

# Research Article (postprint)

This research was conducted within the research project “Sustainable Water Infrastructure Planning”, [http://www.eawag.ch/forschung/sww/schwerpunkte/infrastrukturen/planung\\_wasserinfrastr/index\\_EN](http://www.eawag.ch/forschung/sww/schwerpunkte/infrastrukturen/planung_wasserinfrastr/index_EN). It was funded within the National Research Programme 61 on “Sustainable Water Management” by the Swiss National Science Foundation (<http://www.nrp61.ch>).

## Extension of pipe failure models to consider the absence of data from replaced pipes

Andreas Scheidegger<sup>a,\*</sup>, Lisa Scholten<sup>a,b</sup>, Max Maurer<sup>a,b</sup> and Peter Reichert<sup>a,c</sup>

<sup>a</sup> Eawag, Swiss Federal Institute of Aquatic Science and Technology, Ueberlandstrasse 133, P.O. Box 611, CH-8600 Dübendorf, Switzerland

<sup>b</sup> ETH Zürich, Department of Civil, Environmental and Geomatic Engineering, Institute of Environmental Engineering, Wolfgang-Pauli-Strasse 15, CH-8093 18 Zurich, Switzerland 19

<sup>c</sup> ETH Zürich, Department of Environmental Systems Science, Institute of 20 Biogeochemistry and Pollutant Dynamics, CH-8092 Zurich, Switzerland

\* Corresponding author. E-Mail: [andreas.scheidegger@eawag.ch](mailto:andreas.scheidegger@eawag.ch)

Published in *Water Research* 47 (11): 3696-3705.

View at publisher: <http://www.sciencedirect.com/science/article/pii/S0043135413003461>

**NOTICE:** this is the author's version of a work that was accepted for publication in *Water Research*. Changes resulting from the publishing process, such as editing, corrections, structural formatting, and other quality control mechanisms may not be reflected in this document. Changes may have been made to this work since it was submitted for publication. A definitive version was subsequently published in *Water Research* 47(11): 3696-3705. DOI: 10.1016/j.watres.2013.04.017.

# Extension of pipe failure models to consider the absence of data from replaced pipes

Andreas Scheidegger<sup>a</sup>, Lisa Scholten<sup>a</sup>, Max Maurer<sup>a,b,\*</sup>, Peter Reichert<sup>a</sup>

<sup>a</sup>*Eawag, Swiss Federal Institute of Aquatic Science and Technology, Überlandstrasse 133, 8600 Dübendorf, Switzerland*

<sup>b</sup>*Institute of Environmental Engineering, ETH Zurich, 8093 Zürich, Switzerland*

---

## Abstract

Predictions of the expected number of failures of water distribution network pipes are important to develop an optimal management strategy. A number of probabilistic pipe failure models have been proposed in the literature for this purpose. They have to be calibrated on failure records. However, common data management practices mean that replaced pipes are often absent from available data sets. This leads to a 'survival selection bias', as pipes with frequent failures are more likely to be absent from the data.

To address this problem, we propose a formal statistical approach to extend the likelihood function of a pipe failure model by a replacement model. Frequentist maximum likelihood estimation or Bayesian inference can then be applied for parameter estimation. This approach is general and is not limited to a particular failure or replacement model.

We implemented this approach with a Weibull-exponential failure model and a simple constant probability replacement model. Based on this distribution assumptions, we illustrated our concept with two examples. First, we used simulated data to show how replacement causes a 'survival selection bias' and how to successfully correct for it. A second example with real data illustrates how a model can be extended to consider covariables.

### *Keywords:*

Pipe failure model, replacement model, likelihood, Bayesian inference, survival selection bias

---

\*Corresponding author: max.maurer@eawag.ch

## 1. Introduction

The optimal management strategy for water distribution networks balances issues of water safety, reliability, quality, and quantity, while exploiting the full extent of the useful life of the pipes to achieve economic efficiency (Kleiner and Rajani, 2001). Pipe failure models are one of the key tools to support this management process.

We distinguish between two major applications of pipe failure models: (i) The failure probabilities of the individual pipes are needed for the mid-term maintenance and replacement strategies of the pipe network (ii) For long-term planning, the expected number of failures in the entire system is of interest, but not the specific cause of the failures. It is therefore sufficient to model all deterioration processes lumped together as a function of age, so that less detailed data are required. Applications (i) and (ii) do not require fundamentally different model structures, because models for (ii) can typically be extended to fulfill the needs of (i) by incorporating pipe properties such as material, diameter, etc. to improve pipe-specific predictions.

The model should be calibrated on the basis of failure records of the local system because of differences in the influence factors that are not modeled (e.g. soil properties). Correct calibration can become challenging because the available data typically show some or all of the following properties (see also Figure 1):

- Right censored observations (Figure 1-i): For every pipe in service a right censored observation is available: the time since the last failure or construction until the time of observation. This provides important information, and pipes without recorded failures until the end of the observation period must not be excluded from the calibration process. This issue is considered in the calibration procedures of many time-based failures models (Carrión et al., 2010; Eisenbeis et al., 1999; Gustafson and Clancy, 1999; Mailhot et al., 2000). For models formulated as a counting process (e.g. Economou et al., 2008; Kleiner et al., 2010; Watson et al., 2004), right censoring is not relevant, because the probability of a certain number of failures within a time interval is modeled instead of the time between the failures.
- Left truncation (Figure 1-ii): Left truncation occurs if a pipe was installed before failures were systematically recorded by the utility. As a consequence, it is not known how many failures occurred before the recording period. Only few models (Carrión et al., 2010; Le Gat,

2009; Mailhot et al., 2000) explicitly consider left truncation.

- Absence of replaced pipe data (Figure 1-iii): Frequently, replaced pipes are deleted from the database together with the corresponding pipe failure data because the database was established with the objective of reflecting the current state. This leads to a “selective survival bias” (Renaud et al., 2011), due to the fact that pipes with poor failure histories will be underrepresented in the data set. Hence, ignoring this in the parameter estimation causes systematic errors in the predictions which cannot be reduced merely by increasing the amount of data (Scheidegger et al., 2011).

The intuitive idea to consider the survival selection bias is to assess how likely it is that a pipe similar to the observed has been replaced in the past and correct the likelihood function accordingly. This requires the integration of a *replacement model* that characterizes the probability that a pipe was not replaced, i.e. the chance that a pipe is still in service. Generally, this can be a function of the condition, age and number of failures a pipe has already experienced. The parameters of the replacement model are then estimated jointly with those of the failure model.

To the best of our knowledge only the LEYP model (linear extension of the Yule process) developed in the dissertation of Le Gat (2009) attempts to tackle the selective survival bias. Le Gat (2009) specified the probability of a pipe is not being replaced if a failure occurs as a function of pipe age. The chosen double exponential form allows an analytical evaluation of the likelihood of the LEYP model. In the form presented, however, this approach is difficult to generalize and to transfer to other failure and replacement models. For example, a replacement decision might not depend on age but on the number of previous failures.

In this paper we propose a general framework to derive the likelihood function of a pipe failure model combined with any kind of replacement model to enable unbiased parameter estimation from data sets without historical records. The likelihood function derived in Section 2 has a frequentist interpretation so that the parameters can either be estimated according to the maximum likelihood principle or by Bayesian inference. The latter is favorable in two common cases: (i) Small utilities often have very limited data, either because they have been recording failures only for a short time or simply because they have small networks. However, they typically have dedicated experts with sound practical experience beyond the information in archives. In this situation, carefully elicited expert knowledge can improve

the model performance (Scholten et al., 2013). (ii) The parameters of the replacement model can correlate strongly with those of the failure model and may therefore lead to identifiability problems in a frequentist setting. In a Bayesian framework, this can be circumvented by an informative prior distribution.

The remainder of this paper is structured as follows. In Section 2 we first introduce a universal notation for pipe failure models and then derive the likelihood function for completely and partially observed pipes. On this basis, we illustrate how a failure model can be extended with a replacement model in general. Furthermore, the predictive distributions for the number of failures for pipes with and without failure record are presented. As an example, the equations are derived explicitly for a Weibull-exponential model in Section 3. In the following section, this model is used for two application examples: the first is based on artificial data to highlight the importance of the replacement model. The second illustrates (with real data) how individual pipe properties can be considered. Finally, we discuss the strengths and weaknesses of our approach and point out directions for further research.

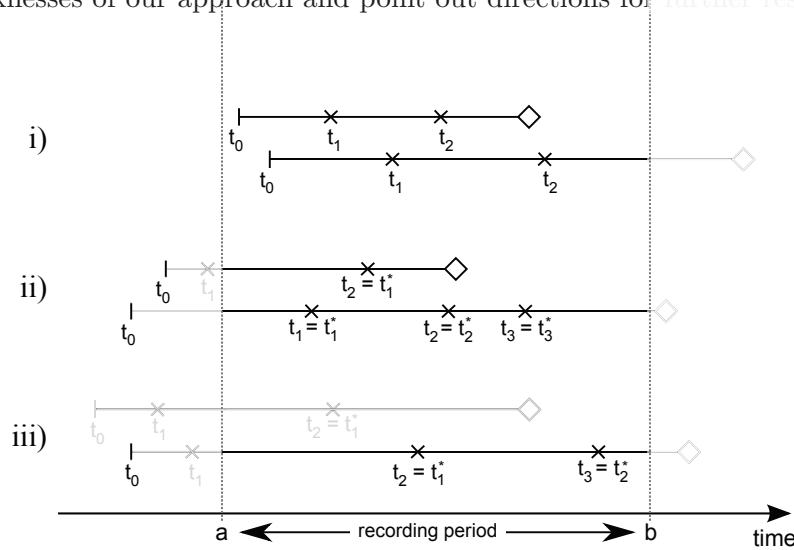


Figure 1: Three scenarios with different data availabilities. The available information is shown in black, the unavailable information in gray.  $a$  marks the beginning and  $b$  the end of the recording period,  $\times$  a failure,  $\diamond$  the replacement,  $t_0$  is the time of construction,  $t_i$  the time of the  $i$ th failure, and  $t_i^*$  the time of the  $i$ th *recorded* failure. i) All failures of the pipes are recorded and data of replaced pipes remain in the data set; ii) Failures before  $a$  are not recorded, the number of failures per pipe is unknown; iii) The total number of failures per pipe is unknown and data of replaced pipes is unavailable.

## 2. Methods

### 2.1. Pipe failure model

As long as a pipe is in operation there is a chance of a failure event. We define a failure as an observable event that requires immediate measures (e.g. a break). Other definitions are possible, depending on the available records. It is assumed that failures are repaired immediately without replacing the pipe.

For a single pipe, the point in time of the  $i$ th failure is denoted by  $t_i$  while  $t_0$  stands for the time of construction. The time when the  $i$ th failure occurs is random and therefore described by a probability density function  $p_i(t|t_0, \dots, t_{i-1}, \boldsymbol{\theta})$  or a survival probability  $\text{Prob}(t_i > t|t_0, \dots, t_{i-1}, \boldsymbol{\theta}) = S_i(t|t_0, \dots, t_{i-1}, \boldsymbol{\theta})$ . Obviously, the  $t_i$ ,  $i \geq 0$  are not independent as the  $i$ th failure cannot occur before the  $(i-1)$ th failure. The vector  $\boldsymbol{\theta}$  represents the parameters of all distributions.

This formulation enables us to express different standard models with the same notation. Models based on a counting process can be written equivalently as time-based models. For example for a homogeneous Poisson process, we would define  $p_i(t|t_0, \dots, t_{i-1}, \boldsymbol{\theta}) = p(t - t_{i-1}|\boldsymbol{\theta}) = \lambda e^{-\lambda(t-t_{i-1})}$  for  $i > 0$ , i.e. the time differences between two failures are all exponentially distributed with the same rate  $\lambda$ .

To statistically estimate the parameters  $\boldsymbol{\theta}$  and for failure predictions, a likelihood function is required. The likelihood is the joint probability (density) of the observed  $n_k$  failures at times  $\mathcal{T}_k = \{t_{k,i} : i = 0, \dots, n_k\}$  for all pipes  $k = 1, \dots, K$  given the model and the parameters.

We define  $a$  as the time at the beginning of the recording period and  $b$  as the end of the recording period. All failures are recorded within this period. The likelihood function for a pipe failure model for two data collection schemes is derived below (compare Figure 1): i) the complete life of the pipe lies within the recording period, and ii) the pipe was built before recording started.

All the following equations apply to a single pipe unless otherwise stated. For the sake of simpler notation, the pipe index  $k$  is omitted in equations that refer to a single pipe.

#### 2.1.1. Likelihood for completely observed pipes

If  $a \leq t_0$ , the recording period covers the complete life of the pipe. For this situation the likelihood of  $n$  failures at times  $\mathcal{T} = \{t_i : i = 0, \dots, n\}$  for

one pipe is formulated as

$$p(\mathcal{T}, n|b, \boldsymbol{\theta}) = \left[ \prod_{i=1}^n p_i(t_i|t_0, \dots, t_{i-1}, \boldsymbol{\theta}) \right] S_{n+1}(b|t_0, \dots, t_n, \boldsymbol{\theta}) \quad (1)$$

where  $\boldsymbol{\theta}$  represents the parameters of the distributions. The factor  $S_{n+1}(b|t_0, \dots, t_n, \boldsymbol{\theta})$  accounts for the fact that there is always a right censored observation available: the time from the last failure (or from construction) until the end of the observation period or the replacement of the pipe.

### 2.1.2. Likelihood for partly observed pipes

If a pipe was built before the observation period began ( $t_0 < a$ ) it is not known how many (if any) failures have occurred before  $a$ .

The likelihood proposed by Mailhot et al. (2000) accounts for this. In the following a distinction must be made between  $t_i^*$ , the point in time of the  $i$ th *recorded* failure and  $t_i$ , the time of the  $i$ th failure which is not necessarily equal to  $t_i^*$ . The  $n$  recorded failures are summarized as  $\mathcal{T}^* = \{t_i^* : i = 0, \dots, n\}$ . Additionally the time of construction  $t_0$  is assumed to be known. For convenient notation we define  $t_0^* := t_0$ . Note that  $p_i(t|\theta_i)$  still stands for the density of the time of the  $i$ th (observed or unobserved) failure.

Mailhot et al. (2000) first derived the joint distribution of the number of non-recorded failures  $m$  and the  $n$  recorded failures at  $\mathcal{T}^*$ . Adapted to our notation and slightly generalized, this is written as

$$\begin{aligned} p(\mathcal{T}^*, m, n|a, b, \boldsymbol{\theta}) = & \int_{t_0}^a \int_{t_1}^a \dots \int_{t_{(m-1)}}^a p_1(t_1|t_0) p_2(t_2|t_0, t_1) \dots p_m(t_m|t_0, \dots, t_{m-1}) \\ & \cdot p_{m+1}(t_1^*|t_0, \dots, t_m) p_{m+2}(t_2^*|t_0, \dots, t_m, t_1^*) \dots p_{m+n}(t_n^*|t_0, \dots, t_m, t_1^*, \dots, t_{n-1}^*) \\ & \cdot S_{m+n+1}(b|t_0, \dots, t_m, t_1^*, \dots, t_n^*) dt_m \dots dt_2 dt_1 \end{aligned} \quad (2)$$

for  $m > 0$ . For no non-recorded failures,  $m = 0$ , the density  $p(\mathcal{T}^*, m = 0, n|a, b, \boldsymbol{\theta})$  takes the form of (1).

The likelihood for a single pipe is then obtained by summing (2) over  $m$ :

$$p(\mathcal{T}^*, n|a, b, \boldsymbol{\theta}) = \sum_{m=0}^{\infty} p(\mathcal{T}^*, m, n|a, b, \boldsymbol{\theta}) \quad (3)$$

## 2.2. Replacement model

The replacement model has to express the probability of the event 'pipe has not been replaced up to time  $b$ ' (abrv. 'not rep.') given its failure history,  $\text{Prob}(\text{'not rep.'}|\mathcal{T}^*, n, a, b, \boldsymbol{\theta})$ .

It is usually more convenient to formulate the replacement model first conditioned on the number of non-recorded failures, i.e.  $\text{Prob}(\text{'not rep.'} | \mathcal{T}^*, n, m, a, b, \boldsymbol{\theta})$ . The unconditional replacement model is then derived as

$$\text{Prob}(\text{'not rep.'}|\mathcal{T}^*, n, a, b, \boldsymbol{\theta}) = \sum_{m=0}^{\infty} \text{Prob}(\text{'not rep.'}|\mathcal{T}^*, n, m, a, b, \boldsymbol{\theta}) \text{Prob}(m|a, \boldsymbol{\theta})$$

where the probability of  $m$  failures before  $a$  is given by

$$\text{Prob}(m|a, \boldsymbol{\theta}) = \int_{t_0}^a \int_{t_1}^a \dots \int_{t_{(m-1)}}^a \left[ \prod_{i=1}^m p_i(t_i|t_0, \dots, t_{i-1}, \boldsymbol{\theta}) \right] S_{m+1}(a|t_0, \dots, t_n, \boldsymbol{\theta}) dt_m \dots dt_2 dt_1$$

Only those replacements that are related to the failure history, i.e.  $\mathcal{T}^*$  and  $n$ , may be represented by the replacement model. Probabilities for independent replacement cancel out in the fraction of (4).

## 2.3. Joint likelihood

If only data of active pipes are available, the likelihood of the pipe failure model and the replacement model must be combined to infer the parameters of the pipe failure model correctly.

The likelihood of the pipe failure model must be conditioned on the event 'pipe has not been replaced up to time  $b$ '. So the likelihood for an *observed* pipe with  $n$  recorded failures at times  $\mathcal{T}^*$  becomes  $p(\mathcal{T}^*, n|a, b, \text{'not rep.'}, \boldsymbol{\theta})$ . Expressed according to the Bayes' theorem, this is

$$p(\mathcal{T}^*, n|\text{'not rep.'}, a, b, \boldsymbol{\theta}) = \frac{p(\mathcal{T}^*, n|a, b, \boldsymbol{\theta}) \text{Prob}(\text{'not rep.'}|\mathcal{T}^*, n, a, b, \boldsymbol{\theta})}{\text{Prob}(\text{'not rep.'}|a, b, \boldsymbol{\theta})} \quad (4)$$

The numerator is the product of the likelihood of the pipe failure model and the replacement model. The denominator of (4) is the probability that a pipe of age  $b$  has not been replaced, which is obtained by marginalization

$$\text{Prob}(\text{'not rep.'}|a, b, \boldsymbol{\theta}) = \sum_{n=0}^{\infty} \int_a^b \int_{t_1^*}^b \dots \int_{t_{(n-1)}^*}^b p(\mathcal{T}^*, n|a, b, \boldsymbol{\theta}) \text{Prob}(\text{'not rep.'}|\mathcal{T}^*, n, a, b, \boldsymbol{\theta}) dt_n^* \dots dt_2^* dt_1^*$$



To obtain the joint likelihood, the likelihoods of the single pipes are multiplied if they are independent.

$$p(\mathcal{T}_1^*, \dots, \mathcal{T}_K^*, n_1, \dots, n_k | \text{'not rep.'}, a, b, \boldsymbol{\theta}) = \prod_{k=1}^K p(\mathcal{T}_k^*, n_k | \text{'not rep.'}, a, b, \boldsymbol{\theta}) \quad (5)$$

Independence is a reasonable assumption if the pipes are aggregated to a sufficient length (see e.g. Gangl, 2008).

#### 2.4. Consideration of covariables

Up to this point, pipes were not distinguished by their properties such as their diameter or material. The same parameter vector  $\boldsymbol{\theta}$  was used for all pipes. Consideration of pipe properties can help to improve the predictions for a specific pipe or pipe group and enables the identification of important deterioration processes. Covariables are incorporated by calculating “individual” parameters  $\boldsymbol{\theta}_k$  for each pipe  $k$  as a function of their properties  $\mathbf{x}_k$ :

$$\boldsymbol{\theta}_k = f(\mathbf{x}_k, \boldsymbol{\theta}, \boldsymbol{\gamma}) \quad (6)$$

where  $\boldsymbol{\gamma}$  are additional parameters of  $f(\cdot)$  that must be estimated together with  $\boldsymbol{\theta}$ . To include qualitative pipe properties (e.g. material) indicator variables are used.

#### 2.5. Parameter inference

Two widely applied approaches to estimate the parameters are frequentist maximum likelihood estimation (MLE) and Bayesian inference. MLE (e.g. Kleiner et al., 2010; Le Gat, 2009) and Bayesian inference (Dridi et al., 2009; Economou et al., 2008; Watson et al., 2004) have frequently been applied for pipe failure models.

The ML estimator is the parameter vector  $\hat{\boldsymbol{\theta}}$  that maximizes the likelihood function.

$$\hat{\boldsymbol{\theta}} = \arg \max_{\boldsymbol{\theta}} p(\mathcal{T}_1^*, \dots, \mathcal{T}_K^*, n_1, \dots, n_k | \text{'not rep.'}, a, b, \boldsymbol{\theta})$$

Large sample properties allow an approximation of the parameter uncertainty (Harrell, 2001).

With Bayesian inference, the distribution of the parameters is calculated given the data and the prior distribution of the parameters  $p(\boldsymbol{\theta})$ . The prior distribution reflects knowledge about the parameters before the calibration.

The proportional relationship (Bernardo and Smith, 2000) is sufficient for numerical calculations:

$$p(\boldsymbol{\theta}|\mathcal{T}_1^*, \dots, \mathcal{T}_K^*, n_1, \dots, n_k, \text{'not rep.'}, a, b) \propto p(\mathcal{T}_1^*, \dots, \mathcal{T}_K^*, n_1, \dots, n_k | \text{'not rep.'}, a, b, \boldsymbol{\theta}) p(\boldsymbol{\theta}) \quad (7)$$

## 2.6. Predictions

In the following the predictive distribution of the number of failures is derived for new pipes and for pipes with a known failure record. Future replacement is purposely not considered in the predictions, to enable the comparison of replacement strategies. The 'pure' predicted failures can then be used directly as input for different replacement strategies.

For the sake of more compact notation, the following predictive distributions are conditioned on the parameters  $\boldsymbol{\theta}$ . Typically, they will be multiplied by the posterior parameter distribution (7) and then marginalized over  $\boldsymbol{\theta}$ .

### 2.6.1. Unconditional predictions

The predictive distribution for a pipe without a failure record is given by the likelihood (1). Typically, interest is limited to the distribution of the number of failures until age  $c$  which is obtained by marginalization of likelihood (1).

$$\text{Prob}(n|c, \boldsymbol{\theta}) = \int_{t_0}^c \int_{t_1}^c \dots \int_{t_{(n-1)}}^c p(\mathcal{T}, n|c, \boldsymbol{\theta}) dt_n dt_{n-1} \dots dt_1 \quad (8)$$

### 2.6.2. Conditional predictions

To predict the future failures of an existing pipe the failures during the observation period must be considered. Therefore we distinguish between the  $n^{(1)}$  observed failures at  $\mathcal{T}^{*(1)}$  and the  $n^{(2)}$  future failures at  $\mathcal{T}^{*(2)}$ . The predictive distribution of  $\mathcal{T}^{*(2)}$  and  $n^{(2)}$  can be expressed by the likelihood for partially observed failures (2)

$$p(\mathcal{T}^{*(2)}, n^{(2)} | \mathcal{T}^{*(1)}, n^{(1)}, a, b, c, \boldsymbol{\theta}) = \frac{p(\mathcal{T}^{*(1)} \cup \mathcal{T}^{*(2)}, n^{(1)} + n^{(2)} | a, b = c, \boldsymbol{\theta})}{p(\mathcal{T}^{*(1)}, n^{(1)} | a, b, \boldsymbol{\theta})} \quad (9)$$

The condition 'not rep.' is not required as it cancels out algebraically. Finally, the distribution of the number of future failures is given by

$$\begin{aligned} \text{Prob}(n^{(2)}|\mathcal{T}^{*(1)}, n^{(1)}, a, b, c, \boldsymbol{\theta}) = \\ \int_b^c \int_{t_1^{(2)}}^c \cdots \int_{t_{(n-1)}^{(2)}}^c p(\mathcal{T}^{*(2)}, n^{(2)}|\mathcal{T}^{*(1)}, n^{(1)}, a, b, c, \boldsymbol{\theta}) dt_{n^{(2)}}^{(2)} dt_{(n^{(2)}-1)}^{(2)} \cdots dt_1^{(2)} \end{aligned} \quad (10)$$

### 3. Example: Weibull-exponential model

While the general description above provides the 'recipe' for the likelihood for a particular model, there is no assurance that the resulting likelihood can be handled algebraically and numerically. In this section, we show how the likelihood for a rather simple pipe failure model that was applied by Mailhot et al. (2000) can be combined with an elementary replacement model.

For the pipe failure model, we assume that the time from construction until the first failure is Weibull distributed, and the time between all following failures exponential with the same rate parameter.

This failure model requires three parameters: the shape parameter  $\theta_1$  and the scale  $\theta_2$  of the Weibull distribution

$$\begin{aligned} p_1(t|t_0, \boldsymbol{\theta}) &= \frac{\theta_1}{\theta_2} \left( \frac{t - t_0}{\theta_2} \right)^{\theta_1 - 1} e^{-[(t - t_0)/\theta_2]^{\theta_1}} \\ S_1(t|t_0, \boldsymbol{\theta}) &= e^{-[(t - t_0)/\theta_2]^{\theta_1}} \end{aligned} \quad (11)$$

and the scale  $\theta_3$  of the exponential distribution

$$\begin{aligned} p_i(t|t_0, \dots, t_{i-1}, \boldsymbol{\theta}) &= p_i(t|t_{i-1}, \boldsymbol{\theta}) = \frac{1}{\theta_3} e^{-(t - t_{i-1})/\theta_3} \\ S_i(t|t_0, \dots, t_{i-1}, \boldsymbol{\theta}) &= S_i(t|t_{i-1}, \boldsymbol{\theta}) = e^{-(t - t_{i-1})/\theta_3} \end{aligned} \quad (12)$$

for all  $i > 1$ .

#### 3.1. Likelihood for completely observed pipes

The distributions defined in (11) and (12) are directly plugged into the general likelihood for completely observed pipes (1). After some algebraic

rearrangements, we obtain

$$p(\mathcal{T}, n|b, \boldsymbol{\theta}) = \begin{cases} e^{-[(b-t_0)/\theta_2]^{\theta_1}}, & n = 0 \\ \frac{\theta_1}{\theta_2} \left( \frac{t_1-t_0}{\theta_2} \right)^{\theta_1-1} e^{-[(t_1-t_0)/\theta_2]^{\theta_1}} \left( \frac{1}{\theta_3} \right)^{n-1} e^{-(b-t_1)/\theta_3}, & n > 0 \end{cases} \quad (13)$$

Note that due to the algebraic form of the exponential distribution and the assumption that the rate parameter remains the same for all  $i > 1$  the likelihood only depends on the time of the first failure and on the number of failures  $n$ .

### 3.1.1. Likelihood for partly observed pipes

Similarly, the likelihood for partly observed pipes must be distinguished for  $n = 0$  and  $n > 0$ . If no failures are observed ( $n = 0$ ), it is

$$p(\mathcal{T}^*, n = 0|a, b, \boldsymbol{\theta}) = e^{-[(b-t_0)/\theta_2]^{\theta_1}} + e^{-(b-a)/\theta_3} \left[ 1 - e^{-[(a-t_0)/\theta_2]^{\theta_1}} \right] \quad (14)$$

and for  $n > 0$

$$p(\mathcal{T}^*, n|a, b, \boldsymbol{\theta}) = \left( \frac{1}{\theta_3} \right)^{n-1} e^{-(b-t_1^*)/\theta_3} \left\{ \frac{\theta_1}{\theta_2} \left( \frac{t_1^*-t_0}{\theta_2} \right)^{\theta_1-1} e^{-[(t_1^*-t_0)/\theta_2]^{\theta_1}} + \frac{1}{\theta_3} e^{(a-t_0)/\theta_3} \left[ 1 - e^{-[(a-t_0)/\theta_2]^{\theta_1}} \right] \right\} \quad (15)$$

See Mailhot et al. (2000) or Pelletier (2000) for a derivation of these equations.

### 3.2. Consideration of replacement

The replacement model is first defined conditioned on  $m$ . The simplest model with a single parameter assumes a constant probability  $\pi$  that a pipe is not replaced if a failure occurs:  $\text{Prob}(\text{'not rep.'}|\mathcal{T}^*, n, m, a, b, \boldsymbol{\theta}) = \pi^{m+n}$ . In the following we assume that  $\pi$  is contained in the parameter vector  $\boldsymbol{\theta}$ . The resulting unconditional replacement model is then

$$\text{Prob}(\text{'not rep.'}|\mathcal{T}^*, n, a, b, \boldsymbol{\theta}) = \sum_{m=0}^{\infty} \pi^{m+n} \text{Prob}(m|a, \boldsymbol{\theta}) \quad (16)$$

where the probability of  $m$  unobserved failures before  $a$  is

$$\begin{aligned} \text{Prob}(m|a, \boldsymbol{\theta}) &= \begin{cases} e^{-[(a-t_0)/\theta_2]^{\theta_1}}, & m = 0 \\ \int_{t_0}^a \int_{t_1}^a \cdots \int_{t_{(m-1)}}^a \frac{\theta_1}{\theta_2} \left( \frac{t_1-t_0}{\theta_2} \right)^{\theta_1-1} e^{-[(t_1-t_0)/\theta_2]^{\theta_1}} \\ \quad \left( \frac{1}{\theta_3} \right)^{m-1} e^{-(a-t_1)/\theta_3} dt_m \cdots dt_2 dt_1, & m > 0 \end{cases} \\ &= \begin{cases} e^{-[(a-t_0)/\theta_2]^{\theta_1}}, & m = 0 \\ \frac{\theta_1}{\theta_2} \int_{t_0}^a \left( \frac{t_1-t_0}{\theta_2} \right)^{\theta_1-1} e^{-[(t_1-t_0)/\theta_2]^{\theta_1}} \left( \frac{1}{\theta_3} \right)^{m-1} e^{-(a-t_1)/\theta_3} \frac{(a-t_1)^{m-1}}{(m-1)!} dt_1, & m > 0 \end{cases} \end{aligned}$$

The integrand depends only on  $t_1$  so the remaining integrals reduce to  $\int_{t_1}^a \cdots \int_{t_{(m-1)}}^a 1 dt_m \cdots dt_2 = \frac{(a-t_1)^{m-1}}{(m-1)!}$ .

Combined with the conditional replacement model, the sum and the factorials can be simplified by recognizing that  $\sum_{k=0}^{\infty} x^k/k! = e^x$  (see also Pelletier et al. (2003), page 83):

$$\begin{aligned} \text{Prob}(\text{'not rep.'}|\mathcal{T}^*, n, a, b, \boldsymbol{\theta}) &= e^{-[(a-t_0)/\theta_2]^{\theta_1}} \\ &\quad + \pi^{n+1} \frac{\theta_1}{\theta_2} \int_{t_0}^a \left( \frac{t_1-t_0}{\theta_2} \right)^{\theta_1-1} e^{-[(t_1-t_0)/\theta_2]^{\theta_1}} e^{-(b-t_1)/\theta_3} \\ &\quad \sum_{m=1}^{\infty} \left( \frac{1}{\theta_3} \right)^{m-1} \pi^{m-1} \frac{(a-t_1)^{m-1}}{(m-1)!} dt_1 \\ &= e^{-[(a-t_0)/\theta_2]^{\theta_1}} \\ &\quad + \pi^{n+1} \frac{\theta_1}{\theta_2} \int_{t_0}^a \left( \frac{t_1-t_0}{\theta_2} \right)^{\theta_1-1} e^{-[(t_1-t_0)/\theta_2]^{\theta_1}} e^{-(b-t_1)/\theta_3} \\ &\quad e^{\pi(a-t_1)/\theta_3} dt_1 \end{aligned} \tag{17}$$

The replacement model (17), and the likelihood of the pipe failure model for partly observed pipes without (14) or with failures (15) are combined in (4) to obtain the conditional likelihood. Although the integrals of the denominator cannot be solved analytically, it can be simplified sufficiently

(details not shown) to allow numerical integration without problems:

$$\begin{aligned}
\text{Prob}(\text{'not rep.'} | a, b, \boldsymbol{\theta}) &= e^{-[(b-t_0)/\theta_2]^{\theta_1}} \\
&+ \frac{\theta_1}{\theta_2} \int_{t_0}^a \left( \frac{t_1 - t_0}{\theta_2} \right)^{\theta_1 - 1} e^{-[(t_1 - t_0)/\theta_2]^{\theta_1}} e^{-(b-t_1)/\theta_3} \pi e^{\pi(a-t_1)/\theta_3} dt_1 \\
&+ \frac{\theta_1}{\theta_2} \int_a^b \left( \frac{t_1^* - t_0}{\theta_2} \right)^{\theta_1 - 1} e^{-[(t_1^* - t_0)/\theta_2]^{\theta_1}} e^{-(b-t_1^*)/\theta_3} \pi e^{\pi(b-t_1^*)/\theta_3} dt_1^* \\
&+ \frac{\theta_1 \pi^2}{\theta_2 \theta_3} \int_{t_0}^a \int_a^b e^{-(b-t_1^*)/\theta_3} e^{\pi(b-t_1^*)/\theta_3} \left( \frac{t_1 - t_0}{\theta_2} \right)^{\theta_1 - 1} \\
&\quad e^{-[(t_1 - t_0)/\theta_2]^{\theta_1}} e^{-(t_1^* - t_1)/\theta_3} e^{\pi(a-t_1)/\theta_3} dt_1^* dt_1 \quad (18)
\end{aligned}$$

### 3.3. Predictions

#### 3.3.1. Unconditional predictions

Monte Carlo samples can be conveniently generated from the likelihood for completely observed pipes (13). From there, it is straightforward to obtain the distribution of the number of failures by sequentially sampling from the Weibull (11) and exponential distributions (12).

#### 3.3.2. Conditional predictions

It is not trivial, however, to sample from the likelihood for partly observed pipes (9). Instead, an expression proportional to  $\text{Prob}(n^{(2)} | \mathcal{T}^{*(1)}, n^{(1)}, a, b, c, \boldsymbol{\theta})$  can be derived, on whose basis a sample can be obtained with importance or Metropolis sampling. The formulations must be distinguished depending on the number of observed  $n^{(1)}$  and predicted  $n^{(2)}$  failures.

If  $n^{(1)} > 0$ :

$$\text{Prob}(n^{(2)} | \mathcal{T}^{*(1)}, n^{(1)}, a, b, c, \boldsymbol{\theta}) \propto p(\mathcal{T}^{*(1)}, n | a, b, \boldsymbol{\theta})_{|n=n^{(1)}+n^{(2)}, b=c} \cdot \frac{(c-b)^{n^{(2)}}}{n^{(2)}!}$$

and if no failures were observed, i.e.  $n^{(1)} = 0$ :

$$\begin{aligned}
&\text{Prob}(n^{(2)} | \mathcal{T}^{*(1)}, n^{(1)}, a, b, c, \boldsymbol{\theta}) \\
&\propto \begin{cases} p(\mathcal{T}^*, n = 0 | a, b, \boldsymbol{\theta})_{|b=c} & n^{(2)} = 0 \\ \int_b^c p(\mathcal{T}^*, n | a, b, \boldsymbol{\theta})_{|n=n^{(2)}, b=c} \cdot \frac{(c-t_1^*)^{n^{(2)}-1}}{(n^{(2)}-1)!} dt_1^* & n^{(2)} > 0 \end{cases}
\end{aligned}$$

For  $p(\mathcal{T}^*, n | a, b, \boldsymbol{\theta})$  and  $p(\mathcal{T}^*, n = 0 | a, b, \boldsymbol{\theta})$  see (14) and (15) for partly observed pipes, respectively.

### 3.4. Implementation

Procedures for inference and prediction were implemented in R (R Development Core Team, 2012) that evokes a Fortran 95 implementation of the likelihood function. Samples of the posterior were obtained with the adaptive Metropolis sampler proposed by Vihola (2011) and implemented in the R-package *adaptMCMC* (Scheidegger, 2012).

## 4. Application examples

### 4.1. Simulated data

Two data sets are simulated to show that replacement causes biased parameter estimations if not considered appropriately.

The data sets have different sample sizes: they consist of the failure records of 100 and 1 000 pipes respectively. The failures were generated on the basis of the distribution assumptions made for the failure model in Section 3 (time to first break is Weibull distributed, time between the following breaks exponential). Replacement was simulated according to the replacement model of Section 3.2. The data sets were then compiled from failures within the observation period of the unreplaced pipes only. Data sets for a 60 years old system were simulated (for parameter values, see Table 1). The recording period was assumed to cover the last 10 years; 39 failures occurred in this period for the small data set and 403 for the large one.

The parameters are inferred with two models: a) the failure model of Section 3 while ignoring replacement, and b) the same failure model combined with the replacement model of Section 3.2. The prior distribution and the summarized posterior based on seven Monte Carlo Markov chains with 100 000 samples each are shown in Table 1 for both models.

The expected number of failures, calculated with equation (8) for unconditional predictions, as a function of the pipe age for newly built pipes, is shown in Figure 2. On Figure 2a) it is clearly seen that the failure frequency is underestimated if the replacement is not considered for the parameter estimation. With increasing sample size only the uncertainty becomes smaller, while the bias remains constant. Figure 2b) shows the result for parameters estimated with the replacement model. Although the uncertainties are greater than in Figure 2a), no systematic deviation is present.

Figure 3 shows all one- and two-dimensional marginals of the posterior parameter distribution. The one-dimensional marginals reveal the shape

of the posterior for each parameter. Dependencies are visible in the two-dimensional marginals. So is a clear correlation apparent between the replacement probability and the two scale parameters. This is important for real applications as it implies that an informative prior for the replacement probability would reduce the uncertainty of the scale parameters considerably.

Table 1: Parameters used for data simulation and estimated values. The posterior is based on seven independent MCMC chains with 100 000 samples each. The marginal posterior distributions are summarized with three numbers: first the mean followed by the 10% and 90% quantiles.

	Data generation	Prior U(l, u)	Posterior without replacement	Posterior with replacement
100 pipes				
$\theta_1$	2	(0.5, 4)	3.03 (2.11, 3.82)	3.05 (2.16, 3.82)
$\theta_2$	30	(1, 250)	36.75 (29.97, 44.90)	31.99 (24.79, 39.60)
$\theta_3$	15	(1, 250)	20.81 (15.43, 26.90)	15.20 (8.36, 22.18)
$\pi$	0.75	(0, 1)	–	0.73 (0.44, 0.96)
1 000 pipes				
$\theta_1$	2	(0.5, 4)	2.28(2.03, 2.54)	2.23 (1.97, 2.49)
$\theta_2$	30	(1, 250)	33.58(31.14, 36.13)	30.26 (25.76, 34.14)
$\theta_3$	15	(1, 250)	19.72(18.14, 21.37)	15.84 (11.29, 19.52)
$\pi$	0.75	(0, 1)	–	0.81 (0.59, 0.97)

#### 4.2. Real data

The second example illustrates one way of modifying the model to incorporate covariables. Covariables can represent quantitative (diameter, ...) or qualitative (material, construction period, ...) properties of an individual pipe. In this example, the influence of the construction period of ductile cast iron pipes is investigated.

Data on the water supply of Lausanne, Switzerland, is used. Instead of using the whole network, the focus is on one characteristic material only, namely, ductile cast iron (DI). It makes up the largest proportion (about 62 %) of the network and can be divided into two generations according to the manufacturing and laying periods: pipes with rather poor protection against outer corrosion (DI1), and pipes with improved corrosion protection (DI2). In Switzerland, DI1 pipes were commonly used until 1980 and were then succeeded by DI2. The proportion of DI2 pipes with recorded failures until the present is low (about 2 %), often making parameter inference a challenge. To reduce the influence of pipe length on the modeling of first and subsequent failures (Fuchs-Hanusch et al., 2012; Gangl, 2008; Poulton



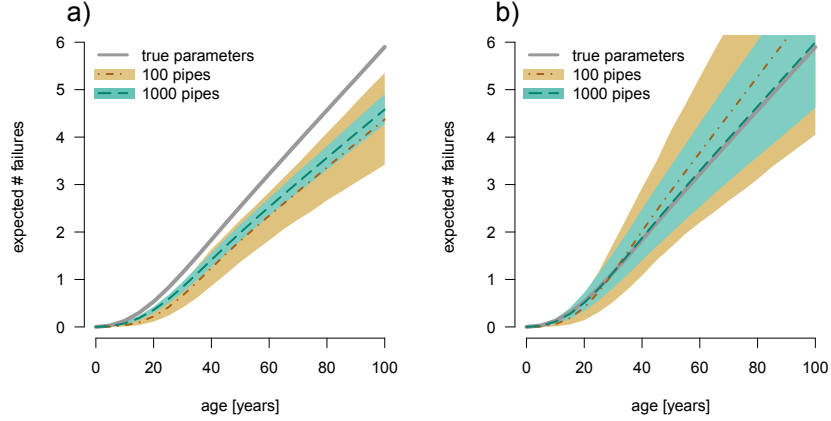


Figure 2: Estimated expected number of failures for a pipe as function of age. The estimations are based on simulated data with 100 and 1000 pipes respectively. For Figure a) the parameters were estimated without consideration of a replacement model. Figure b) shows the estimation with the replacement model. The dashed lines show the mean and the shaded areas indicate the 80%-credibility interval of the expected number of failures.

et al., 2007), the approach of Gangl (2008) was used. Gangl (2008) suggests forming 100 to 200 m long pipe units from neighboring pipes with equal diameters, materials and laying years. This is based on an analysis of the distances between subsequent failures, which are usually below 100 m and no longer than 200 m. Thus, short pipe segments were merged to 444 segments of DI1 pipes and to 2636 segments of DI2 pipes. As no spatial information was available, the merging was based on the construction year and diameter only. The average length of the merged segments is 143.1 m. Pipe failures were systematically recorded over ten years. Failures within the first year after installation were removed, because they are attributed to installation deficiencies and not to structural ageing. The record contains then 116 failures of DI1 pipes and 82 failures of DI2 pipes.

The qualitative information about the construction period is modeled with the help of indicator variables. For each pipe  $k$ , “individual” parameters  $\theta_k$  are computed as described in equation (6). In this case we choose  $f(\cdot)$  as

$$\theta_k = f(\mathbf{x}_k, \boldsymbol{\theta}, \boldsymbol{\gamma}) = (\theta_1, \gamma^{x_k} \theta_2, \gamma^{x_k} \theta_3, \theta_4)^T$$

where  $x_k$  is the indicator variable that equals one if pipe  $k$  is a DI2 pipe and zero otherwise. Accordingly,  $\theta_2$  and  $\theta_3$  can be interpreted as scale parameters for DI1 pipes and  $\gamma\theta_2$  and  $\gamma\theta_3$  as scales for DI2 pipes.

The same uniform priors as for the first example (Table 1) were used for

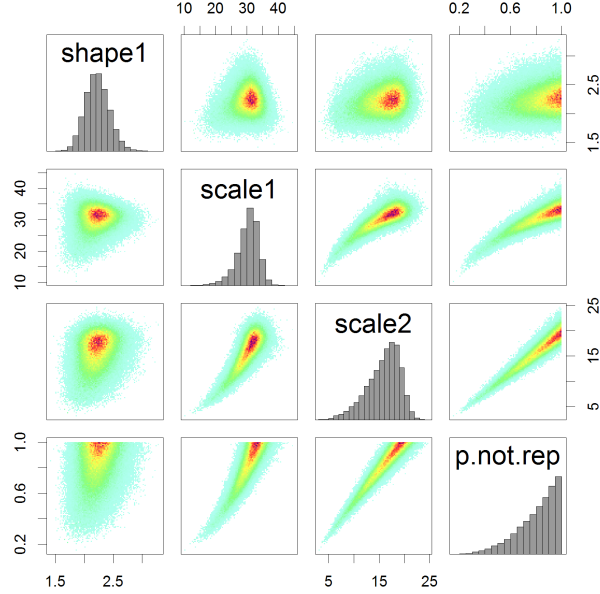


Figure 3: Marginals of the posterior distribution based on simulated data of 1 000 pipes and uniform priors. Warm colors denote regions with high probability density.

$\theta$ , and a gamma distribution for  $\gamma$  with mode one and a standard deviation of five.

As in the previous example we inferred the parameters without (Figure 4a) and with (Figure 4b) consideration of the replacement model. The resulting expected number of failures as function of pipe age is shown for both pipe generations in Figure 4a) and 4b). The corresponding posterior parameter distributions are summarized in Table 2.

As expected, the model predicts a higher failure rate if it corrects for pipe replacement. However, the predictions have larger uncertainties due to the additional parameter. The probability  $\pi$  that a pipe is not replaced after a failure is estimated within a reasonable range (see Table 2). Pipes of the first generation have a considerably higher risk of failures. This is in line with observations from practice in Switzerland; the lack of corrosion protection and the grounding of electrical appliances on the water lines until the 1990s led to increased corrosion and a large number of failures of DI1 pipes (Kappeler et al., 2010). Differences in the failure behavior of pipes

of different installation periods were frequently observed (e.g. Kleiner and Rajani, 1999; Mailhot et al., 2000). Because the data contain fewer DI1 pipes, the estimation is more uncertain than that of DI2.

Table 2: The resulting posterior parameter distributions for data from Lausanne inferred with and without replacement model. The posteriors are based on seven independent MCMC chains with 100 000 samples each. The posterior marginal distributions are summarized by three numbers: first the mean followed by the 10% and 90% quantiles.

	without repl. model	with repl. model
$\theta_1$	1.80 (1.52, 2.10)	1.80 (1.53, 2.09)
$\theta_2$	73.47 (63.82, 84.56)	65.88 (54.88, 77.35)
$\theta_3$	14.55 (11.98, 17.39)	13.05 (10.20, 16.08)
$\gamma$	1.80 (1.50, 2.13)	1.84 (1.52, 2.20)
$\pi$	—	0.89 (0.76, 0.99)

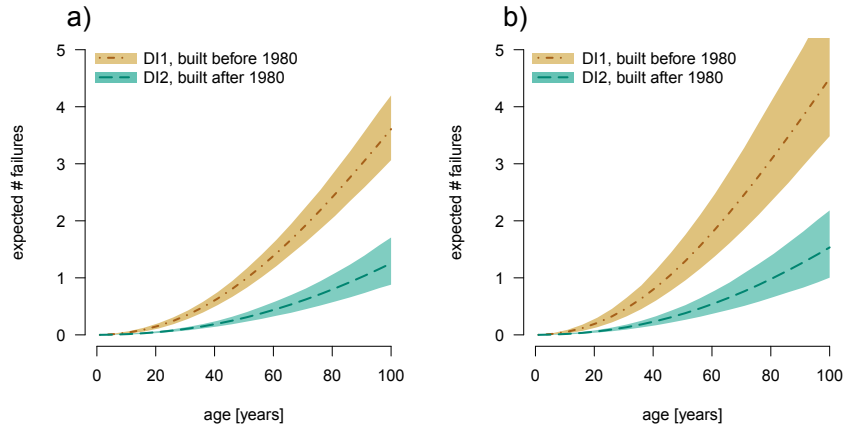


Figure 4: Expected number of failures for two generations of ductile iron pipes of the Lausanne water supply network. For Figure a) the parameters were estimated without consideration of a replacement model. Figure b) shows the estimation with the replacement model. The dashed lines show the mean and the shaded areas indicate the 80%-credibility interval of the expected number of failures.

## 5. Discussion

We demonstrated in Section 2 how a failure based pipe model can be extended by a replacement model. It is important to realize, that we did not propose a new pipe failure model. The intention was to present a procedure to avoid selective survival biases by adopting existing (or future) models. Therefore, we first introduced a generic notation, exceeding the “general

framework for water main break modeling” of Mailhot et al. (2003). Furthermore, we also derived all the equations for left truncated observations, including the likelihood function and the predictive distribution.

Our approach is limited to pipe models that (a) are failure based, (b) consider—if known—the failure history of a specific pipe for predictions, and (c) allow probabilistic statements about parameter and prediction uncertainty, i.e models based on a likelihood function. Condition (a) excludes life-span models (Herz, 1995; Scholten et al., 2013) as they intentionally lump failure behavior and the often consequential replacement together. Regression models (Boxall et al., 2007; Kleiner and Rajani, 1999) and simple proportional hazard (Cox) models (Carrión et al., 2010; Gangl, 2008) do not fulfill condition (b) and are therefore of limited use for prediction. Purely data driven algorithms (Giustolisi et al., 2006; Jafar et al., 2010) often do not fulfill condition (c) and therefore do not fit into a probabilistic framework. Models excluded here may nevertheless be influenced by the survival selection bias.

In Section 3 we exemplified our approach with a Weibull-exponential model. This model was chosen because it has been successfully applied (Mailhot et al., 2000) and has manageable complexity. However, as any parametric distribution the Weibull has some limitations. In particular, the hazard rate begins at zero (if shape parameter  $> 1$ ) and therefore installation failures and the probability of third party damages (typically caused by construction activities, Thomson and Wang, 2009) of young pipes cannot be modeled.

We demonstrated in the first example with artificially generated data that for datasets without historic data (containing information about replaced pipes) consideration of a replacement model is crucial for reliable predictions. In these cases, ignoring the replacement of pipes leads to a bias that cannot be reduced by increasing the sample size. This is especially significant in well maintained networks in which substantial replacements were made in the past. In these cases the failure rates are strongly underestimated. Missing data on replaced pipes is very common for many networks that we encounter here in Switzerland, and we suspect that it is equally common elsewhere (e.g. Le Gat, 2009).

The second example was based on a data set of ductile iron pipes from a real water supply network. The results show the expected behavior: (i) the first generation ductile pipes have a clearly higher failure rate, and (ii) the dataset with fewer observations shows larger uncertainties. This also illustrates a possible approach to extending the model by covariables. They

give the model more flexibility to fit the data. An alternative is to group pipes in homogeneous sets and fit a model independently to each of those. However, incorporating covariables has the advantage that interactions can be revealed and that, in total, fewer parameters need be inferred.

Inevitably, the likelihood function becomes more complicated if a replacement model is included, in particular if the data are left truncated (common for many European water networks). However, a replacement model may not be required, if data of replaced pipes are available. For some models the representation as counting process is more practicable. Instead of translating such models into the time domain, it might be more feasible to modify the presented approach accordingly. It is important to realize that the data availability and ultimately the data collection scheme determines the correct likelihood function. Therefore, the modeler must have a clear understanding of how the data have been collected and managed. This information is usually not directly evident from the data.

Replacement models do not have to represent decisions that are independent of the failure record, for example replacement decisions that are based purely on the age of the pipes. They do not lead to a bias in the parameter inference as terms representing independent decisions cancel out algebraically. Considering such replacement strategies in the replacement model would add unnecessary complexity.

The simplest possible replacement model was chosen for the examples, and it is certainly not suitable for all data sets. Therefore, the development of a more general replacement model could be helpful. For example replacement decisions may depend on the pipe age when failures occur or replacement strategies could change over time. Furthermore, in the context of Bayesian inference, how the prior distribution of the replacement model parameters is elicited optimally should be investigated. A good elicitation method to obtain informative priors from other utilities and/or expert elicitation would be critical for applications to small utilities with scarce data.

Selecting the most appropriate model remains a challenging task that cannot be automated. In many utilities, the data handling was guided by daily operation requirements and not with failure modeling in mind. This results in data sets that require the application of an adequate model, suited to the particular data management characteristics. This problem is not exclusive to drinking water pipes. For example, a similar approach to sewer deterioration models is developed by (Egger et al., submitted). Adapting the failure models to specific data sets is the only option, in the absence of more standardized data management strategies.

## 6. Conclusions

- Pipe failure models are important tools for the management of water distribution networks. The calibration of such models is often complicated by common practices of data handling. A frequent problem is that many available data sets contain records of pipes in service but not of replaced pipes. Calibration without explicit accounting for this practice can lead to considerably biased predictions.
- To correct for such biases, we propose an approach to modify the likelihood function of failure models. The key idea is to combine the likelihood function of the failure model with a probabilistic replacement model.
- As past replacement and data management practices are different for every network, a failure model must be adapted to a specific data set. The approach presented here is formulated generally and is therefore applicable to many—existing or future—pipe failure models and different replacement models.
- The concept is illustrated explicitly for a Weibull-exponential model in combination with a simple replacement model. Furthermore, we show how models can be extended to consider covariables.
- The code of the model used in the examples is freely available on request.

## 7. Acknowledgments

We thank Sébastien Apothéloz and Eauservice Lausanne for their cooperation and contribution of data.

## 8. References

### References

- Bernardo, J. M., Smith, A. F. M., 2000. Bayesian Theory, 1st Edition. Wiley.
- Boxall, J. B., O'Hagan, A., Pooladsaz, S., Saul, A. J., Unwin, D. M., 2007. Estimation of burst rates in water distribution mains. In: Proceedings of the Institution of Civil Engineers-Water Management. Vol. 160. pp. 73–82.

- Carrión, A., Solano, H., Gamiz, M. L., Debón, A., 2010. Evaluation of the reliability of a water supply network from right-censored and left-truncated break data. *Water resources management* 24 (12), 2917–2935.
- Dridi, L., Mailhot, A., Parizeau, M., Villeneuve, J., 2009. Multiobjective approach for pipe replacement based on Bayesian inference of break model parameters. *Journal of Water Resources Planning and Management* 135 (5), 344–354.
- Economou, T., Kapelan, Z., Bailey, T., 2008. A zero-inflated Bayesian model for the prediction of water pipe bursts. In: *Proceedings of Water Distribution Systems Analysis 2008*.
- Egger, C., Scheidegger, A., Reichert, P., Maurer, M., submitted. Sewer deterioration modeling with data lacking historical records.
- Eisenbeis, P., Røstum, J., Le Gat, Y., 1999. Statistical models for assessing the technical state of water networks – Some European experiences. In: *Proceedings of the AWWA Annual Conference*.
- Fuchs-Hanusch, D., Kornberger, B., Friedl, F., Scheucher, R., 2012. Whole of life cost calculations for water supply pipes. *Water Asset Management International* 8, 19–24.
- Gangl, G., 2008. Rehabilitationsplanung von Trinkwassernetzen. Ph.D. thesis, Technische Universität Graz.
- Giustolisi, O., Laucelli, D., Savic, D. A., 2006. Development of rehabilitation plans for water mains replacement considering risk and cost-benefit assessment. *Civil Engineering and Environmental Systems* 23 (3), 175–190.
- Gustafson, J. M., Clancy, D. V., 1999. Modeling the occurrence of breaks in cast iron water mains using methods of survival analysis. In: *Proc. 1999 Annual Conference of the AWWA, Chicago*.
- Harrell, F. E., 2001. *Regression Modeling Strategies with Applications to Linear Models, Logistic Regression, and Survival Analysis*. Springer, New York.
- Herz, R., 1995. Alterung und Erneuerung von Infrastrukturbeständen – ein Kohortenüberlebensmodell (Aging and renewal of infrastructure – a cohort survival model). *Jahrbuch für Regionalwissenschaft* 14–15, 5–29.

- Jafar, R., Shahrour, I., Juran, I., 2010. Application of artificial neural networks (ANN) to model the failure of urban water mains. *Mathematical and Computer Modelling* 51 (9-10), 1170–1180.
- Kappeler, J., Neubert, K., Woll, L., 2010. Benchmarking für Wasserversorgungen: Erkenntnisse für die Branche. *GWA* 90 (4), 305–314.
- Kleiner, Y., Rajani, B., 1999. Using limited data to assess future needs. *American Water Works Association* 91 (7), 47–61.
- Kleiner, Y., Rajani, B., 2001. Comprehensive review of structural deterioration of water mains: statistical models. *Urban Water* 3 (3), 131–150.
- Kleiner, Y., Rajani, B., et al., 2010. I-WARP: individual water main renewal planner. *Drinking Water Engineering and Science* 3, 71–77.
- Le Gat, Y., 2009. Une extension du processus de yule pour la modélisation stochastique des événements récurrents. Application aux défaillances de canalisations d’eau sous pression. Ph.D. thesis, Cemagref Bordeaux, Paris-tech.
- Mailhot, A., Pelletier, G., Noël, J., Villeneuve, J., 2000. Modeling the evolution of the structural state of water pipe networks with brief recorded pipe break histories: Methodology and application. *Water Resources Research* 36 (10), 3053–3062.
- Mailhot, A., Poulin, A., Villeneuve, J. P., 2003. Optimal replacement of water pipes. *Water resources research* 39 (5), 1136–1150.
- Pelletier, G., 2000. Impact du remplacement des conduites d’aqueduc sur le nombre annuel de bris. Ph.D. thesis, Université du Québec, Institut national de la recherche scientifique.
- Pelletier, G., Mailhot, A., Villeneuve, J. P., 2003. Modeling water pipe breaks – three case studies. *Journal of water resources planning and management* 129, 115.
- Poulton, M., Le Gat, Y., Brémond, B., 2007. The impact of pipe segment length on break predictions in water distribution systems. In: *Leading Edge conference on Strategic Asset Management (LESAM)*. Lisbon, Portugal.
- R Development Core Team, 2012. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna,



Austria.

URL <http://www.R-project.org/>

Renaud, E., Le Gat, Y., Poulton, M., 2011. Using a break prediction model for drinking water networks asset management: From research to practice. In: Leading Edge conference on Strategic Asset Management (LESAM). Mühlheim an der Ruhr, Germany.

Scheidegger, A., 2012. adaptMCMC: Implementation of a generic adaptive Monte Carlo Markov Chain sampler. R package version 1.0.3.

Scheidegger, A., Hug, T., Rieckermann, J., Maurer, M., 2011. Network condition simulator for benchmarking sewer deterioration models. *Water Research* 45 (16), 4983 – 4994.

Scholten, L., Scheidegger, A., Maurer, M., Reichert, P., 2013. Combining expert knowledge and local data for improved service life modeling of water supply networks. *Journal of Environmental Modelling & Software* 42, 1–16.

Thomson, J., Wang, L., 2009. Condition assessment of ferrous water transmission and distribution systems state of technology review report. Tech. rep., EPA, Office of Research and Development, National Risk Management Research Laboratory.

Vihola, M., 2011. Robust adaptive metropolis algorithm with coerced acceptance rate. *Statistics and Computing*.

URL <http://www.springerlink.com/content/672270222w79h431/>

Watson, T. G., Christian, C. D., Mason, A. J., Smith, M. H., Meyer, R., 2004. Bayesian-based pipe failure model. *Journal of Hydroinformatics* 6 (4), 259–264.