

Modeling the evolution of the structural state of water pipe networks with brief recorded pipe break histories: Methodology and application

Alain Mailhot, Geneviève Pelletier, Jean-François Noël,
and Jean-Pierre Villeneuve

Institut National de la Recherche Scientifique, Sainte-Foy, Québec, Canada

Abstract. Water managers are concerned about the poor structural state of their water pipe networks and are confronted with the lack of tools to analyze their data in order to evaluate both the present and future state of their networks. Researchers have developed statistical models with the annual number of pipe breaks as an indicator of the structural state of their networks. Unfortunately, most statistical modeling strategies do not perform well when used in municipalities with short recorded pipe break histories. Pipe breaks must have been recorded since the installation of the first pipes, and this is rarely the case in most municipalities. This paper presents a methodology to estimate calibration parameters of statistical models in municipalities with short recorded pipe break histories. The application of this methodology to a municipality with a century-old water pipe network and a recorded pipe break history of 21 years is also presented.

1. Introduction

The last decades have seen an important number of studies related to the structural state and aging of urban infrastructures in North America. Roads, sidewalks, sewers, water facilities, bridges, and other infrastructures are reported to be in poor condition and deteriorating rapidly. In 1985 the Federation of Canadian Municipalities (FCM) published a report on the state of municipal infrastructures based on a survey in many Canadian municipalities. The cost of updating these infrastructures was estimated to be around 18 billion dollars [Desbiens, 1997]. In 1995 a new survey undertaken by McGill University in conjunction with FCM concluded that the cost had risen to 44 billion dollars for all of Canada [Siddiqui and Mirza, 1996].

“Out of sight, out of mind” aggravates the problem for underground water infrastructure. In the United States, two milestones were at the origin of the realization that major problems existed with older water distribution networks. The first was the passage of the Safe Drinking Water Act of 1974, which mandated that the U.S. Environmental Protection Agency be concerned with the supply of drinking water to consumers. The second was the publication of the report by Choate and Walter [1981], which attempted to draw attention to the “infrastructure crisis.” With pipe breaks and sewer overflows occurring more frequently, water managers are asking questions about the state of their water and sewer networks, concerned that these infrastructure systems did not receive adequate preventive maintenance.

Although it is quite clear that major interventions are needed on most networks, budget constraints make it essential to better plan and optimize these interventions. The management of water networks, in terms of optimizing the cost of intervention by better planning repair, rehabilitation, and re-

placement of pipes, has become very complex, and researchers have developed methodologies to help water managers in that task [e.g., Goulter *et al.*, 1993; Kleiner *et al.*, 1998a, 1998b; Villeneuve *et al.*, 1998]. Before cost optimization can be achieved, one must first develop modeling strategies to evaluate the present state and predict the evolution of the structural state of networks using available (or easily obtainable) data. The performance of a modeling strategy will depend on the identification of good indicators of the structural state. The resulting pipe break model can serve both as a diagnostic tool (e.g., identification of pipes with high breakage rates) and an optimization tool (e.g., best replacement strategies), but also, when coupled with an economic assessment model, it becomes a powerful tool for decision making for water managers [O’Day *et al.*, 1986].

Survival analysis is a statistical method used for analysis of time-to-failure data. The main advantage of survival analysis is that it takes into account information from pipes that have failed one or more times during the observation period as well as pipes that have not failed during the same period. The main disadvantage is that this type of analysis requires detailed information on the characteristics of pipe segments in the network and a long history of pipe breaks. Time to failure between successive breaks was observed to be very different than time to failure from pipe installation to first break [e.g., Clark *et al.*, 1982, 1988; Andreou *et al.*, 1987]. The survival time between different break orders needs to be described by different statistical distributions since they exhibit different breakage behaviors. The most commonly used statistical distributions for time-to-failure data associated with pipe breaks are the exponential and the Weibull distributions [e.g., Andreou *et al.*, 1987; Eisenbeis, 1994]. In order to use survival analysis, time to failure must be known for the entire sequence of pipe breaks, that is, from pipe installation to first break, from first break to second, etc. This is rarely the case, especially for municipalities with older water pipe networks, since their recorded pipe break histories rarely cover the entire history of their network.

Copyright 2000 by the American Geophysical Union.

Paper number 2000WR900185.
0043-1397/00/2000WR900185\$09.00

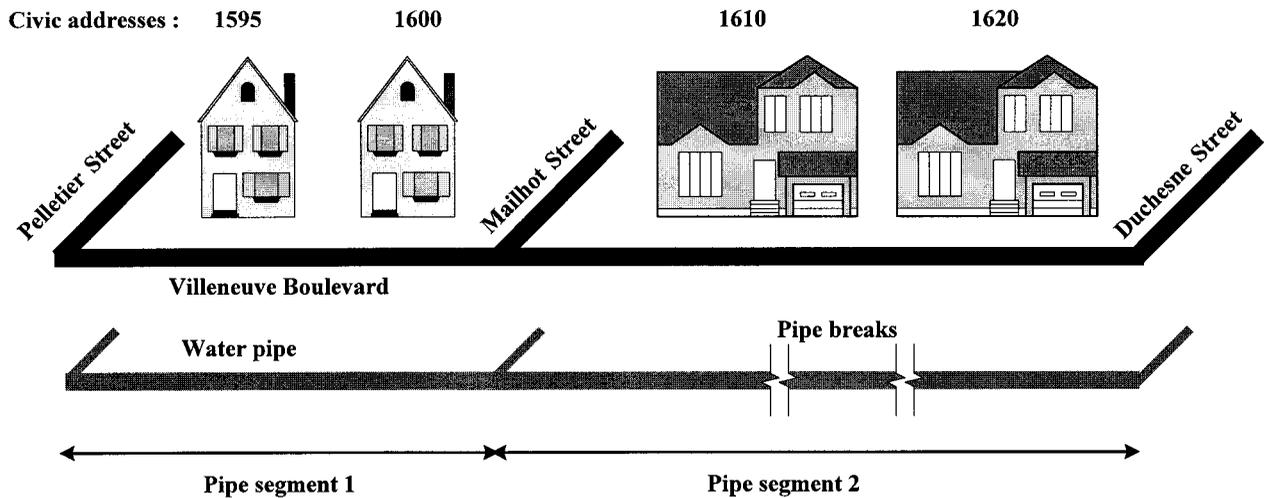


Figure 1. Example of pipe segments.

Therefore time to failure is only known for breaks that happen within an observation period that usually only covers recent years. Knowledge of the order of breaks (first, second, third, etc.) is only available if the pipe was installed during the time covered by the observation period. This considerably limits the applicability of standard survival analysis in modeling the structural state of water pipe networks.

Many researchers have used survival analysis successfully to evaluate past and present breakage rates and to predict breakage trends in the near future. For example, *Karaa and Marks* [1990] used Cox's proportional hazard model to analyze the city of New Haven's data divided into six installation periods. In Canada, pioneer work has been done by I. C. Goulter's team on the city of Winnipeg's water pipe network's data. *Kettler and Goulter* [1985] analyzed the breakage rates of grey cast iron and asbestos-cement pipes in relation to pipe diameter and pipe age. *Goulter and Kazemi* [1988] observed that 22% of pipe breaks occurred within 1 m from one another, and 46% occurred within 20 m. *Goulter et al.* [1993] developed a methodology to quantify the temporal and spatial groupings of these pipe breaks. *Jacobs and Karney* [1994] developed two indicators of the reliability of a water pipe network: first, the probability of having a day without a pipe break and, second, the probability of having an independent break (not in the vicinity of a recent break). In Germany, *Herz* [1996] developed his own statistical distribution and used a cohort survival model to estimate breakage and renewal rates on 10 European networks. In France, *Eisenbeis* [1994] used Cox's proportional hazard model to analyze breakages rates on two urban networks with recorded pipe break histories of 40 and 54 years.

To our knowledge the only two studies that dealt with brief recorded pipe break histories are the ones by *Kleiner and Rajani* [1999] and *Eisenbeis* [1994]. These researchers obtained parameter values for different categories of pipes from a general model developed with recorded pipe break histories covering many decades. The general model used by *Kleiner and Rajani* [1999] is based on an exponential relationship developed by *Shamir and Howard* [1979], while, as mentioned above, *Eisenbeis* [1994] used Cox's proportional hazard model.

In this paper, we present a methodology which extends the applicability of survival analysis to networks with brief recorded pipe break histories. The proposed methodology is

different from that of *Eisenbeis* [1994] and *Kleiner and Rajani* [1999], since it is formally exact and it explicitly takes into account the fact that we do not know how many and when pipe breaks occurred during the nonrecorded period. It is then possible to include the "older" part of the network in the analysis and to compare the evolution of its structural state with the "younger" part. Comparison of different pipe break models is also possible.

First, we present basic concepts and some terminology, and then we elaborate on mathematical developments and calculations of the likelihood function. Application of this methodology to the municipality of Chicoutimi (Province of Quebec, Canada) is then presented, and we finally show how two pipe break models can be compared in order to estimate which one is the most significant statistically.

2. Modeling Water Pipe Network Structural State: Basic Elements

2.1. Discretization of the Network Into Pipe Segments

The water pipe network must be discretized into pipe segments. A pipe segment is a series of pipes with relatively homogeneous characteristics, in terms of what affects the rate of pipe degradation, that usually runs from one street corner to the next (see Figure 1). Previous studies have highlighted the fact that pipe characteristics, such as pipe diameter and type of material, have an impact on the breakage rate [*Clark et al.*, 1982; *Andreou et al.*, 1987]. For example, *Kettler and Goulter* [1985] have observed that pipe failure rates are much higher for smaller diameters, probably because of the thinner pipe walls. A pipe segment is also the smallest element where a pipe break can be located.

2.2. Modeling Strategy

Once the pipe network has been discretized into pipe segments, one must specify the type of probability distribution that will be used to describe each break order. In the present study, two types of probability distribution were considered for modeling time between breaks: a two-parameter Weibull distribution and a one-parameter exponential distribution (Table 1). These two distributions were applied to different orders of breaks in various combinations as depicted in Table 2. The

Table 1. Equations for the Different Functions of the Exponential and Weibull Distributions

Distribution	Probability Density Function	Survival Function	Hazard Function
Exponential	$\kappa \exp(-\kappa t)$	$\exp(-\kappa t)$	κ
Weibull	$\kappa p (\kappa t)^{p-1} \exp[-(\kappa t)^p]$	$\exp[-(\kappa t)^p]$	$\kappa p (\kappa t)^{p-1}$

pipe break model defined by the combination of these two distributions was then used to evaluate the evolution of breakage probability with time for each pipe segment. Equations for the survival function $F(t)$, the hazard function $\lambda(t)$, and probability density function $f(t)$ for these two distributions are presented in Table 1 (for a definition of these functions see *Kalbfleisch and Prentice* [1980] or *Cox and Oakes* [1994]). It is important to note that the developed methodology is applicable no matter the types of distributions considered to describe the different break orders. Table 2 presents the different pipe break models considered for our case study. The number of calibration parameters depends on the distributions used and on the number of break orders considered.

Survival analysis is then used to build the likelihood function and to estimate model parameters. The main advantage of survival analysis in modeling pipe breaks is that it takes into account the occurrence of one or many breaks, and the absence of breaks on pipe segments, when estimating the parameters of the distributions. When a pipe segment has yet to fail at the time of analysis, its time to failure is said to be right-censored [Miller, 1981]. The likelihood function at the time of analysis t is built considering both the break probability for pipe segments that have failed and the survival probability for pipe segments that have not [Miller, 1981; Kalbfleisch and Prentice, 1980].

2.3. Specific Problems Associated With Brief Recorded Pipe Break Histories

In order to apply standard survival analysis, it is necessary to have the entire break record since the installation of the first pipes. This is not the case for municipalities with brief recorded pipe break histories. Figure 2 illustrates this problem for a pipe segment laid in 1950. If pipe breaks have been recorded at least since 1950, then we know that the pipe break in 1985 is the first break and the one in 1990 is the second. However, if pipe breaks began to be recorded in 1976, it is not possible to know if the pipe break observed in 1985 is actually

the first break. One, many, or no breaks could have occurred during the 1950–1976 period. Referring to Figure 2, it should be noted that unknown information “left” of 1976 cannot be associated to the so-called “left censoring” [Miller, 1981]. Left censoring could only be used if the number of breaks that occurred before 1976 was known, without knowing when these breaks did occur. Knowing the exact order of breaks is key to successfully using survival analysis for modeling pipe breaks, and since this information is not available in municipalities with brief recorded pipe break histories, a new methodology was developed to use at best the information at hand during the observation period.

3. Description of the Proposed Methodology

Our goal is to develop a likelihood function that, once maximized, will give us the values of the calibration parameters of the distributions associated with the different break orders. The time step used is 1 year. Figure 3 illustrates the case of a pipe segment i that failed β times during the observation period, at times t_1, t_2, \dots , and t_β , and failed α times during the nonrecorded history, at times t'_1, t'_2, \dots , and t'_α . Time $t = 0$ is the time of installation of pipe segment i , T_b is the time when recording began, and T_a is the time of analysis (present time). The n th break order (there are $\beta + \alpha$ break orders during a pipe segment’s life) is associated with a statistical distribution and its corresponding probability density function $f_n(t)$, survival function $F_n(t)$, and hazard function $\lambda_n(t)$.

Probability $P(\alpha, \beta; \{t_j\})$ for a pipe segment i is defined as the probability of occurrence of β breaks at times t_1, t_2, \dots , and t_β during the observation period, considering that the pipe segment already failed α times during the nonrecorded history at unknown times t'_1, t'_2, \dots , and t'_α . In terms of the different statistical functions describing the break orders, this probability is defined by

Table 2. List and Characteristics of Models Considered for the Municipality of Chicoutimi

Models	Break Order	Distribution	Number of Parameters
Weibull-Exponential (W-E)	first break	Weibull (κ_1, p_1)	3
Weibull-Exponential-Exponential (W-E-E)	second break and up	exponential (κ_2)	4
	first break	Weibull (κ_1, p_1)	
	second break	exponential (κ_2)	
Weibull-Weibull-Exponential (W-W-E)	third break and up	exponential (κ_3)	5
	first break	Weibull (κ_1, p_1)	
	second break	Weibull (κ_2, p_2)	
Weibull-Weibull-Exponential Exponential (W-W-E-E)	third break and up	exponential (κ_3)	6
	first break	Weibull (κ_1, p_1)	
	second break	Weibull (κ_2, p_2)	
	third break	exponential (κ_3)	
	fourth break and up	exponential (κ_4)	

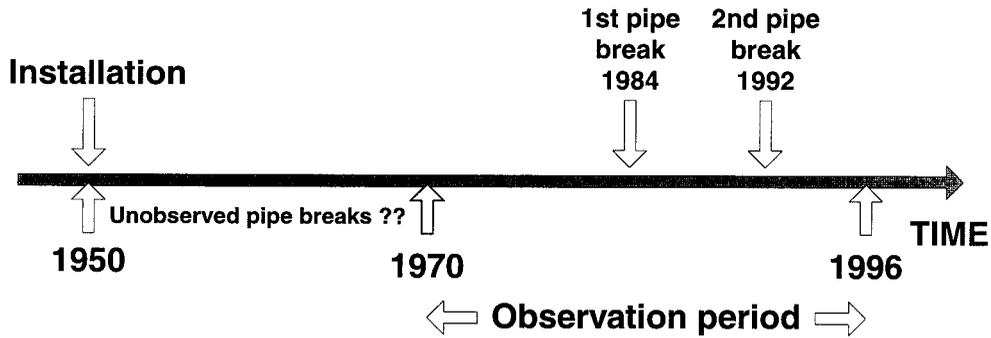


Figure 2. Example of a brief recorded pipe break history for a given pipe segment.

$$P(\alpha, \beta; \{t_j\}) = \int_0^{T_b} dt'_1 f_1(t'_1) \int_{t'_1}^{T_b} dt'_2 f_2(t'_2 - t'_1) \dots \\
 \cdot \int_{t'_{\alpha-1}}^{T_b} dt'_\alpha f_\alpha(t'_\alpha - t'_{\alpha-1}) f_{\alpha+1}(t_1 - t'_\alpha) \dots \\
 \cdot f_{\alpha+\beta}(t_\beta - t_{\beta-1}) F_{\alpha+\beta+1}(T_a - t_\beta) dt_1 \dots dt_\beta. \quad (1)$$

Breaks that occur during the nonrecorded history do so in a sequence of arbitrary times $\{t'_1, t'_2, \dots, t'_\alpha\}$. Since these times are unknown, the probability of occurrence of the first nonrecorded break is equal to the integral of the probability density function associated with this break order, between the time of installation of the pipe segment ($t = 0$) and the time T_b when recording began. Each following integral in the equation represents the probability of failure for each higher break order during the nonrecorded history. Obviously, the sequence of breaks must respect the constraint $0 \leq t'_1 \leq t'_2 \leq t'_3 \dots \leq t'_\alpha$.

The probability of occurrence of the first recorded break is equal to the probability density function corresponding to break order $\alpha + 1$. Each following probability density function in the equation corresponds to the probability of failure of each higher break order during the observation period (up to the last recorded break order $\alpha + \beta$). The last part of the equation is the survival function which is used to estimate the probability of not having a $(\alpha + \beta + 1)$ break between the time of occurrence of the last recorded break and the time of analysis T_a . This expression, though quite difficult to estimate in its general form, becomes much simpler for particular cases, as the one presented hereafter.

Once probability $P(\alpha, \beta; \{t_j\})$ is estimated for each pipe segment (for clarity, we omit for now pipe segment indices), we must calculate probability $P'(\beta; \{t_j\})$, which corresponds to the probability of observing the break sequence t_1, t_2, \dots, t_β during the observation period, no matter how many and when

breaks occurred during the nonrecorded period. In terms of $P(\alpha, \beta; \{t_j\})$ this probability is defined as

$$P'(\beta; \{t_j\}) = \sum_{\alpha=0}^{\infty} P(\alpha, \beta; \{t_j\}). \quad (2)$$

For each pipe segment i we know the number of breaks during the observation period β_i , times when these breaks occurred $\{t_j\}_i$, time when recording began T_{bi} , and time of analysis T_{ai} . Function $P'_i(\beta_i; \{t_j\}_i)$ can therefore be calculated. The likelihood function $L(\{\gamma_k\})$, which depends on the k calibration parameters γ_k is obtained by calculating the product of this probability over every pipe segment in the water pipe network:

$$L(\{\gamma_k\}) = \prod_i P'_i(\beta_i; \{t_j\}_i). \quad (3)$$

Values for the calibration parameters are obtained by maximizing the likelihood function. The number of parameters and the complexity of the likelihood function depend on the distributions used to describe each break order and on the number of break orders considered. Here is an example of calculations for the model consisting of a Weibull distribution for the first break and an exponential distribution for all subsequent breaks.

Time to failure for the first break is modeled by a Weibull distribution with two parameters, κ_1 and p_1 , and time to failure for all subsequent breaks is modeled by an exponential distribution with one parameter, κ_2 . Therefore, in this example, three parameters must be calibrated by maximizing the likelihood function. Probability density functions for these distributions are presented in Table 1.

First, we must calculate probability, $P(\alpha, 0)$, for each pipe segment. This corresponds to the probability for a given pipe segment to experience no breaks during the observation period and α breaks during the nonrecorded history. From the general

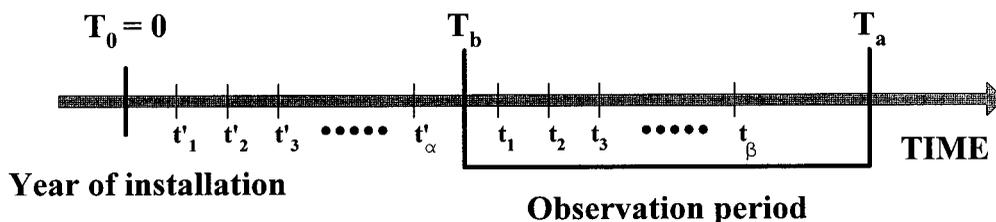


Figure 3. Definition of the variables used to estimate the likelihood function for a given pipe segment.

expression presented in (1) and considering the model in this example, the probability of observing no breaks in the observation period and no breaks during the nonrecorded history, $P(0, 0)$, is

$$P(0, 0) = F_1(T_a) = \exp[-(\kappa_1 T_a)^{p_1}]. \quad (4)$$

For $P(1, 0)$ we have

$$\begin{aligned} P(1, 0) &= \int_0^{T_b} dt'_1 f_1(t'_1) F(T_a - t'_1) \\ &= p_1 \kappa_1^{p_1} \exp(-\kappa_2 T_a) \\ &\quad \cdot \int_0^{T_b} dt' (t')^{p_1-1} \exp[-(\kappa_1 t')^{p_1}] \exp(\kappa_2 t'). \end{aligned} \quad (5)$$

This expression can be generalized to define the probability $P(\alpha, 0)$, when $\alpha \geq 1$, and we find

$$\begin{aligned} P(\alpha, 0) &= p_1 \kappa_1^{p_1} \kappa_2^{\alpha-1} \exp(-\kappa_2 T_a) \int_0^{T_b} dt' (t')^{p_1-1} \\ &\quad \cdot \exp[-(\kappa_1 t')^{p_1}] \exp(\kappa_2 t') \left[\frac{(T_b - t')^{\alpha-1}}{(\alpha-1)!} \right]. \end{aligned} \quad (6)$$

It can be shown that probability $P(\alpha, \beta; \{t_j\})$ can be expressed as a function of $P(\alpha, 0)$, when $\alpha \geq 1$, and we find

$$P(\alpha, \beta; \{t_j\}) = \kappa_2^\beta P(\alpha, 0). \quad (7)$$

According to (7), when a pipe break has occurred during the nonrecorded period, the probability to observe a pipe break during the observation period does not depend on the time when the nonrecorded pipe break occurred. This is related to the nature of the exponential distribution associated with subsequent pipe breaks. It can also be shown that probabilities of the type $P(0, \beta; \{t_j\})$ can be expressed in the general form for $\beta \geq 1$:

$$\begin{aligned} P(0, \beta; \{t_j\}) &= p_1 \kappa_1^{p_1} \kappa_2^{\beta-1} t_1^{p_1-1} \exp[-\kappa_2(T_a - t_1)] \\ &\quad \cdot \exp[-(\kappa_1 t_1)^{p_1}]. \end{aligned} \quad (8)$$

It should be noted that the only t_j that appears in the right-hand side of (8) is the time of occurrence of the first recorded break t_1 . Again, this is related to the nature of the exponential distribution chosen to model time to failure of higher order of breaks.

The expression for probability $P'(\beta; \{t_j\})$ which is, as a reminder, the probability of observing β breaks during the observation period at times $\{t_j\}$, no matter how many and when breaks have occurred during the nonrecorded history, is then

$$P'(\beta; \{t_j\}) = P(0, \beta; \{t_j\}) + \sum_{\alpha=1}^{\infty} P(\alpha, \beta; \{t_j\}). \quad (9)$$

Using (7), this expression can be written in the following form:

$$P'(\beta; \{t_j\}) = P(0, \beta; \{t_j\}) + \kappa_2^\beta \sum_{\alpha=1}^{\infty} P(\alpha, 0). \quad (10)$$

The first term is given by (8). Considering (6) for $P(\alpha, 0)$, we can show that this sum reduces to

$$\sum_{\alpha=1}^{\infty} P(\alpha, 0) = \exp[\kappa_2(T_b - T_a)] \{1 - \exp[-(\kappa_1 T_b)^{p_1}]\}. \quad (11)$$

In the case of $T_b = 0$ (recording begins with the installation of the pipe segment; therefore the entire break history is known), the break probability of (11) reduces to zero, as expected. Summing both terms of (10), we obtain in the case of $\beta \geq 1$

$$\begin{aligned} P'(\beta; t_1) &= \kappa_2^{\beta-1} \exp(-\kappa_2 T_a) \\ &\quad \cdot \{p_1 \kappa_1^{p_1} t_1^{p_1-1} \exp[-(\kappa_1 t_1)^{p_1}] \exp(\kappa_2 t_1) \\ &\quad + \kappa_2 \exp(\kappa_2 T_b) [1 - \exp(-(\kappa_1 T_b)^{p_1})]\}. \end{aligned} \quad (12)$$

It should be noted that in this example, this probability only depends on the number of breaks during the observation period β and on the time of occurrence of the first recorded break t_1 . The probability, $P'(0)$, of no breaks occurring during the observation period can be written as

$$\begin{aligned} P'(0) &= \exp[-(\kappa_1 T_a)^{p_1}] \\ &\quad + \exp[\kappa_2(T_b - T_a)] \{1 - \exp[-(\kappa_1 T_b)^{p_1}]\}. \end{aligned} \quad (13)$$

The likelihood function can therefore be written as

$$L(p_1, \kappa_1, \kappa_2) = \prod_{i|\beta=0} P'_i(0) \prod_{i|\beta \geq 1} P'_i(\beta_i; t_{i1}). \quad (14)$$

The subscript i refers to pipe segments. The first term accounts for pipe segments for which no breaks have been recorded ($\beta = 0$), and the second term accounts for pipe segments that have failed once or more ($\beta \geq 1$). Probabilities for the first term are calculated using (13), and probabilities for the second term are calculated with (12). It should be noted that times T_b and T_a appearing in both (12) and (13) vary for each pipe segment. Developing (14) and taking the logarithm of $L(p_1, \kappa_1, \kappa_2)$, we finally obtain the following expression:

$$\begin{aligned} \ln [L(p_1, \kappa_1, \kappa_2)] &= \sum_{i|\beta=0} \ln \{ \exp[-(\kappa_1 T_{ai})^{p_1}] \\ &\quad + \exp[\kappa_2(T_{bi} - T_{ai})] [1 - \exp(-(\kappa_1 T_{bi})^{p_1})] \} \\ &\quad + n' \ln \kappa_2 - \kappa_2 \sum_{i|\beta \geq 1} T_{ai} \\ &\quad + \sum_{i|\beta \geq 1} \ln \{ p_1 \kappa_1^{p_1} t_{i1}^{p_1-1} \exp[-(\kappa_1 t_{i1})^{p_1}] \exp(\kappa_2 t_{i1}) \\ &\quad + \kappa_2 \exp(\kappa_2 T_{bi}) [1 - \exp(-(\kappa_1 T_{bi})^{p_1})] \}. \end{aligned} \quad (15)$$

In this expression, $n' = (n_b + n_0 - n_t)$, where n_b is the total number of recorded breaks for the entire water pipe network, n_0 is the number of pipe segments that have not failed during the observation period, and n_t is the total number of pipe segments. Optimization of the likelihood function will give the values of calibration parameters, p_1 , κ_1 , and κ_2 , that once injected in the model, will best approximate pipe break behavior. In this example, and for all the models considered, no analytical form could be derived to obtain the optimal values of calibration parameters. Therefore it was necessary to use numerical methods of optimization to find the optimal set of calibration parameters that maximized the likelihood functions.

Even if (15) seems complex, it is easily applied to a given network. The likelihood function only depends on model parameters, p_1 , κ_1 , and κ_2 , and for each pipe segment on the

Table 3. Example of Input File Needed to Estimate the Likelihood Function (Equation (15))

Pipe Number	Year of Installation	Number of Breaks Observed	Year of First Observed Break	Time of Beginning of Observation Period (T_b) ^a	Time of Analysis (T_a) ^a	Time of First Observed Break (t_1)
1	1950	1	1971	20	49	21
2	1972	0	...	0	18	...

^aWe suppose, in this example, that the observation period begins in 1970 and that the time analysis is year 1999 for all pipe segments.

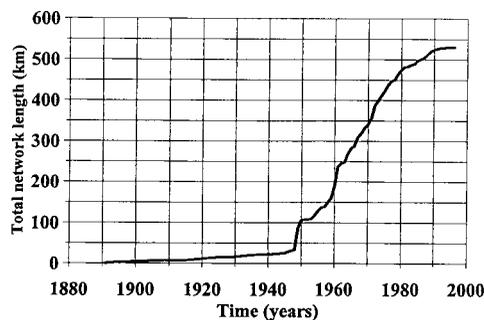
times when the observation period began T_{bi} , on the times of analysis T_{ai} , on the number of breaks observed β_i , and on the times of the first recorded break in the observation period t_{1i} . Table 3 shows how the input values in (15) can be easily estimated and formatted from basic information on pipe segments. It is important to note that T_{bi} is equal to zero for pipe segments installed during the observation period. Pipe segments that have been replaced during the observation period must also be included in the input file, as one entry for the replaced pipe segment and another for the new pipe segment. In that case, time of analysis for the replaced pipe and time of installation of the new one are given by the time of replacement.

Using a simple pipe break model, this example illustrates the different steps used in the calculation of the likelihood function. Section 4 presents the application of the proposed methodology to the municipality of Chicoutimi's water pipe network. Many other pipe break models were considered for Chicoutimi, but the calculations for each model will not be presented here since the methodology is similar. It is important to note, however, that the expressions and calculations, although straightforward, become more and more demanding as the pipe break models considered become more complex [Pelletier, 2000].

4. Application to the Municipality of Chicoutimi

4.1. Water Pipe Network Characteristics and Recorded Pipe Break History

Chicoutimi is a municipality of around 65,000 inhabitants, located on the shores of the Saguenay River, 150 km north of Quebec City. In 1996 the total length of Chicoutimi's water pipe network was approximately 352 km. The water pipe network was discretized into 2096 pipe segments, and information on six characteristics of each pipe segment was saved in a database, namely, pipe diameter, pipe material, year of instal-

**Figure 4.** Evolution of the total network length for the municipality of Chicoutimi.

lation, pipe length, soil type, and land use above pipe. More than 89% of pipe segments are under 300 m in length, with an average length of around 150 m. Figure 4 shows the evolution of total pipe length since the first pipes were laid in 1891. The municipality experienced a major growth in 1949, and urban development remained rapid until recent years.

Systematic recording of pipe breaks began in 1976. The observation period covers 21 years (1976–1996). There are some data missing for part of the year in 1978. Figure 5 presents the evolution of the number of pipe breaks during the observation period. There is a net increase with time. The total number of breaks archived is 2289, but only 1719 breaks could be associated with a single pipe segment. These breaks all occurred on pipes and joints. Breaks on all other parts of the water pipe network, like fire hydrants and valves, are excluded from the analysis. In 1996, 163 pipe breaks occurred on the water pipe network. This corresponds globally to a break rate of 46 breaks per 100 km, which is considered high and a symptom of a degraded network according to a study done by McDonald *et al.* [1994] of the National Research Council of Canada. The average break rate in municipalities in Ontario (Canada) was estimated at 25 breaks per 100 km [Canada Mortgage and Housing Corporation, 1992] and was estimated at 13 breaks per 100 km in a sample of American cities [American Water Works Association, 1994].

4.2. Models Considered and Parameter Estimation Using Likelihood Functions

Table 2 presents the different models considered for the municipality of Chicoutimi. Four models were examined, with an increasing number of parameters. For each of these models the likelihood function was derived. A numerical method of optimization was used, namely, Powell's algorithm [Press *et al.*, 1988], to estimate the values of the calibration parameters. Validation of the likelihood function formulations and of the

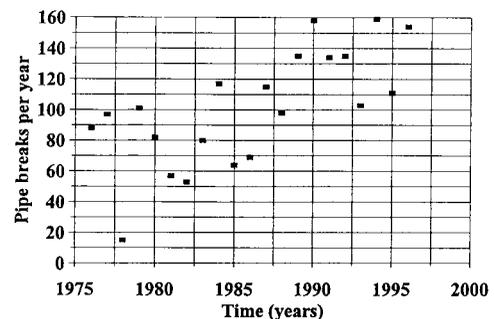
**Figure 5.** Annual number of recorded breaks from 1976 to 1996 in the municipality of Chicoutimi.

Table 4. Results of the W-E Model for Different Pipe Segment Installation Periods

Installation Periods	p_1	κ_1	κ_2
1976–1996	1.157	0.017	0.168
1970–1996	1.206	0.018	0.148
1965–1996	1.394	0.024	0.182
1960–1996	1.479	0.025	0.205
1949–1996	1.241	0.018	0.161
1891–1996	1.053	0.015	0.147

method used for maximization were done by considering synthetic data generated with the proposed models. A sequence of pipe breaks was randomly generated for every pipe of a network. An “unrecorded” period was then assumed, and maximization of the likelihood function was performed using information from the recorded period only. Resulting parameter values were compared to the actual ones used to generate the pipe break records, and they were found to be in close agreement.

After this verification the method was applied to Chicoutimi’s data. At first, because of the complexity of the functions to optimize and the number of parameters (up to six), optimization was performed with only the pipe segments installed from the beginning of the recording period (1976–1996). For these pipe segments, “standard” survival analysis can be used, and therefore it was possible to verify that the likelihood functions of the different models considered were formulated appropriately.

4.3. Results

Tables 4 to 7 present the different values of the calibration parameters for all the models considered. Analyses were performed on pipe segments installed during these six periods: 1976–1996, 1970–1996, 1965–1996, 1960–1996, 1949–1996, and 1891–1996. The total number of pipe segments considered (including replaced and new pipes) in each case is 679, 1033, 1243, 1510, 2184, and 2404, respectively.

Tables 4–7 show that parameter values vary noticeably for different installation periods for all the models considered. Hence taking into account only pipe segments laid during the years covered by the observation period when estimating parameter values will lead to significantly different results than when all pipe segments are used for the analysis. This suggests, as expected, that the installation period or another time-related factor (e.g., pipe material) has a strong impact on the probability of pipe failure.

Figure 6 presents the hazard function of the Weibull distribution associated with the first break order for some installation periods. Adding pipe segments installed in the 1960s (in-

Table 5. Results of the W-E-E Model for Different Installation Periods

Installation Periods	p_1	κ_1	κ_2	κ_3
1976–1996	1.155	0.017	0.076	0.331
1970–1996	1.193	0.018	0.078	0.246
1965–1996	1.369	0.024	0.082	0.291
1960–1996	1.461	0.025	0.075	0.342
1949–1996	1.249	0.020	0.048	0.284
1891–1996	1.154	0.018	0.038	0.261

Table 6. Results of the W-W-E Model for Different Installation Periods

Installation Periods	p_1	p_2	κ_1	κ_2	κ_3
1976–1996	1.155	0.924	0.017	0.070	0.347
1970–1996	1.200	0.999	0.018	0.077	0.255
1965–1996	1.380	1.013	0.024	0.080	0.301
1960–1996	1.469	1.043	0.025	0.074	0.352
1949–1996	1.318	0.921	0.020	0.046	0.283
1891–1996	1.231	0.779	0.019	0.030	0.260

stallation periods 1960–1996 and 1965–1996) has a strong influence on the shape of the hazard function: The risk of failure is much higher. It is important to note that the number of pipes installed during the 1960s (422 pipe segments) is comparable to the number of pipes installed in the 1950s (329 pipe segments) and in the 1970s (454 pipe segments). Since it is not only the weight of the number of pipes installed in the 1960s that can justify the higher failure risk, one must conclude that there is some unknown risk factor associated with this period. It could be related to some pipe characteristics or to some other factors such as the installation techniques used, the quality of pipe material, etc. The lowest risk of failure is found when all pipe segments are included (installation period 1891–1996). The value of this function associated with all subsequent breaks is also higher for installation periods 1960–1996 and 1965–1996 (0.205 and 0.182, respectively) compared to all other installation periods (all values below 0.170). The lowest risk of subsequent breaks is also associated with the 1891–1996 installation period.

Comparing parameter values obtained for the different models (Tables 4–7), we observe that values of p_1 are significantly different from one. It should be noted that the exponential distribution is a special case of the Weibull distribution with $p = 1$. Therefore the first break order (parameters p_1 and κ_1) is better described by a Weibull than an exponential distribution. The only exception is for period 1891–1996 for the Weibull-exponential (W-E) model (Table 4) since the p_1 value is very close to 1. In that case an exponential distribution could be used. For the second break the p_2 values (Tables 6 and 7) are often smaller than 1, meaning that the risk of failure decreases with time, which would not be expected. However, in all cases but one (installation period 1891–1996 with the Weibull-Weibull-exponential model), the values of p_2 cannot be considered significantly different from 1. Therefore an exponential distribution should be considered to model second breaks. In the case of exception the use of a Weibull distribution is justified. Section 5 presents a simple method to evaluate if Weibull-Weibull-exponential (W-W-E) and Weibull-

Table 7. Results of the W-W-E-E Model for Different Installation Periods

Installation Periods	p_1	p_2	κ_1	κ_2	κ_3	κ_4
1976–1996	1.157	0.922	0.017	0.070	0.166	0.521
1970–1996	1.184	0.991	0.018	0.079	0.130	0.374
1965–1996	1.342	0.993	0.023	0.084	0.140	0.407
1960–1996	1.404	1.008	0.025	0.081	0.119	0.494
1949–1996	1.243	0.955	0.020	0.061	0.071	0.438
1891–1996	1.123	0.917	0.018	0.055	0.049	0.420

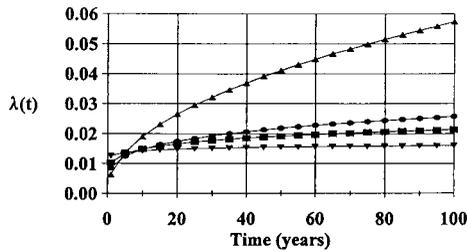


Figure 6. First-order break hazard function for the Weibull-exponential model. The different curves correspond to the following periods: (square) 1976–1996, (triangle) 1960–1996, (circle) 1949–1996, and (inverted triangle) 1891–1996.

exponential-exponential (W-E-E) models are significantly different, permitting us to determine if a Weibull distribution is a better choice than an exponential distribution to model second break orders.

Is there a gain to be made by distinguishing between the second, third, and fourth break orders? Discriminating between the second breaks and all subsequent breaks for the W-E-E model leads to significantly different values for the κ_2 and κ_3 parameters compared to the κ_2 parameter of the W-E model. However, parameter values for κ_2 of the W-E model are roughly equal to the midvalue between the κ_2 and κ_3 parameter values of the W-E-E model, which is consistent with the fact that it is the average behavior that is modeled.

The same conclusion can be drawn for parameter values for κ_3 of the W-W-E model compared to the κ_3 and κ_4 parameter values of the Weibull-Weibull-exponential-exponential (W-W-E-E) model. With the W-W-E-E model the values of κ_4 (fourth break order) are roughly 3 times higher than the values of parameter κ_3 (third break order). This could be symptomatic of the fact that a group of pipe segments is largely responsible for the high number of subsequent breaks observed in recent years. This hypothesis should be further examined. Globally, we observe that as we have more breaks on a pipe segment, the probability of subsequent breaks increases (in Table 7 $\kappa_4 \geq \kappa_3 \geq \kappa_2$). This has been reported by many authors [Clark et al., 1982; Goulter and Kazemi, 1988]. Therefore there is a gain to be made by distinguishing between higher break orders as long as the number of pipes in the sample used to estimate the parameter values is statistically significant.

5. Discrimination Between W-W-E and W-E-E Models

To evaluate if W-W-E and W-E-E models are significantly different, we verified the hypothesis $p_2 = 1$ for the W-W-E model. Given θ , a vector of parameters, the likelihood ratio $R(\theta)$, which can be used for inference about θ , was calculated [Kalbfleisch and Prentice, 1980]. The statistic $-2 \log R(\theta_0)$ has an asymptotic χ^2 distribution if $\theta = \theta_0$, and an approximate 95% confidence interval for θ can be obtained. Refer to Kalbfleisch and Prentice [1980] for details. Table 8 presents the minimum and maximum values of the 95% confidence interval for the p_2 parameter of the W-W-E model.

When the value 1 for parameter p_2 is included in the 95% confidence interval, the difference between a Weibull and an exponential distribution for the second break order is not significant at that confidence level. As presented in Table 8, the only case where the difference is significant is for the installa-

tion period 1891–1996. Also, since p_2 is significantly smaller than 1, it means that the hazard function decreases with time; thus, as time goes by, the probability of having a second pipe break on a given pipe segment decreases. This behavior is atypical, not consistent with the physical processes going on. Many hypotheses can be put forward. Among these it is possible that the Weibull distribution is inappropriate to describe the second break order for older pipe segments or is inappropriate to describe long-term behavior for these pipes. In order to verify this, a more precise analysis would be necessary, and that is beyond the scope of this paper. From this example we see that likelihood ratio is useful for comparing different types of models and for estimating if the available data make it possible to discriminate between them.

6. Summary and Conclusion

Rare are the municipalities that have recorded pipe breaks since the installation of the first pipes of their water pipe networks. In the case study presented here, the water pipe network of Chicoutimi is more than a century old, but the recorded pipe break history only covers 21 years (1976–1996). Different modeling strategies have been proposed in the literature to describe the physical degradation of pipes with time. The most often used indicator of the physical degradation is the annual number of pipe breaks on the network. Many of the strategies used to model pipe breaks are based on statistical methods, especially survival analysis, to model pipe break behavior and estimate the calibration parameters of those models. Using “standard” survival analysis is not possible in the case where the recorded pipe break history does not cover the entire history of the water pipe network. The main reason is that the exact order of breaks that occurred during the observation period is unknown. Survival analysis necessitates stratification of the times to failure depending on the exact break order so that each break order can be modeled with an appropriate distribution. To use survival analysis, one would need to use only pipe segments that were installed during the observation period, which is often only a small portion of the information available.

To model the evolution of the number of pipe breaks in a municipality with a brief recorded pipe break history, we have developed a rigorous approach inspired by survival analysis that permits the estimation of calibration parameters of a statistical pipe break model while integrating all the available information, even though the pipe break history is only partially known for most pipe segments. The approach is based on likelihood functions that once maximized yield the optimal set of calibration parameters for the pipe break model considered. For the application in the municipality of Chicoutimi, four pipe

Table 8. The 95% Confidence Interval for the p_2 Parameter of the W-W-E Model for Different Installation Periods

Installation Periods	Average	Minimum	Maximum
1976–1996	0.924	0.692	1.195
1970–1996	0.999	0.825	1.196
1960–1996	1.043	0.915	1.185
1949–1996	0.968	0.853	1.107
1891–1996	0.779	0.711	0.924

break models were considered, using different probability distributions for different orders of breaks. The distributions used are the two-parameter Weibull and the one-parameter exponential distributions. The four pipe break models, presented in Table 2, discriminate between two to four orders of breaks and include three to six calibration parameters. For each of these pipe break models we have estimated the optimal values of parameters considering different pipe installation periods. This was a useful procedure to verify possible errors in defining the functions and also to assess the impact of the different installation periods on the analysis. As a matter of fact, the procedure enabled us to identify the 1960s as a critical period for Chicoutimi, since pipe segments installed during that time failed more rapidly than all other pipe segments. These results show that two or even three models of pipe break behavior should be used: one for pipe segments installed before the 1960s, one for those installed in the 1960s, and finally one for after the 1960s. There can be many possible explanations for the change of behaviors, but the most probable causes are related to the pipe materials and installation techniques used. Many authors have also observed that pipe breakage varies considerably between installation periods [e.g., *Kleiner and Rajani*, 1999] and thus should be taken into account by using different parameter sets (or even different distributions) for different installation periods.

It is also possible to assess, on a statistical basis, the relative performance of a Weibull distribution compared to an exponential distribution for a particular break order. An example was presented for the W-W-E and the W-E-E models. The likelihood ratio was used to calculate a 95% confidence interval on the value of the p parameter which distinguishes a Weibull from an exponential distribution, since $p = 1$ in the latter case. In the presented example, we could conclude that the use of a Weibull distribution was only justified in the case of the 1891–1996 installation period, which represents the entire history of the water pipe network. In fact, it is interesting to note that very old pipe segments seem to have a much better break record than “younger” pipe segments. The reason for this behavior has yet to be elucidated.

The application of the developed approach to other municipalities with brief recorded pipe break histories is currently being done. We are further investigating the impact of the duration of the recorded history on the quality of the results, and we plan to assess the impact of different risk factors (e.g., pipe material and pipe diameter) on the analysis.

Notation

$f(t)$	probability density function.
n_0	number of pipe segments that have not failed during the observation period.
n_b	total number of recorded breaks.
n_t	total number of pipe segments.
p	“scale” parameter of the Weibull distribution.
$F(t)$	survival function.
L	likelihood function.
$P(\alpha, \beta; \{t_j\})$	probability to have α nonrecorded pipe breaks no matter when these breaks occurred and β pipe breaks in the observation period at times $\{t_j\}$.
$P'(\beta; \{t_j\})$	probability to have β pipe breaks in the observation period at times $\{t_j\}$ no matter

	how many and when breaks occurred during the nonrecorded period.
t_j	time of recorded break j .
t'_j	time of nonrecorded break j .
T_a	time of analysis (present time).
T_b	time of the beginning of the observation period.
κ	“shape” parameter of the Weibull or the exponential distributions.
$\lambda(t)$	hazard function.

Acknowledgments. This research was financed in part by the Ministry of Municipal Affairs of Quebec. The authors would like to thank Normand Bouchard of the municipality of Chicoutimi for providing all the useful data and his expertise on the water pipe network and the reviewers of the paper for their useful comments.

References

- American Water Works Association, An assessment of water distribution systems and associated research needs, report, Am. Water Works Assoc. Res. Found., Denver, Colo., 1994.
- Andreou, S. A., D. H. Marks, and R. M. Clark, A new methodology for modelling break failure patterns in deteriorating water distribution systems: Theory, *Adv. Water Resour.*, 10(3), 2–10, 1987.
- Canada Mortgage and Housing Corporation, Urban infrastructure in Canada, report, Ottawa, Ont., Canada, 1992.
- Choate, P., and S. Walter, America in ruins, report, Counc. of State Plann. Agencies, Washington, D. C., 1981.
- Clark, R. M., C. L. Stafford, and J. A. Goodrich, Water distribution systems: A spatial and cost evaluation, *J. Water Resour. Plann. Manage.*, 108(3), 243–256, 1982.
- Clark, R. M., R. G. Eilers, and J. A. Goodrich, Distribution system: Cost of repair and replacement, in *Proceedings of the Conference on Pipeline Infrastructure*, edited by B. A. Bennett, pp. 428–440, Am. Soc. of Civ. Eng., Reston, Va., 1988.
- Cox, D. R., and D. Oakes, *Analysis of Survival Data*, 201 pp., Chapman and Hall, New York, 1994.
- Desbiens, M.-E., La problématique des infrastructures urbaines: Sa résolution par les voies de la méthodologie et de la technologie, in *Annual Conference of the Canadian Society of Civil Engineering (CSCE)*, edited by P. Labossière, R. Cryslar, and D. T. Lau, pp. 37–43, Can. Soc. Civ. Eng., Montreal, Que., Canada, 1997.
- Eisenbeis, P., Modélisation statistique de la prévision des défaillances sur les conduites d'eau potable, in *Études du CEMAGREF, Sér. Equip. Environ.* 17, CEMAGREF, Bordeaux, France, 1994.
- Goulter, I. C., and A. Kazemi, Spatial and temporal groupings of water main pipe breakage in Winnipeg, *Can. J. Civ. Eng.*, 15, 91–97, 1988.
- Goulter, I. C., J. Davidson, and P. Jacobs, Predicting water-main breakage rates, *J. Water Resour. Plann. Manage.*, 119(4), 419–436, 1993.
- Herz, R. K., Ageing processes and rehabilitation needs of drinking water distribution networks, *J. Water SRT Aqua*, 45(5), 221–231, 1996.
- Jacobs, P., and B. Karney, GIS development with application to cast iron water main breakage rate, paper presented at the 2nd International Conference on Water Pipeline Systems, BHR Group Ltd., Edinburgh, Scotland, May, 1994.
- Kalbfleisch, J. D., and R. L. Prentice, *The Statistical Analysis of Failure Time Data*, Wiley Series in Probability and Mathematical Statistics, 321 pp., John Wiley, New York, 1980.
- Karaa, F. A., and D. H. Marks, Performance of water distribution networks: Integrated approach, *J. Performance Constr. Facil.*, 4(1), 51–67, 1990.
- Kettler, A. J., and I. C. Goulter, An analysis of pipe breakage in urban water distribution networks, *Can. J. Civ. Eng.*, 12, 286–293, 1985.
- Kleiner, Y., and B. B. Rajani, Using limited data to assess future needs, *J. Am. Water Works Assoc.*, 91(7), 47–62, 1999.
- Kleiner, Y., B. J. Adams, and J. S. Rogers, Long-term planning methodology for water distribution system rehabilitation, *Water Resour. Res.*, 34(8), 2039–2051, 1998a.
- Kleiner, Y., B. J. Adams, and J. S. Rogers, Selection and scheduling of rehabilitation alternatives for water distribution systems, *Water Resour. Res.*, 34(8), 2053–2061, 1998b.

- McDonald, S., L. Daigle, and G. Félio, Réseaux d'aqueduc et systèmes d'égouts, report, Infrastruct. Lab., Inst. for Res. in Construct., Natl. Res. Counc. of Can., Ottawa, Ont., 1994.
- Miller, R. G., *Survival Analysis*, 238 pp., John Wiley, New York, 1981.
- O'Day, D. K., R. Weiss, S. Chiavari, and D. Blair, Water main evaluation for rehabilitation/replacement, report, Am. Water Works Assoc. Res. Found., Denver, Colo., 1986.
- Pelletier, G., Impact du remplacement des conduites d'aqueduc sur le nombre annuel de bris, Ph.D. thesis, Inst. Natl. de la Rech. Sci., Sainte-Foy, Qué., Canada, 2000.
- Press, W. H., B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes in C*, 735 pp., Cambridge Univ. Press, New York, 1988.
- Shamir, U., and C. Howard, Analytic approach to scheduling pipe replacement, *J. Am. Water Works Assoc.*, 71(5), 248–258, 1979.
- Siddiqui, S., and S. Mirza, Canadian municipal infrastructure—The present state, in *Annual Conference of the Canadian Society of Civil Engineering (CSCE)—1st Transportation Specialty Conference*, pp. 519–530, Can. Soc. of Civ. Eng., Montreal, Que., Canada, 1996.
- Villeneuve, J.-P., S. Duchesne, A. Mailhot, E. Musso, and G. Pelletier, Évaluation des besoins des municipalités québécoises en réfection et construction d'infrastructures d'eaux—Rapport final—INRS-Eau, research report, Inst. Natl. de la Rech. Sci., Sainte-Foy, Qué., Canada, 1998.
-
- A. Mailhot, J.-F. Noël, G. Pelletier, and J.-P. Villeneuve, Institut National de la Recherche Scientifique, INRS-Eau 2800, rue Einstein, C. P. 7500, Sainte-Foy, Québec, Canada C1V 4C7. (Alain_Mailhot@inrs-eau.quebec.ca)

(Received June 21, 1999; revised April 13, 2000; accepted June 6, 2000.)