 2):

$$S_j\left(t,X\right) = \Pr\left(T > t\right) = \int_t^{\infty} f_j\left(t,X\right) dt \tag{2}$$

At the moment of inspection, only the physical and functional characteristics, including the deterioration state and the age of the pipe are available, but not the time at which it entered its current deterioration state and the previous ones. For example, if a pipe in the fourth deterioration state is inspected 60 years after its installation, 60 years is the total time the pipe was in the first, second, third and fourth states. For this reason the pdf and survival functions for cumulative residence times need to be developed.

The equation for f_{01}, the pdf of the sum of residence times in the first and second states, is given in Equation 3:

$$f_{01}(t,X) = f_0 * f_1(t,X) = \int_{\tau=-\infty}^{\infty} f_0(\tau,X) f_1(t-\tau,X)\, d\tau \qquad (3)$$

where * is the convolution product. As f_0, f_1 and f_{01} are defined only for positive real numbers, then $\tau \geq 0$ and $(t - \tau) \geq 0$, and consequently $\tau \leq t$, thus the previous integral can be simplified to:

$$f_{01}(t,X) = \int_{\tau=-\infty}^{\infty} f_0(\tau,X) f_1(t-\tau,X)\, d\tau = \int_{\tau=0}^{t} f_0(\tau,X) f_1(t-\tau,X)\, d\tau \qquad (4)$$

The corresponding survival function, Equation 5, gives the probability that a pipe will be in the second state or lower (i.e., in the first or second state) at time T:

$$S_{01}(T,X) = \int_{t=T}^{+\infty} f_{01}(t,X)\,dt = 1 - \int_{t=0}^{T} f_{01}(t,X)\,dt = 1 - \int_{t=0}^{T}\int_{\tau=0}^{t} f_0(\tau,X) f_1(t-\tau,X)\, d\tau\, dt \qquad (5)$$

Similarly for f_{012}, the pdf of the sum of the time in the first, second and third states, is given by Equation 6:

$$f_{012}(t,X) = f_{01} * f_2(t,X) = \int_{\tau=0}^{t} f_{01}(\tau,X) f_2(t-\tau,X)\, d\tau \qquad (6)$$

And the probability that a pipe will be in the fourth state or lower (i.e. either in the first, second, third or fouth state) at time T is as presented in Equation 7:

$$S_{012}(T,X) = 1 - \int_{t=0}^{T} \int_{\tau=0}^{t} f_{01}(\tau,X) f_2(t-\tau,X)\, d\tau\, dt \tag{7}$$

The probability that a pipe with age t and having characteristics corresponding to covariates X will be in the first deterioration state is expressed in our model as (Equation 8):

$$P_0\left(t,X\right) = S_0\left(t,X\right) = e^{-k_0 e^{\beta_0 X} t} \tag{8}$$

with: k_0, β_0 = model parameters associated with the residence time in the first condition state.

The probability that a pipe of age t and having characteristics X is in the second, third or fourth (final) structural state (respectively $P_1(t,X)$, $P_2(t,X)$ and $P_3(t,X)$) is computed from the survival functions for cumulative residence times, as given in Equations 9, 10 and 11 respectively.

$$P_1\left(t,X\right) = S_{01}(t) - S_0\left(t,X\right) = \frac{k_0 e^{\beta_0 X} e^{-k_0 e^{\beta_0 X} t} - k_0 e^{\beta_0 X} e^{-k_1 e^{\beta_1 X} t}}{k_1 e^{\beta_1 X} - k_0 e^{\beta_0 X}} \tag{9}$$

$$P_2\left(t,X\right) = S_{012}(t) - S_{01}\left(t,X\right) = \frac{\begin{pmatrix} k_0 k_1 k_2 e^{\beta_0 X + \beta_1 X + \beta_2 X} e^{-k_0 e^{\beta_0 X} t} - k_0 {k_1}^2 e^{\beta_0 X + 2\beta_1 X} e^{-k_0 e^{\beta_0 X} t} \\ - k_0 k_1 k_2 e^{\beta_0 X + \beta_1 X + \beta_2 X} e^{-k_1 e^{\beta_1 X} t} + {k_0}^2 k_1 e^{2\beta_0 X + \beta_1 X} e^{-k_1 e^{\beta_1 X} t} \\ + k_0 {k_1}^2 e^{\beta_0 X + 2\beta_1 X} e^{-k_2 e^{\beta_2 X} t} - {k_0}^2 k_1 e^{2\beta_0 X + \beta_1 X} e^{-k_2 e^{\beta_2 X} t} \end{pmatrix}}{\left(k_2 e^{\beta_2 X} - k_1 e^{\beta_1 X}\right)\left(k_1 e^{\beta_1 X} - k_0 e^{\beta_0 X}\right)\left(k_2 e^{\beta_2 X} - k_0 e^{\beta_0 X}\right)} \tag{10}$$

$$P_3\left(t,X\right) = 1 - S_{012}(t) = 1 - \frac{\begin{pmatrix} k_1 {k_2}^2 e^{\beta_1 X + 2\beta_2 X} e^{-k_0 e^{\beta_0 X} t} - {k_1}^2 k_2 e^{2\beta_1 X + \beta_2 X} e^{-k_0 e^{\beta_0 X} t} \\ - k_0 {k_2}^2 e^{\beta_0 X + 2\beta_2 X} e^{-k_1 e^{\beta_1 X} t} + {k_0}^2 k_2 e^{2\beta_0 X + \beta_2 X} e^{-k_1 e^{\beta_1 X} t} \\ + k_0 {k_1}^2 e^{\beta_0 X + 2\beta_1 X} e^{-k_2 e^{\beta_2 X} t} - {k_0}^2 k_1 e^{2\beta_0 X + \beta_1 X} e^{-k_2 e^{\beta_2 X} t} \end{pmatrix}}{\left(k_2 e^{\beta_2 X} - k_1 e^{\beta_1 X}\right)\left(k_2 e^{\beta_2 X} - k_0 e^{\beta_0 X}\right)\left(k_1 e^{\beta_1 X} - k_0 e^{\beta_0 X}\right)} \tag{11}$$

where: k_1 and β_1 = model parameters associated with the residence time in the second deterioration state; k_2 and β_2 = model parameters associated with the residence time in the third deterioration state.

2.2.1.2 Choice of the covariates and verification of the absence of correlations between them

Factors (covariates) integrated in the Cox model should not be correlated, thus a correlation test should be used to identify possible correlations among the factors before constructing the Cox model. In the work presented here, the Spearman method was used to determine the correlations among the factors. As described in many handbooks on statistics (e.g., Sheskin 2003), this test measures the degree of association, linear or not, between two variables, even for those which are ordinal.

Also, the factors that should be included in *X*, the vector of explanatory variables, should be factors for which extensive data are available on the studied network and for which an impact on sewer structural deterioration is suspected a priori. The factors selected for the analysis presented here were those for which data are generally available for Canadian sewer networks and that could affect the overall performance and structural state of the pipes. These factors are grouped into two categories: physical and functional factors. The first category includes general pipe characteristics such as the age, diameter, length, material, and slope, while the last concerns the type of network. The selected factors have been frequently identified as influential factors in previous studies (see Table 1). Pipe age is not included in the vector of explanatory variables since it appears explicitly in the Cox model as the variable *t*.

Finally, the covariates can be quantitative and/or qualitative. However, qualitative variables require specific coding (as ordinal or binary variables) to include them in the model. In the present case study, the material type and network type were qualitative variables. In order to evaluate their potential correlations with other factors, and to enable their integration into the Cox model, these variables required coding in the form of variable indicators X_i. The number of variable indicators (i) varied according to the number of categories included in the same variable, with i equal to one less than the number of categories (Klein and Moeschberger, 2003). For example, within Network B, sewer type was coded as follows: combined sewer, X_1 = 1 and X_2 = 0; storm water sewer, X_1 = 0 and X_2 = 1; and sanitary sewer, X_1 = 0 and X_2 = 0. All codes for the qualitative variables are provided in the supplementary material (Table S-1).

2.2.1.3 Calibration of the Cox model parameters

Before using the model, all of its parameters k_0, k_1, k_2, β_0, β_1 and β_2 should be estimated, based on the condition state observed during televisual inspections for a sample of representative sewer pipes. They remain specific to each sewer system and must be adjusted according to the inspection results, but may be subsequently verified using a cross-validation method as carried out in Duchesne *et al.* (2013). In the present study, the calibration of the Cox model was performed using the maximum likelihood method. This consisted of estimating the values of the parameters that maximized the likelihood function given in Equation 12.

$$L = \prod_{k\in cd_0}^{n_{cd_0}} P_0(t_k, X_k) \prod_{k\in cd_1}^{n_{cd_1}} P_1(t_k, X_k) \prod_{k\in cd_2}^{n_{cd_2}} P_2(t_k, X_k) \prod_{k\in cd_3}^{n_{cd_3}} P_3(t_k, X_k) \qquad (12)$$

with t_k = age of inspected pipe k (years); X_k = values of the covariates for inspected pipe k; cd_j = all inspected pipes for which the state was equal to j at inspection at age t_k; and ncd_j = the number of pipes in the set cd_j.

3.2.1.4 Determination of significance of each covariate

In the proposed method, the statistical significance level of parameters in the vectors β_0, β_1 and β_2 is tested with the likelihood ratio test (Thiombiano 2013; Klein and Moeschberger 2003). This test verifies if the coefficients corresponding to the factors integrated into the model (i.e. elements of β_0, β_1 and β_2) are significantly different from zero. It is based on the calculation of the distance λ between the logarithm of the likelihood function, calculated with the $\beta(X_i)$ values estimated during calibration (different from zero; unrestricted model, β_{unr}), and the logarithm of the likelihood function calculated with $\beta(X_i)$, the value of the coefficient for the analyzed factor, forced to zero (restricted model, β_{res}) (Equation 13).

$$\lambda = 2\left[LL(\beta_{unr}) - LL(\beta_{res})\right] \tag{13}$$

Under the null hypothesis ($\beta(X_i) = 0$), λ follows a Chi-square law (χ^2_α) with the number of degrees of freedom equal to the number of imposed constraints for this hypothesis (number of parameters forced to zero) (Thiombiano 2013; Klein and Moeschberger 2003). If the probability that the null hypothesis is valid is greater than the chosen significance level ($\alpha = 0.05$ here), then the restricted model is accepted. In the opposite case, the unrestricted model is accepted and the coefficient corresponding to the evaluated factor is judged significantly different from zero.

2.2.2 Comparison of results with those of two simpler common methods

The results of the above method for the evaluation of the significant factors, which should be considered to predict the structural deterioration of sewer pipes, were compared to those of two simpler methods. The first method (Figure 1, method B) involves the separation of pipes into cohorts sharing common characteristics, according to the factors analyzed, and the comparison of their deterioration curves as computed with the Cox model without covariates. Table 4 summarizes the cohorts created for this method and their characteristics. Each cohort should contain a sufficient number of pipes in order to be able to establish significant statistical relationships between the age and structural state; for this reason, the impact of the slope could not be analyzed using this method. Accordingly, for Networks A and B, respectively, twelve and eleven models were calibrated (one for each cohort). These models were then used to calculate, for each cohort, the probability of being in each of the considered structural states, with respect to the age of the pipes. This kind of analysis is similar to those conducted by Ana *et al.* (2008), Baur and Herz (2002), Duchesne *et al*. (2013), and Micevski *et al*. (2002).

- ➢ **Insert Table 4 here**

The second method (Figure 1, method C) involves evaluating whether the distributions of the values for each factor are significantly different among the structural states. When the distribution for a given factor differs according to the structural state of the pipes, it can be considered to be influential to the process of structural deterioration. This evaluation was performed using the Kruskal-Wallis test statistic (details are found in Sheskin 2003), which allows for the comparison of the distributions of two or more samples of different sizes. However, this test can only be used to compare the distributions of quantitative data. Therefore, only the impacts of age, diameter, length, and slope were evaluated with this method for the pipes of Network A. This also held true for Network B, except for the slope, which was not available.

3. Results and discussion

3.1 Statistical significance of factors according to the proposed method using the Cox model

The first step before applying the Cox model is to evaluate the correlation between covariates. The results of the Spearman test are presented in Table 5 for Networks A and B.

- ➢ **Insert Table 5 here**

These results demonstrate that correlations exist between some of the factors for the two networks. For Network A, the pipe diameter, type of network, and type of material are correlated. For Network B, correlations are noted between the pipe diameter and network type, as well as between the pipe diameter and material, and between the material and network type (combined or other).

Following these observations, the pertinence of each covariate in the Cox model was first evaluated for Networks A and B without accounting for the observed correlations (section 3.1.1). Next, the impacts of these correlations were estimated (section 3.1.2). In both cases and as mentioned above, the statistical significance of each factor was evaluated using the likelihood ratio test.

3.1.1 Statistical significance of factors in Cox model without consideration of correlations between covariates

In this section, a “global” Cox model that integrates all of the evaluated factors for each network is established. The statistical significance of the coefficients corresponding to each covariate for Networks A and B are presented in Table 6 (in this table, for all coefficients, *p_value* = 1 and $\chi^2_{0.05} = 3.84$).

- ➢ **Insert Table 6 here**

As reported in the two tables, none of the coefficients are significantly different from zero (p_value = 1). This suggests that the factors evaluated for the two networks do not significantly impact the aging process of the pipes. However, the lack of statistical significance for some of these factors may be the result of correlations among the variables or correlations between the variables and the age of the pipes (Table 5). The following section describes how these correlations were taken into account.

3.1.2 Statistical significance of factors in Cox model with consideration of correlations between covariates

To evaluate the impacts of the correlations between factors on the results of the likelihood ratio test, several other models were tested. These models only integrated the covariates that were not correlated. Thus, based on the previously identified correlations, two models were tested for Network A. The first model only takes into account the covariates diameter, length, and slope. The second model incorporates the covariates network type and material, in the place of the pipe diameter. For Network B, once again based on the previously identified correlations, three models were created: 1) the first includes the diameter and length; 2) the second includes the material and length; and 3) the third includes the type of network and the length. Results of the likelihood ratio tests for the different models for Networks A and B are presented in Table 7 (in this table, for all coefficients, *p_value* = 1 and $\chi^2_{0.05} = 3.84$).

- **Insert Table 7 here**

Again, despite the inclusion of correlations between certain covariates of the two networks, the likelihood ratio test demonstrated that none of the covariates except for age were influential in the structural deterioration process of the pipes in these networks (all p_values = 1). Similar results were obtained by testing models that incorporate only the age of the pipes and one covariate at a time.

Based on these results, the Cox model without covariates was finally used to estimate the probability that a pipe will be in structural state 0, 1, 2-3, or 4-5 as a function of its age, for Networks A and B. Figure 2 shows a comparison of the estimated and observed proportions for Network A. This figure demonstrates the ability of the model to reproduce the current structural condition of the pipes. However, this visual comparison provides only a partial estimation of this ability, because each point represents a different number of pipes. The value of the likelihood function (used previously in the likelihood ratio test to compare the different models) is a better estimator of this ability. A slight overestimation of the probabilities for young pipes to be in state 0 can be noted in Figure 2; this is due to the fact that the curve must pass through one for 0-year-old pipes. In addition, the probabilities for older pipes (61-70 years old) to be in deterioration states 0, 2-3, and 4-5 seem to be less well estimated than for other pipe ages; this is due to the limited number of inspected pipes in this age range.

- ➢ **<u>Insert Figure 2 here</u>**

3.2 Comparison of structural deterioration of pipes separated according to evaluated factors

The results of the Cox model without covariates, separating the pipes into cohorts, are presented in this section. Because of the large number of cohorts, only the probabilities associated with the final (4-5) state are presented. Figure 3 illustrates the probabilities that pipes in Network A will be in state 4-5. Note that in Figure 3b, the probability curves for the two sewer type cohorts overlap, whereas in Figure 3d, the curves corresponding to the pipes with lengths less than 60 m and those with lengths between 60 and 120 m also overlap. Results for Network B are provided in the supplementary material (Figure S-1; in this figure, the probability curves for the combined and stormwater sewers overlap in panel (b) and the two curves overlap in panel (d)).

- **<u>Insert Figure 3 here</u>**

As shown in Figure 3, the application of the Cox model without covariates to different cohorts of pipes in Network A demonstrates that the structural conditions of pipes evolve similarly for different cohorts. Older pipes are more likely to be damaged, regardless of either their physical or functional characteristics. However, for the majority of factors defining the cohorts, slight differences can be noted for the probabilities that pipes will be in state 4-5 over time. Most of these differences are very small. However, the largest differences are found for the type of material for the pipes in Network A (Figure 3c). Therefore, this factor seems to affect the deterioration process in the pipes of Network A, a priori. There are also marked differences for the pipes of Network B, classified by diameter, in relation to the probabilities that the pipes will be in state 4-5 (Figure S-1a); the diameter can therefore be considered to be a potentially influential factor for Network B. This evaluation method remains visual, and the results greatly depend on the amount of data for each cohort used for model calibration.

3.3 Comparison of distributions of factors classified by structural state

Table 8 presents the results of the Kruskal-Wallis test for comparison of distributions of the studied factors between very good and poor structural condition states (0 and 4-5) for Networks A and B. Box diagrams showing the distributions of the factors for Networks A and B for the same two structural states, 0 and 4-5, are presented in the supplementary material (Figure S-2).

- **<u>Insert Table 8 here</u>**

Based on a significance threshold of 0.05, it is evident that the pipe age has a significant impact on the structural deterioration process for the pipes in the studied networks (p_value < 0.05; the pipes in state 4-5 tend to be older than those in state 0). This method also highlights the possible impact that a pipe's slope has on the aging of pipes in Network A; the pipes in state 4-5 have greater mean slopes than those in state 0. However, this method, like the preceding one, does not take into account the possible correlations among factors. Additionally, the use of this method is limited by its dependence on the type of available data.

3.4 Comparison of results obtained from three methods

All of the assessments performed with the Cox model combined with the likelihood ratio test demonstrate that age alone can explain, in a significant manner, the structural conditions of the pipes in Networks A and B. The addition of other factors as covariates in the Cox model does not improve the prediction of the structural states of these pipes over time. This includes factors that were determined to be influential in the aging process using the two simpler methods. However, the identification of some influential factors with the simpler methods (e.g., the material type for Network A and pipe diameter for Network B) may be the result of correlations among the variables or correlations between the variables and the age of the pipes (Table 5). Indeed, the two simpler methods do not take into account the possible correlations among factors. Additionally, the use of a method that compares the distributions of factors among the different structural states (using, for example, the Kruskal-Wallis test) cannot take into account the possible correlation between the age and some influential factors; it is also limited by its dependence on the type of available data (quantitative and continuous). The Cox model method presents the following advantages over the two simpler methods: 1) it does not require grouping data together according to a given characteristic (which in some cases may reduce the amount of data and thus hamper a statistical analysis), and 2) it can treat both quantitative and qualitative data.

The fact that no significant factors other than age were found using this method, including factors that are often identified as important by other researchers and other methods (e.g., diameter, material, and type of network), may be explained by the amounts and types of data available to assess the impact of each of these factors. For example, the majority of the inspected pipes in Networks A and B were made of concrete, which makes it difficult to effectively evaluate the impact of the type of pipe material on the deterioration process of the pipes. If one assumes that Networks A and B, and the data that are available to characterize them, are representative of most Canadian wastewater systems, age would remain the only significant factor that would need to be taken into account to model the structural deterioration of these networks.

4. Conclusion

In this article, a new robust and standardized method, based on the use of the Cox model, was proposed to identify the most influential factors, which should be taken into account when modeling the structural deterioration of sewer pipes. To the best of our knowledge, this was the first time that the Cox model has been used to model the structural deterioration of sewer pipes. A calibration method was also proposed to apply the Cox model to the mathematical representation of a series of successive degradation states. Then, the impacts of physical and functional factors on the structural deterioration of pipes of two Canadian sewer networks were assessed using this method and two simpler ones. Results of this comparison highlighted the importance of evaluating interactions and correlations among factors. Use of simpler methods for the identification of the significant influential factors should thus be avoided.

Within the two networks examined here, pipe age was unanimously identified by the three methods as the main factor influencing sewer pipe deterioration state. This was particularly evident using the Cox model with covariates, which demonstrated that taking the age alone into account could provide satisfactory predictions of the structural states of the pipes of the studied networks. If databases were available that included, for example, information on the structural states of several pipes made of different types of material and having a wide range of different ages, the results obtained could have been quite different. However, considering that the networks studied are representative of Canadian networks, and the data included variables that are generally available for these networks, it is unlikely that the integration of factors other than age in structural deterioration models would significantly benefit these networks.

The model and assessment methods presented in this article are useful tools for planning the renewal of sewer pipes. In addition to the structural deterioration of the network, this planning should ideally take into account the evolution of the hydraulic performances of sewers, which is all the more critical in the context of climate change. Future work should focus on the integration of these aspects in order to improve renewal procedures, reduce the costs associated with them, and improve the overall performance of sewage systems.

Acknowledgements

The authors wish to thank the City of Quebec and another anonymous municipality for providing the data used in this study. The authors are also grateful to Prof. Alain Mailhot from INRS who helped with the development of the likelihood function.

References

Ana, E.V., and Bauwens, W. 2010. Modeling the structural deterioration of urban drainage pipes: the state-of-the-art in statistical methods. Urban Water Journal, **7**(1): 47-59. doi: 10.1080/15730620903447597.

Ana, E.V., Bauwens, W., Pessemier, M., Thoeye, C., Smolders, S., Boonen, I., and De Gueldre, G. 2009. An investigation of the factors influencing sewer structural deterioration. Urban Water Journal, **6**(4): 303-312. doi**:** 10.1080/15730620902810902.

Ana, E.V., Bauwens, W., Pessemier, M., Thoeye, C., Smolders, S., Boonen, I., and De Gueldre, G. 2008. Investigating the effects of specific sewer attributes on sewer ageing–a Belgian case study. *In* 11th International Conference on Urban Drainage, Edinburgh, UK. 31 August - 5 September 2008.

Andreou, S.A., Marks, D.H., and Clark, R.M. 1987a. A new methodology for modelling break failure patterns in deteriorating water distribution systems: Theory. Advances in Water Resources, **10**(1): 2-10. doi**:** 10.1016/0309-1708(87)90002-9.

Andreou, S.A., Marks, D.H., and Clark, R.M. 1987b. A new methodology for modelling break failure patterns in deteriorating water distribution systems: Applications. Advances in Water Resources, **10**(1): 11-20. doi: 10.1016/0309-1708(87)90003-0.

Ariaratnam, S.T., El-Assaly, A., and Yang, Y. 2001. Assessment of infrastructure inspection needs using logistic models. Journal of Infrastructure Systems, **7**(4): 160-165. doi: 10.1061/(ASCE)1076-0342(2001)7:4(160).

Baik, H.S., Jeong, H.S., and Abraham, D.M. 2006. Estimating transition probabilities in Markov chain-based deterioration models for management of wastewater systems. Journal of Water Resources Planning and Management, **132**(1): 15-24. doi: 10.1061/(ASCE)0733-9496(2006)132:1(15).

Baur, R., Zielichowski-Haber, W., and Kropp, I. 2004. Statistical analysis of inspection data for the asset management of sewer networks. *In* Proceedings of the 19th EJSW on Process Data and Integrated Urban Water Modeling, Lyon, France. March 2004.

Baur, R., and Herz, R. 2002. Selective inspection planning with ageing forecast for sewer types. Water Science and Technology, **46**(6): 389-396.

CERIU 2004. Manuel de standardisation des observations - Inspections télévisées des conduites et regards d'égout (in French). Centre d'expertise et de recherche en infrastructures urbaines, Montréal, Canada.

Chughtai, F., and Zayed, T. 2008. Infrastructure condition prediction models for sustainable sewer pipelines. Journal of Performance of Constructed Facilities, **22**(5): 333-341. doi: 10.1061/(ASCE)0887-3828(2008)22:5(333).

Duchesne, S., Villeneuve, J.-P., Beardsell, G., Toumbou, B., and Bouchard, K. 2011. Modélisation de la détérioration structurale des conduites d'égout de la ville de Québec (in French). Research Report No R1287. Institut National de la Recherche Scientifique, Université du Québec, Québec, Canada.

Duchesne, S., Beardsell, G., Villeneuve, J.-P., Toumbou, B., and Bouchard, K. 2013. A survival analysis model for sewer pipe structural deterioration. Computer-Aided Civil and Infrastructure Engineering, **28**(2): 146-160. doi: 10.1111/j.1467-8667.2012.00773.x.

Fuchs-Hanusch, D., Günther, M., Möderl, M., and Muschalla, D. 2015. Cause and effect oriented sewer degradation evaluation to support scheduled inspection planning. Water Science and Technology, **72**(7): 1176-1183. doi: 10.2166/wst.2015.320.

Harvey, R.R., and McBean, E.A. 2014. Predicting the structural condition of individual sanitary sewer pipes with random forests. Canadian Journal of Civil Engineering, **41**(4): 294-303. doi: 10.1139/cjce-2014-0037.

Horold, S., and Baur, R. 1999. Modelling sewer deterioration for selective inspection planning: Case study Dresden. *In* Proceedings of the 13th European Junior Scientist Workshop. Switzerland. 8-12 September 1999.

Klein, J.P., and Moeschberger, M.L. 2003. Survival Analysis: Techniques for Censored and Truncated Data - 2nd edition. Springer, New York, NY.

Kleiner, Y., Sadiq, R., and Rajani, B. 2006. Modelling the deterioration of buried infrastructure as a fuzzy Markov process. Journal of Water Supply: Research and Technology-AQUA, **55**(2): 67-80. doi: 10.2166/aqua.2006.074.

Konig, A. 2005. CARE-S WP2 External Corrosion Model Description. SINTEF Technology and Society, Trondheim, Norway.

Le Gat, Y. 2008. Modelling the deterioration process of drainage pipelines. Urban Water Journal, **5**(2): 97-106. doi: 10.1080/15730620801939398.

Mailhot, A., Duchesne, S., Musso, E., and Villeneuve, J.-P. 2000. Modélisation de l'évolution de l'état structural des réseaux d'égout: application à une municipalité du Québec (in French). Canadian Journal of Civil Engineering, **27**(1): 65-72.

Micevski, T., Kuczera, G., and Coombes, P. 2002. Markov model for storm water pipe deterioration. Journal of Infrastructure Systems, **8**(2): 49-56. doi: 10.1061/(ASCE)1076-0342(2002)8:2(49).

Rokstad, M.M., and Ugarelli, R.M. 2015. Evaluating the role of deterioration models for condition assessment of sewers. Journal of Hydroinformatics, **17**(5): 789-804. doi: 10.2166/hydro.2015.122.

Ross, S.M. 2000. Introduction to probability models - 7th ed. Academic Press, San Diego, CA.

Salman, B., and Salem, O. 2012. Modeling failure of wastewater collection lines using various section-level regression models. Journal of Infrastructure Systems, **18**(2): 146-154. doi: 10.1061/(ASCE)IS.1943-555X.0000075.

Serpente, R.F. 1994. Understanding the modes of failure for sewers. *In* Urban Drainage Rehabilitation Programs and Techniques; Selected Papers on Urban Drainage Rehabilitation from 1988-1993 Water Resource Planning and Management Division Conference Sessions. *Edited by* W.A. Macaitis. ASCE, New York, NY, pp. 86-100.

Sheskin, D.J. 2003. Handbook of Parametric and Nonparametric Statistical Procedures: Third Edition. Chapman & Hall/CRC, New York, NY.

Thiombiano, T. 2013. Économétrie des modèles dynamiques : 2^{e} édition (in French). L'Harmattan, Paris, France.

Tran, D.H., Ng, A.W.M., Perera, B.J.C., Burn, S., and Davis, P. 2006. Application of probabilistic neural networks in modelling structural deterioration of stormwater pipes. Urban Water Journal, **3**(3): 175-184. doi: 10.1080/15730620600961684.

Tscheikner-Gratl, F., Mikovits, C., Rauch, W., and Kleidorfer, M. 2014. Adaptation of sewer networks using integrated rehabilitation management. Water Science and Technology, **70**(11): 1847-1856. doi: 10.2166/wst.2014.353.

Ugarelli, R.M., Selseth, I., Le Gat, Y., Rostum, J., and Krogh, A.H. 2013. Wastewater pipes in Oslo: from condition monitoring to rehabilitation planning. Water Practice and Technology, **8**(3-4): 487-494. doi: 10.2166/wpt.2013.051.

Wirahadikusumah, R., Abrahamm D., and Iseley, T. 2001. Challenging issues in modeling deterioration of combined sewers. Journal of Infrastructure Systems, **7**(2): 77-84. doi: 10.1061/(ASCE)1076-0342(2001)7:2(77).

WRc 1994. Manual of Sewer Condition Classification – 3rd Edition. Water Resources Center, Swindon, UK.

Younis, R., and Knight, M.A. 2010. A probability model for investigating the trend of structural deterioration of wastewater pipelines. Tunnelling and Underground Space Technology, **25**(6): 670-680. doi: 10.1016/j.tust.2010.05.007.

Table 1. Summary of studies that assessed the impacts of different variables on the sewer structural deterioration process

Authors	Model type used	City of application	Evaluated factors	Influential factors	Statistical tests used
Ariaratnam et al. (2001)	Logistic regression	Edmonton * (Canada)	Age; Diameter; Depth; Sewer type; Material; Interactions between factors (Age; Diameter; Sewer type)	Age; Diameter; Sewer type	Wald Likelihood ratio
Micevski et al. (2002)	Markov model	Newcastle ** (Australia)	Diameter; Material; Soil type; Exposure classification (distance from the coastline); Hydraulic performance	Diameter; Material; Soil type; Exposure classification	Chi squared
Baur and Herz (2002)	Cohort survival model	Dresden* (Germany)	Period of construction; Material; Location relative to other infrastructure (road network); Diameter; Slope; Sewer type; Shape; Sewer function (minor or major network)	Material; Period of construction; Location relative to other infrastructure	Visual comparison
Baik et al. (2006)	Markov and ordered probit model	San Diego † (United States)	Age; Length; Diameter; Material; Slope	Age; Length; Diameter; Slope	Measure of overall statistical fit: ρ^2 (for ordered probit model)
Tran et al. (2006)	Multiple discriminant analysis regression	Greater Dandenong** (Australia)	Age; Diameter; Depth; Slope; Location (reserve, under road, under nature strip, under easement); Trees root presence; Hydraulic performance; Soil type; Soil moisture index at dry condition	According to the chi squared test and stepwise method: Hydraulic performance According to ANOVA 1: Slope; Depth	ANOVA 1 Stepwise method Univariate Analysis with Chi squared
Chughtai and Zayed (2008)	Multiple regression	Niagara Falls and Pierrefonds (Canada)	Age; Diameter; Length; Material; Depth; Slope; Bedding factor; Location relative to other infrastructure (road network category); Interactions between some factors	Concrete pipes: Age associated to material; Location relative to other infrastructure; Depth; Depth associated to Bedding factor; Bedding factor Asbestos cement pipes: Age; Depth	t-test

Authors	Model type used	City of application	Evaluated factors	Influential factors	Statistical tests used
				associated to Length PVC pipes: Age; Length; Location relative to other infrastructure; Bedding factor; Diameter associated to Depth	
Ana et al. (2008)	Cohort survival model	Leuven* (Belgium)	Period of construction; Diameter; Length; Shape; Material; Slope; Depth; Location relative to other infrastructure (road network)	Period of construction; Length; Material	Visual comparison
Ana et al. (2009)	Multiple logistic regression	Leuven* (Belgium)	Age; Period of construction; Diameter; Length; Shape; Material; Sewer type; Slope; Depth; Location relative to other infrastructure (road network)	Age; Material; Length	Wald Likelihood ratio
Younis and Knight (2010)	Ordinal regression	Niagara Falls † (Canada)	Age; Material; Interaction between Age and Material	Material; Interaction between Age and Material (Age is influent only for concrete pipes)	Wald
Salman and Salem (2012)	Multinomial logistic regression	Cincinnati †† (United States)	Age; Diameter; Length; Material; Sewer type; Slope; Location relative to other infrastructure (road network); Depth; Interactions between factors: 27 two-way interactions (except interaction between Sewer type and Material) and 5 three-way interactions	Age; Diameter; Length; Material; Sewer type; Slope; Location relative to other infrastructure; Depth; 13 two-way interactions and 2 three-way interactions	Wald (for factors) Stepwise method (for interactions between factors)

Authors	**Model type used**	**City of application**	**Evaluated factors**	**Influential factors**	**Statistical tests used**
	Binary logistic regression		Age; Diameter; Length; Material; Sewer type; Slope; Location relative to other infrastructure (Road network); Depth; Interactions between factors: 20 two-way interactions (except interaction between Sewer type and Material)	Age; Diameter; Length; Slope; Material; Sewer type; Location relative to other infrastructure; Depth; 7 two-way interactions	Wald (for factors) Stepwise method and Likelihood ratio (for interactions between factors)
Ugarelli et al. (2013)	GompitZ deterioration modeling tool using the combination of Markov and survival model	Oslo †† (Norway)	Diameter; Sewer type; Period of construction; Road traffic; Soil type; Tramway proximity; Trees presence; Interactions between factors	For concrete pipes up to 600 mm: Diameter; Sewer type; Period of construction; Soil type; Trees presence; Interaction between Sewer type and trees presence	Likelihood ratio
Rokstad and Ugarelli (2015)	GompitZ deterioration modeling tool (non-homogeneous Markov Chain model)	Oslo * (Norway)	Diameter; Sewer type; Period of construction; Road traffic; Bedding factor; Trees presence	For all pipes material : Diameter; Sewer type; Period of construction; Road traffic; Bedding factor; Trees presence	Chi squared
Fuchs-Hanusch et al. (2015)	Logistic regression models	Unnamed * (Austria)	Material; Vintage (period of construction); Sewage type; Profile type; Width; Height; Length; Depth	Material; Length; Width; Vintage; Profile type	Likelihood ratio
Tscheikner-Gratl et al. (2014)	Binary logistic regression	Alpine city ††† (Austria)	Age; Diameter; Material Slope; Length; Shape	Age; Diameter; Slope; Length; Shape	Wald

* : Total network: sanitary, stormwater and combined sewers
** : Stormwater sewer
† : Sanitary sewer
†† : Sanitary and combined sewers
††† : Combined sewer

Table 2. Characteristics of sewer pipes in Networks A and B

	Network A		**Network B**	
Installation year	**Whole network (%)**	**Inspected pipes (%)**	**Whole network (%)**	**Inspected pipes (%)**
1900-1929	5.0	7.0	0.7	0.0
1930-1949	6.1	8.1	1.7	1.7
1950-1969	25.4	32.5	44.32	55.6
1970-1989	36.6	36.9	33.9	35.7
1990-2012	21.8	13.6	19.2	6.8
Unknown	4.8	1.6	0.0	0.0
Sewer type	**Total length (km)**	**Inspected length (km)**	**Total length (km)**	**Inspected length (km)**
Combined	586	383	442	70
Stormwater	1,972	53	424	43
Sanitary	2,067	1,213	420	46
Diameter	**Total length (km)**	**Inspected length (km)**	**Total length (km)**	**Inspected length (km)**
≤ 300 mm	2,252	1,209	426	23
300 - 600 mm	1,236	344	570	81
> 600 mm	636	93	290	52
Unknowm	501	3	-	-
Material	**Total length (km)**	**Inspected length (km)**	**Total length (km)**	**Inspected length (km)**
PVC	471	273	-	-
Reinforced concrete	2,167	1,212	-	-
Concrete	-	-	911	120
Other	267	115	361	39
Unknown	1,720	49	14	-

Table 3. Proportion of pipes in each structural state in different age ranges for Networks A and B

Network A

Age (years)	Proportion of pipes in state 0 (%)	Proportion of pipes in state 1 (%)	Proportion of pipes in state 2-3 (%)	Proportion of pipes in state 4-5 (%)
0 to 10	83	9	7	2
11 to 20	80	9	10	1
21 to 30	71	14	13	1
31 to 40	66	15	16	3
41 to 50	57	20	19	4
51 to 60	45	24	25	5

Network B

Age (years)	Proportion of pipes in state 0 (%)	Proportion of pipes in state 1 (%)	Proportion of pipes in state 2-3 (%)	Proportion of pipes in state 4-5 (%)
0 to 10	54	15	27	4
11 to 20	47	24	29	0
21 to 30	41	18	40	1
31 to 40	37	15	47	1
41 to 50	28	13	58	2
51 to 60	14	14	70	1
61 to 70	0	0	100	0

Table 4. Descriptions of analyzed cohorts for the two sewer networks

Factors	Network A		Network B	
	Cohort name	**Description**	**Cohort name**	**Description**
All factors	A1	All inspected pipes (*n* = 16,896)	B1	All inspected pipes (*n* = 2,380)
Diameter	A2	Inspected pipes with diameter < 300 mm (*n* = 13,209)	B2	Inspected pipes with diameter < 300 mm (*n* = 353)
	A3	Inspected pipes with diameter between 300 and 600 mm (*n* = 3,007)	B3	Inspected pipes with diameter between 300 and 600 mm (*n* = 1,336)
	A4	Inspected pipes with diameter ≥ 600 mm (*n* = 679)	B4	Inspected pipes with diameter ≥ 600 mm (*n* = 691)
Sewer type	A5	Inspected sanitary pipes (*n* = 13,544)	B5	Inspected combined pipes (*n* = 1,055)
	A6	Inspected combined and stormwater pipes (*n* = 3,352)	B6	Inspected stormwater pipes (*n* = 637)
			B7	Inspected sanitary pipes (*n* = 688)
Material	A7	Inspected PVC pipes (*n* = 2,882)	B8	Inspected concrete pipes (*n* = 1,808)
	A8	Inspected reinforced concrete pipes (*n* = 12,900)	B9	Inspected pipes of all other materials (asbestos cement, vitrified clay, PVC, and cast iron) (*n* = 572)
	A9	Inspected pipes of all other materials (asbestos cement, non-reinforced concrete, corrugated steel, brick, cast iron, pipe reinforced with glass fiber, polyethylene, steel) (*n* = 1,114)		
Length	A10	Inspected pipes with length < 60 m (*n* = 6,938)	B10	Inspected pipes with length < 60 m (*n* = 1,172)
	A11	Inspected pipes with length between 60 and 120 m (*n* = 9,221)	B11	Inspected pipes with length ≥ 60 m (*n* = 1,208)
	A12	Inspected pipes with length ≥ 120 m (*n* = 737)		

Table 5. Spearman correlation coefficients for pipe age and other covariates

Network A

	Age	**Diameter**	**Length**	**Sewer type X_1**	**Slope**	**Material X_1**	**Material X_2**
Age	1	0.24	0.06	0.37	0.01	0.14	-0.19
Diameter		1	0.02	0.49	0.07	0.30	-0.29
Length			1	-0.06	0.01	0.08	-0.09
Sewer type X_1				1	0.18	0.07	-0.06
Slope					1	-0.10	0.15
Material X_1						1	-0.81
Material X_2							1

Network B

	Age	**Diameter**	**Length**	**Material X_1**	**Sewer type X_1**	**Sewer type X_2**
Age	1	0.13	0.10	0.24	0.41	-0.27
Diameter		1	0.14	0.53	0.28	0.23
Length			1	-0.01	0.02	-0.02
Material X_1				1	0.47	0.05
Sewer type X_1					1	-0.54
Sewer type X_2						1

Table 6. Results of likelihood ratio test for the global Cox models

Network A																		
	Coefficients for the global model																	
	Diameter			**Length**			**Sewer type X_1**			**Slope**			**Material X_1**			**Material X_2**		
Estimated parameters	β_0	β_1	β_2	β_0	β_1	β_2	β_0	β_1	β_2	β_0	β_1	β_2	β_0	β_1	β_2	β_0	β_1	β_2
	0.00	0.14	0.00	0.07	0.00	0.03	0.47	0.00	0.00	0.08	0.00	0.41	0.27	0.00	0.00	0.01	0.51	0.88
λ	-32.90			-30.58			-333.71			-4.77			-303.74			-150.02		

Network B															
	Coefficients for the global model														
	Diameter			**Length**			**Material X_1**			**Sewer X_1**			**Sewer X_2**		
Estimated parameters	β_0	β_1	β_2	β_0	β_1	β_2	β_0	β_1	β_2	β_0	β_1	β_2	β_0	β_1	β_2
	0.00	0.00	0.02	0.00	0.07	0.01	0.06	0.00	0.00	0.17	0.13	0.04	0.13	0.65	0.01
λ	0.00			-2.32			-4.54			-26.69			-54.77		

Table 7. Results of likelihood ratio test for the simplified models

	Network A														
	Model 1														
	Diameter			**Length**			**Slope**								
Estimated parameters	β_0	β_1	β_2	β_0	β_1	β_2	β_0	β_1	β_2						
	0.08	0.03	0.03	0.49	0.00	0.00	0.03	0.01	0.56						
λ	-43.77			-38.53			-12.95								
	Model 2														
	Length			**Sewer type X_1**			**Slope**			**Material X_1**			**Material X_2**		
Estimated parameters	β_0	β_1	β_2	β_0	β_1	β_2	β_0	β_1	β_2	β_0	β_1	β_2	β_0	β_1	β_2
	0.06	0.01	0.02	0.48	0.00	0.02	0.04	0.00	0.28	0.22	0.00	0.08	0.02	0.38	0.87
λ	-22.67			-346.56			-0.08			-207.43			-143.32		

	Network B								
	Model 1								
	Diameter			**Length**					
Estimated parameters	β_0	β_1	β_2	β_0	β_1	β_2			
	0.02	0.02	0.02	0.00	0.06	0.00			
λ	-0.87			-1.93					
	Model 2								
	Length			**Sewer type X_1**			**Sewer type X_2**		
Estimated parameters	β_0	β_1	β_2	β_0	β_1	β_2	β_0	β_1	β_2
	0.00	0.08	0.00	0.20	0.14	0.00	0.15	0.65	0.00
λ	-2.60			-34.71			-58.03		
	Model 3								
	Length			**Material X_1**					
Estimated parameters	β_0	β_1	β_2	β_0	β_1	β_2			
	0.00	0.07	0.00	0.19	0.00	0.00			
λ	-2.54			-40.09					

Table 8. Results of Kruskal-Wallis test for comparison between very good and poor structural condition states

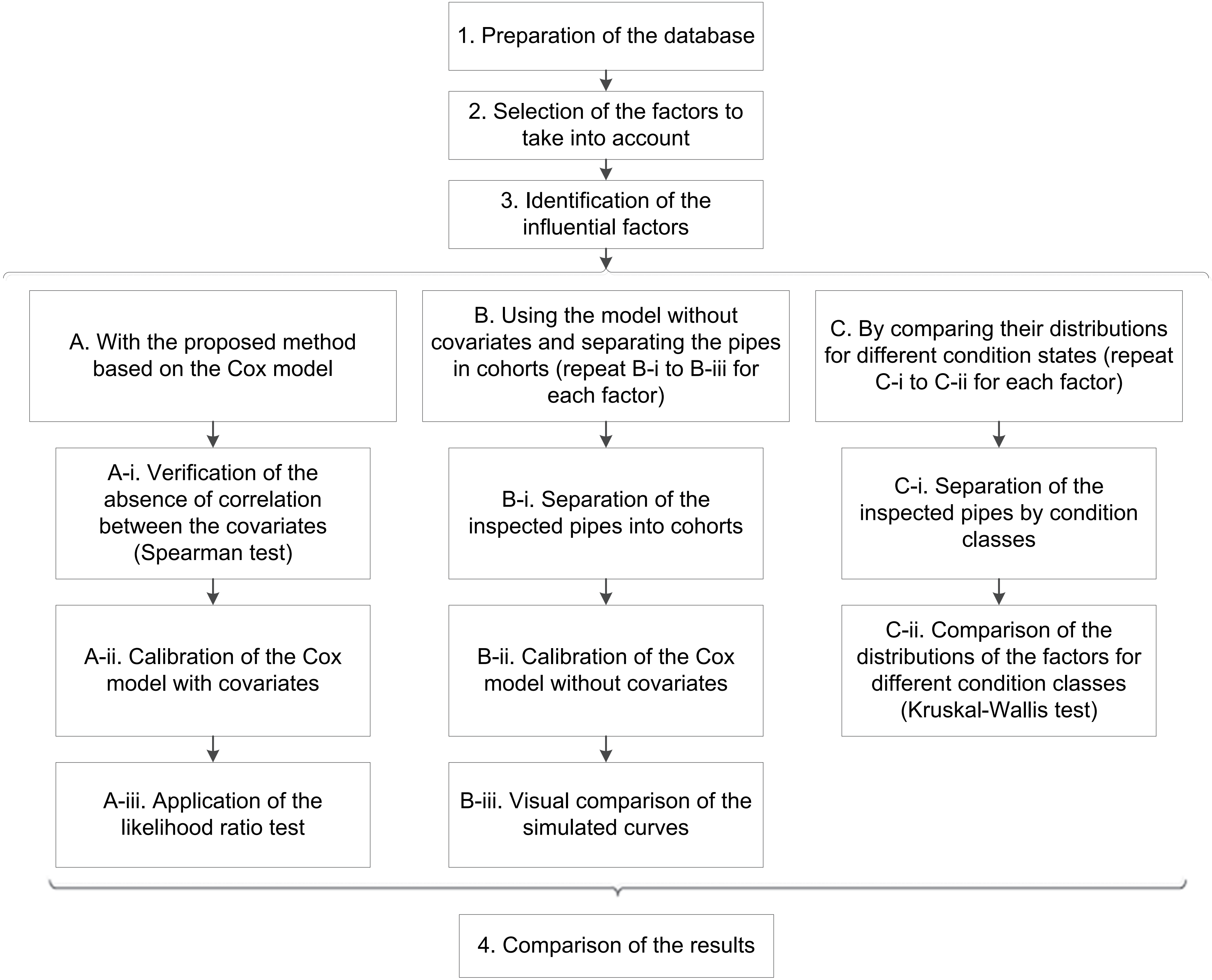
Network	p_value for the compared factors			
	Age	Diameter	Length	Slope
A	1.241E-38	0.215	0.094	0.013
B	0.007	0.965	0.931	

Figure captions

Figure 1. Illustration of methodology

Figure 2. Proportion of pipes assigned to structural states 0 (very good), 1 (good), 2-3 (fair), and 4-5 (poor) for Network A

Figure 3. Probability that pipes of Network A will be in state 4-5 (poor structural state), classified by (a) diameter, (b) sewer type, (c) material and (d) length


1. Preparation of the database
2. Selection of the factors to take into account
3. Identification of the influential factors
A. With the proposed method based on the Cox model
B. Using the model without covariates and separating the pipes in cohorts (repeat B-i to B-iii for each factor)
C. By comparing their distributions for different condition states (repeat C-i to C-ii for each factor)
A-i. Verification of the absence of correlation between the covariates (Spearman test)
B-i. Separation of the inspected pipes into cohorts
C-i. Separation of the inspected pipes by condition classes
A-ii. Calibration of the Cox model with covariates
B-ii. Calibration of the Cox model without covariates
C-ii. Comparison of the distributions of the factors for different condition classes (Kruskal-Wallis test)
A-iii. Application of the likelihood ratio test
B-iii. Visual comparison of the simulated curves
4. Comparison of the results

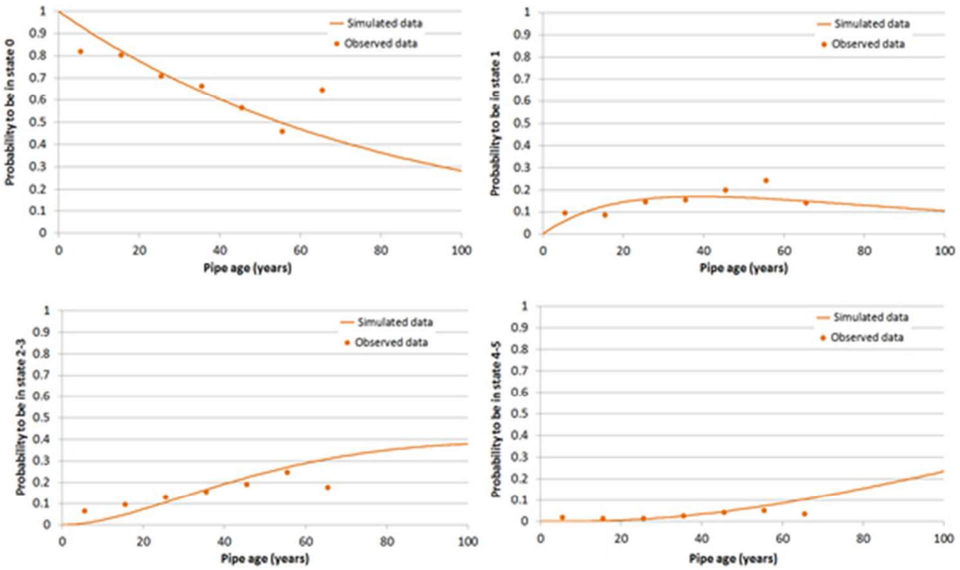


Figure 2. Proportion of pipes assigned to structural states 0 (very good), 1 (good), 2-3 (fair), and 4-5 (poor) for Network A.

50x30mm (300 x 300 DPI)

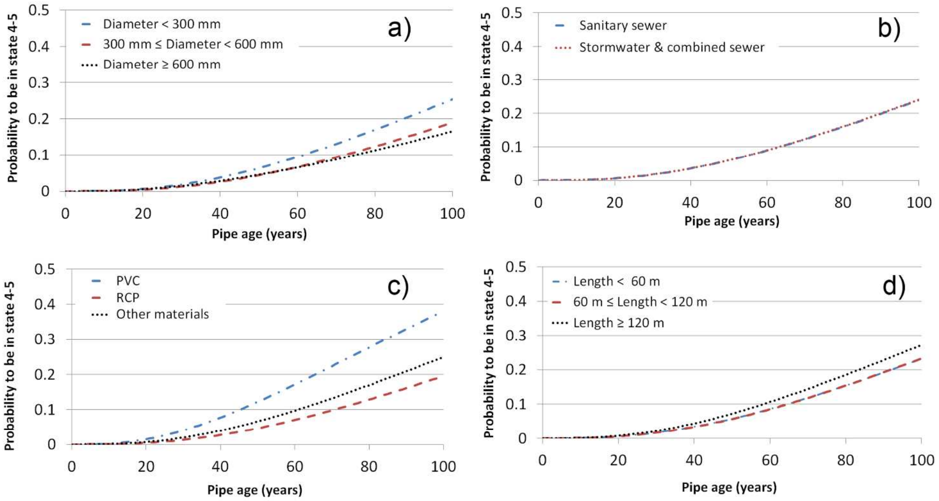


Figure 3. Probability that pipes of Network A will be in state 4-5 (poor structural state), classified by (a) diameter, (b) sewer type, (c) material and (d) length

186x101mm (300 x 300 DPI)