Research article

SEWER DETERIORATION MODELING WITH CONDITION DATA LACKING HISTORICAL DATA

C. Egger, A. Scheidegger, P.Reichert and M. Maurer

Published in Water Research, Volume 47, issue 17, 1 November 2013, Pages 6762-6779

SEWER DETERIORATION MODELING WITH CONDITION DATA LACKING HISTORICAL RECORDS

C. Egger [a,b*], A. Scheidegger [a], P.Reichert [a,c] and M. Maurer [a,b]

[a] *Eawag, Swiss Federal Institute of Aquatic Science and Technology, Überlandstrasse 133, P.O. Box 611, 8600 Dübendorf, Switzerland*

[b] *ETH Zürich, Department of Civil, Environmental and Geomatic Engineering, Institute of Environmental Engineering, Wolfgang-Pauli-Strasse 15, CH-8093 Zürich, Switzerland*

[c] *ETH Zürich, Department of Environmental System Science, Institute of Biogeochemistry and Pollutant Dynamics, CH-8092 Zürich, Switzerland*

Abstract

Accurate predictions of future conditions of sewer systems are needed for efficient rehabilitation planning. For this purpose, a range of sewer deterioration models has been proposed which can be improved by calibration with observed sewer condition data. However, if datasets lack historical records, calibration requires a combination of deterioration and sewer rehabilitation models, as the current state of the sewer network reflects the combined effect of both processes. Otherwise, physical sewer lifespans are overestimated as pipes in poor condition that were rehabilitated are no longer represented in the dataset. We therefore propose the combination of a sewer deterioration model with a simple rehabilitation model which can be calibrated with datasets lacking historical

[*] Corresponding author: Eawag, Überlandstrasse 133, P.O. Box 611, CH-8600 Dübendorf, Switzerland., phone: +41-58-765 5055, fax: +41-58-765 5389, E-mail address: christoph.egger@eawag.ch

information. We use Bayesian inference for parameter estimation due to the limited information content of the data and limited identifiability of the model parameters. A sensitivity analysis gives an insight into the model's robustness against the uncertainty of the prior. The analysis reveals that the model results are principally sensitive to the means of the priors of specific model parameters, which should therefore be elicited with care. The importance sampling technique applied for the sensitivity analysis permitted efficient implementation for regional sensitivity analysis with reasonable computational outlay. Application of the combined model with both simulated and real data shows that it effectively compensates for the bias induced by a lack of historical data. Thus, the novel approach makes it possible to calibrate sewer pipe deterioration models even when historical condition records are lacking. Since at least some prior knowledge of the model parameters is available, the strength of Bayesian inference is particularly evident in the case of small datasets.

Keywords: Deterioration model, rehabilitation model, data management, survival selection bias, likelihood, Bayesian inference

Nomenclature

Symbol	Description
a	Unit used for year
t	Pipe age
τ_k	Age of pipe k at the last available inspection
D_k	Age of pipe k when the dataset used for inference was lastly updated
R_k	Age of pipe k when rehabilitation started

$i = 1,\ldots,m$	Index of condition states (1=best, m= worst)
$c_{k\tau}$	Condition state (CS) of pipe k observed at its age τ
$C(t)$	Condition state of a pipe at its age t
T_i	(Random) pipe age at transition from CS i to CS $i+1$
t_i	Realization of T_i
T_0, t_0	Pipe age at construction
$S_i(t)$	Survival function, probability that a pipe section of age t is in CS i or better
λ_i	Rate of pipe replacement rehabilitation (a^{-1}) if a pipe is in condition state i
$\boldsymbol{\theta}$	Parameter vector of the likelihood functions $L_1(\boldsymbol{\theta})$ and $L_2(\boldsymbol{\theta})$
M_j	Hyperparameter, mean of the prior of θ_j
Σ_j	Hyperparameter, standard deviation of the prior of θ_j
N	Number of pipes of a sewer network with condition records used for model parameter inference
$N^{\exp}$	Number of pipes built per year
P_i^{reh}	Probability that a pipe in CS i is rehabilitated within one year

1 Introduction

Reliable forecasts of sewer network conditions facilitate proactive, far-sighted sewer asset management which in turn furthers an optimal balance between (future) expenses and system performance. Generally, sewer pipe conditions are related to different aspects of system performance. They constitute potential health and environmental hazards as a result of leaking pipes (Rutsch et al., 2008) as well as flooding and other hazards due to collapsing and malfunctioning pipes (Saegrov, 2005). Hence, given a specific, desired service level, knowledge about future sewer pipe conditions allows strategies for maintenance, surveillance and rehabilitation of sewers to be adjusted (Kleiner, 2001), and better estimates to be made for future operational and investment costs.

A wide range of models has been developed aiming to predict pipe deterioration. An overview of deterioration models differing in their mathematical approaches, requirements on data and mode of calibration is given by Kleiner and Rajani (2001), Rajani and Kleiner (2001) and Tran (2007). Tran (2007) classifies sewer deterioration models into (i) deterministic, (ii) artificial intelligence and (iii) statistical types, while Ana and Bauwens (2010) and Tran (2007) conclude that statistical models relying on a probabilistic relationship between model input and output data are the most feasible and most frequently applied approaches. This is basically due to the complexity of the processes responsible for sewer deterioration and the impossibility of making sufficient mechanistic parameters available over time and space. The most basic data needed to describe deterioration are condition records and corresponding pipe ages. Other explanatory variables, such as diameter, laying depth, etc., may also be relevant (Ana et al., 2009; Müller, 2002).

Unsuccessful model calibration is often attributed to a lack of suitable data for this task as discussed in Ana and Bauwens (2010) and Schmidt (2009). If datasets lack historical records, they represent the combined effect of deterioration and rehabilitation. Consequently, model calibration requires the use of a model that accounts for all of these processes. It was demonstrated by Scheidegger et al. (2011) that naively calibrating a deterioration model on the basis of such data leads to a significant overestimation of physical lifespans. Datasets without historical records are common in practice because operators are primarily interested in the current state of the sewer network and are unaware of the usefulness of data about its deterioration history. Lack of the following information hinders the calibration of sewer deterioration models:

(i) Condition records of sewer pipes which have been replaced by new ones. The corresponding records are discarded from the database and replaced by the new information.
(ii) Condition ratings of renovated or repaired sewer pipes from the time before such action. The repaired or renovated pipes are reassessed. On the basis of this reassessment, the condition rating prior to repair or renovation is overwritten by a new (and usually better) condition rating. Thus, certain records must be excluded from the analysis.
(iii) The renovation or repair of a pipe is not recorded. The pipe is consequently assigned a better state in the subsequent condition assessment.

When a sewer deterioration model is calibrated without accounting for rehabilitation, cases (i) and (ii) lead to a ‘survival selection bias’, since pipe rehabilitation depends on the pipe condition so that slow-aging pipes are overrepresented in the dataset. The data actually available suggests longer physical lifespans. Case (iii) causes a bias simply because the condition of a pipe was improved without recording the improvement. To our experience, sewer systems which underwent extensive rehabilitation in the past are in very good overall condition including older parts of the system. The datasets comprise often very few records of pipes in poor condition. This indicates that the sewer population is affected by the selective effect of rehabilitation. In the present paper, we address the lack of historical data according to cases (i) and (ii). Similar issues have been addressed in the context of failure prediction in water supply networks (Le Gat, 2009; Scheidegger et al., 2013). However, the proposed approaches are not readily transferable to the modeling of sewer pipe deterioration based on condition states.

We introduce a likelihood function for parameter estimation which takes account of the fact that the observations used for inference refer exclusively to pipes which have not been rehabilitated in the past. To derive this likelihood function, we combine a sewer deterioration model with a rehabilitation model. This allows us to derive probabilities of current states affected both by deterioration and rehabilitation. Generally, our principal approach could be applied to describe deterioration of other infrastructure if conditions are measured on an ordinal scale.

To account for the poor identifiability of such a combined model, we apply Bayesian inference for parameter estimation, which allows us to combine the data with prior knowledge of model parameters. This enables us to benefit from datasets despite their limited information content.

In this paper we focus on the systematic error created by the lack of historical data. For the sake of simplicity we ignore explanatory variables in our model. These improve model predictions for specific pipe cohorts (Ana et al., 2009). Neglecting explanatory variables may cause a bias for individual pipes but not on average for the whole pipe population. However, an extension of the model with explanatory variables can easily be implemented. The effect of input data uncertainty has been discussed elsewhere (Scheidegger and Maurer, 2012).

The remaining parts of this paper are organized as follows: the general concept of the combined deterioration - rehabilitation model is introduced in Section 2. In Section 3, an example is given with specific distribution assumptions. In Section 4, the model behavior is analyzed. Model results from applications with real data are presented in Section 1. Finally, we discuss the potential of our approach and give an outlook on further research needs on this topic.

2 Model description

In this section, a general formulation of the combined sewer deterioration and rehabilitation model is proposed. We use a statistical deterioration model due the probabilistic nature of pipe deterioration. The combined model is designed to be calibrated by using only the last available observed condition states of sewer pipes which have not been rehabilitated so far. After calibration, the deterioration model can be used on its own for predicting future deterioration of the sewer pipes as it would happen without further rehabilitation.

The condition of a pipe is described by an ordinal condition rating with m condition states (CS). These condition ratings are usually determined on the basis of closed circuit television (CCTV) surveys (DIN EN 13508-1, 2013) and a coding system such as specified in DIN EN 13508-2 (2011). New pipes are always assumed to be in the best CS, $C=1$. We denote the pipe age when the transition from CS $i \geq 1$ to CS $i+1$ occurs as T_i and define the age at construction as $T_0 = t_0 = 0$. The random variable T_i, $1 \leq i \leq m-1$, is characterized by the probability density function $p_i(t_i \mid t_1,...,t_{i-1},\boldsymbol{\theta})$ that is parameterized with the parameters included in $\boldsymbol{\theta}$. The deterioration model is completely defined by these probability density functions for all i and the parameter values $\boldsymbol{\theta}$. Figure 1 illustrates the relevant variables for the deterioration of a sewer pipe over the course of its lifespan introduced in Sections 2 and 3.

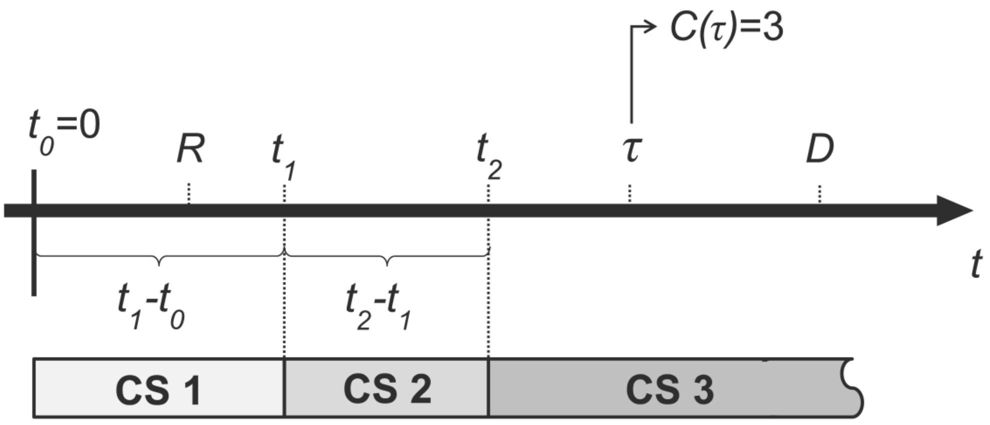

Figure 1. Time line of a deteriorating sewer pipe section introducing relevant variables of the proposed sewer deterioration model. The ages t_i are realizations of the random variable T_i.

To infer the parameters statistically, a likelihood function for the observed variables has to be formulated. This function is the probability of observing the data, given a model and its parameters. As the model is completely defined by the probability density functions of the random variables T_i, such a likelihood function can be derived from these densities. This will be done for the condition states at a given time in Section 2.1 for the case where all historical data for parameter inference is available or where no rehabilitation has taken place so far. We refer to this likelihood as the *unconditioned likelihood function*. In Section 2.2, this function is extended by a rehabilitation model to make it suitable for inference with data when historical condition records are lacking. We call this likelihood *conditioned likelihood function*.

2.1 Unconditioned likelihood function

Since only snapshot observations of pipe conditions are available, the ages T_i are not observable. The CS of a pipe at age t is $C(t)$. If a pipe is inspected at age τ, the data consists of the observed CS, c_τ. Therefore the likelihood for a single pipe k must be formulated as the probability $P(C(\tau_k)=c_{k\tau}\,|\,\boldsymbol{\theta})$. Assuming that each pipe deteriorates independently of others, the joint likelihood $L_1(\boldsymbol{\theta})$ of all N pipes becomes:

$$L_1(\boldsymbol{\theta})=\prod_{k=1}^{N}P(C(\tau_k)=c_{k\tau}\,|\,\boldsymbol{\theta}) \tag{1}$$

In the following, we derive the likelihood for a single pipe. The index k and the parameter vector $\boldsymbol{\theta}$ are then discarded if they do not have to be addressed explicitly.

We derive the probability $P(C(\tau)=c_\tau)$ by integrating the joint probability density $p(t_1,...,t_{m-1}\,|\,\boldsymbol{\theta})$ of the transition ages T_i, $1\le i\le m-1$,

$$p(t_1,...,t_{m-1}\,|\,\boldsymbol{\theta})=p_1(t_1\,|\,\boldsymbol{\theta})\,p_2(t_2\,|\,t_1,\boldsymbol{\theta})...p_{m-1}(t_{m-1}\,|\,t_1,...,t_{m-2},\boldsymbol{\theta}) \tag{2}$$

over adequate sets of transition ages. We distinguish three cases, depending on whether the pipe at age τ is in (1) the best condition state 1, (2) a condition state between 1 and m, or (3) the worst condition state m:

(1) $C(\tau)=1$: As $T_1>\tau$ the following applies:

$$P(C(\tau)=1)=P(T_1>\tau)=\int_{\tau}^{\infty} p_1(t_1)dt_1 \tag{3}$$

(2) $C(\tau)=i$ and $1<i<m$: As $T_{i-1}\leq\tau<T_i$ we can write:

$$\begin{aligned}&P(C(\tau)=i,1<i<m)=P(T_{i-1}\leq\tau<T_i)\\&=\int_0^{\tau}...\int_{t_{i-2}}^{\tau}\int_{\tau}^{\infty} p_1(t_1)...p_{i-1}(t_{i-1}\,|\,t_1,...,t_{i-2})\,p_i(t_i\,|\,t_1,...,t_{i-1})dt_i dt_{i-1}...dt_1\end{aligned} \tag{4}$$

(3) In case of $C(\tau)=m$ the following applies:

$$P(C(\tau)=m)=P(T_{m-1}\leq\tau)=1-\sum_{i=1}^{m-1}P(C(\tau)=i) \tag{5}$$

2.2 Conditioned likelihood function

For parameter inference, we are restricted to exclusively using the condition data of pipes $k,\ k=1...N$ which have not been rehabilitated before having reached age D_k. D_k is the age of a pipe k at which the asset dataset was lastly updated (in the following, the pipe index k is again discarded wherever possible). Sewer pipe rehabilitation is not independent of the CS. We therefore need to condition the likelihood introduced in the previous section by the fact that condition records used for inference refer exclusively to pipes which have not been rehabilitated before their age D, that is $P(C(\tau)=c_\tau\,|\,\mathrm{NR}(D),\boldsymbol{\theta})$. (The event that a pipe has not been rehabilitated before reaching age D is abbreviated by $\mathrm{NR}(D)$). According to Eq. (1), the joint likelihood $L_2(\boldsymbol{\theta})$ of all N pipes becomes:

$$L_2(\boldsymbol{\theta}) = \prod_{k=1}^{N} P(C(\tau_k) = c_{k\tau} \mid \mathrm{NR}(D_k), \boldsymbol{\theta}) \tag{6}$$

We rewrite the conditional probability $P(C(\tau) = c_\tau \mid \mathrm{NR}(D), \boldsymbol{\theta})$ using Bayes' theorem. This yields probabilities which permit easier interpretation:

$$P(C(\tau) = c_\tau \mid \mathrm{NR}(D), \boldsymbol{\theta}) = \frac{P(\mathrm{NR}(D) \mid C(\tau) = c_\tau, \boldsymbol{\theta}) \cdot P(C(\tau) = c_\tau \mid \boldsymbol{\theta})}{\sum_{j=1}^{m} P(\mathrm{NR}(D) \mid C(\tau) = j, \boldsymbol{\theta}) \cdot P(C(\tau) = j \mid \boldsymbol{\theta})} \tag{7}$$

$P(\mathrm{NR}(D) \mid C(\tau) = c_\tau, \boldsymbol{\theta})$ is the probability that a pipe has not been rehabilitated before D given $C(\tau) = c_\tau$ and $\boldsymbol{\theta}$. A model is required for this probability which we term the rehabilitation model. One possible rehabilitation model is introduced in Section 3.3.1. The probability $P(C(\tau) = c_\tau \mid \boldsymbol{\theta})$ is the likelihood as described by Eq. (1).

3 Model example for three condition states

3.1 Model assumptions

In the following, an example of the model is described on the basis of the principles introduced in the previous chapter.

For this example, the following assumptions are made:

(i) Three condition states are used ($m = 3$)

(ii) T_1 is Weibull distributed with the parameters shape α and scale β:

$$p_1(t_1) = \frac{\alpha}{\beta}\left(\frac{t_1}{\beta}\right)^{\alpha-1} e^{-(t_1/\beta)^\alpha} \tag{8}$$

(iii) The time span $T_2 - T_1$ a pipe spends in CS 2 does not depend on the age of the pipe and is further assumed to be exponentially distributed. Given these assumptions, T_2 for given t_1 is exponentially distributed as well with the single parameter scale μ:

$$p_2(t_2 \mid t_1) = \frac{1}{\mu} e^{-((t_2 - t_1)/\mu)} \tag{9}$$

(iv) $R_k \geq T_0$ is the age of pipe k at which rehabilitation was established.

(v) The applied replacement model introduced in Section 3.3.1 is parameterized by age-invariant, condition-state-dependent rehabilitation rates λ_i.

We would like to emphasize that we consider the combination of the processes deterioration and rehabilitation in a single model designed to improve the prediction of sewer pipe deterioration as the main innovation of our approach, and not the individual models themselves. Various stochastic deterioration models exist which describe the (random) variable T_i (Baur and Herz, 2002; Micevski et al., 2002; Mishalani and Madanat, 2002). They mainly differ in the distribution of T_i and may comprise further features such as the consideration of additional factors such as the pipe diameter and material. The suggested combinations of Weibull and exponential distributions have been used successfully to describe stepwise survival processes as done by (Mailhot et al., 2000) for modeling subsequent breaks of water supply pipes. We abstained from using a simpler approach such as age or time invariant transition probabilities. These are not appropriate to describe the aging of sewer pipes as discussed in Trujillo Alvarez (1995). This increases the mathematical efforts, but results in an applicable model.

Our principal concept is flexible, and individual aspects of the combined model as suggested above can therefore be altered or substituted by other models in case they prove to be more suitable for a given case.

3.2 Unconditioned likelihood function

Having framed the model as outlined above, we can now specify the equations needed to calculate the unconditioned likelihood according to Eq. (1) with the parameters $\boldsymbol{\theta} = (\alpha, \beta, \mu)^T$, which in turn is required to calculate the conditioned likelihood specified by Eq. (6).

The probability $P(C(\tau) = c_\tau \mid \boldsymbol{\theta})$ is calculated using Eqs. (3-5) and Eqs. (8-9) (the parameter vector $\boldsymbol{\theta}$ is discarded):

$$P(C(\tau)=1) = P(T_1 > \tau) = \int_\tau^\infty p_1(t_1)dt_1 = e^{-(\tau/\beta)^\alpha} \tag{10}$$

$$\begin{aligned} P(C(\tau)=2) &= P(T_1 \le \tau < T_2) = \int_0^\tau p_1(t_1) \int_\tau^\infty p_2(t_2 \mid t_1) dt_2 dt_1 \\ &= \int_0^\tau \frac{\alpha}{\beta}\left(\frac{t_1}{\beta}\right)^{\alpha-1} e^{-(t_1/\beta)^\alpha - ((\tau - t_1)/\mu)} dt_1 \end{aligned} \tag{11}$$

$$P(C(\tau)=3) = 1 - P(C(\tau)=1) - P(C(\tau)=2) \tag{12}$$

3.3 Conditioned likelihood function

The equations derived in Section 3.2 and the rehabilitation model proposed below enables us to calculate the conditioned likelihood $L_2(\boldsymbol{\theta})$ according to Eq. (6) with the extended parameter vector $\boldsymbol{\theta} = (\alpha, \beta, \mu, \lambda_1, \lambda_2, \lambda_3)^T$.

3.3.1 Rehabilitation model

Pipe rehabilitation depends on various factors such as (i) condition state (ii) pipe age, (iii) lack of hydraulic capacity, (iv) coordinated rehabilitation projects involving other infrastructure than sewers and (v) budget restraints. However, we assume a simple model describing rehabilitation exclusively dependent on the CS which we suppose being the major driver for rehabilitation. Further factors could be included but would be intricate to identify based on the available information. Specifically, this model describes the CS-dependent probability P_i^{reh} that a pipe is rehabilitated within one year once rehabilitation started. We further assume that this probability is age-invariant. From this probability we derive a constant rehabilitation rate λ_i for each CS *i* using the following equation:

$$\lambda_i = -\log\left(1 - P_i^{reh}\right)\mathrm{a}^{-1} \tag{13}$$

Formally, the rehabilitation model describes the functional survival of the pipes, i.e. the probability that a pipe with age t has not been rehabilitated. Therefore, the rehabilitation rate can be interpreted as a hazard rate. In a first step, we consider the probability $P(\mathrm{NR}(\tau) \mid C(\tau) = c_\tau, \boldsymbol{\theta})$ that a pipe has not been rehabilitated before age τ given $C(\tau) = c_\tau$ and $\boldsymbol{\theta}$. This probability depends on the pipe ages T_i at which transitions occurred, and on τ. If we knew the ages $\{T_1, T_2\}$, we could specify the rehabilitation rate $\lambda(t)$ for this pipe at age t, $0 \le t \le \tau$ as:

$$\lambda(t \mid T_1, T_2, R) = \begin{cases} 0 & t < R \\ \lambda_1 & t \ge R, t < T_1 \\ \lambda_2 & t \ge R, T_1 \le t < T_2 \\ \lambda_3 & t \ge R, T_2 \le t \end{cases} \tag{14}$$

The probability $P(\mathrm{NR}(\tau) \mid \tau, T_1, T_2, R, \boldsymbol{\theta}_{4:6})$ that a pipe has not been rehabilitated before τ given (i) τ, (ii) the ages $\{T_1, T_2\}$, (iii) R and (iv) $\boldsymbol{\theta}_{4:6} = (\lambda_1, \lambda_2, \lambda_3)^T$ can then be calculated:

$$\begin{aligned} P(\mathrm{NR}(\tau) \mid \tau, T_1, T_2, R, \boldsymbol{\theta}_{4:6}) &= e^{-\int_R^\tau \lambda(t \mid T_1, T_2, R)\,dt} \\ &= \begin{cases} 1 & \tau \le R \\ e^{-\lambda_1(\tau - R)} & R \le \tau \le T_1 \\ e^{-\lambda_2(\tau - R)} & T_1 \le R \le \tau \le T_2 \\ e^{-\lambda_1(t_1 - R) - \lambda_2(\tau - t_1)} & R \le T_1 \le \tau \le T_2 \\ e^{-\lambda_3(\tau - R)} & T_1 \le T_2 \le R \le \tau \\ e^{-\lambda_2(t_2 - R) - \lambda_3(\tau - t_2)} & T_1 \le R \le T_2 \le \tau \\ e^{-\lambda_1(t_1 - R) - \lambda_2(t_2 - t_1) - \lambda_3(\tau - t_2)} & R \le T_1 \le T_2 \le \tau \end{cases} \end{aligned} \tag{15}$$

Since the ages $\{T_1, T_2\}$ are unknown and only $C(\tau)$ is given, we must multiply Eq. (15) by Eq. (2), the joint probability density of the ages T_i and integrate between the bounds of integration given in Table 1 for specific observed CS $C(\tau) = c_\tau$. To condition on $C(\tau) = c_\tau$ we divide by $P(C(\tau) = c_\tau)$ or respective the joint probability $p(t_1, t_2 \mid \boldsymbol{\theta})$, see Eqs.(2, 8-9), integrated between the bounds given in Table 1. This yields $P(\mathrm{NR}(\tau) \mid \mathrm{C}(\tau) = \mathrm{c}_\tau, \tau, R, \boldsymbol{\theta})$:

$$P(\mathrm{NR}(\tau) \mid \mathrm{C}(\tau)=\mathrm{c}_\tau, \tau, R, \boldsymbol{\theta}) = \frac{\int_{t_1^u}^{t_1^o}\int_{t_2^u}^{t_2^o} P(\mathrm{NR}(\tau) \mid \tau, T_1, T_2, R, \boldsymbol{\theta}_{4:6})\, p(t_1, t_2, \boldsymbol{\theta})\, dt_2\, dt_1}{\int_{t_1^u}^{t_1^o}\int_{t_2^u}^{t_2^o} p(t_1, t_2, \boldsymbol{\theta})\, dt_2\, dt_1} \tag{16}$$

Table 1. Bounds of integration for $c_\tau = \{1,2,3\}$ to be used for the integrals of Eq. (16).

c_τ	t_1^u	t_1^o	t_2^u	t_2^o
1	τ	∞	t_1	∞
2	0	τ	τ	∞
3	0	τ	t_1	τ

In the case of $R \geq \tau$, this probability is independent of $C(\tau)$ and we can write:

$$P(\mathrm{NR}(\tau) \mid R \geq \tau, \boldsymbol{\theta}) = 1 \tag{17}$$

In the case of $R \leq \tau$ and $C_{obs,\tau} = 1$, we obtain the following expression:

$$P(\mathrm{NR}(\tau) \mid C(\tau) = 1, R \leq \tau, \boldsymbol{\theta}) = e^{-\lambda_1(\tau - R)} \tag{18}$$

If $R \leq \tau$ and $i = \{2,3\}$, we need further to consider that the transition(s) from CS $i-1$ to CS i have taken place either before or after a pipe has reached age R, see Eq. (15).

The distinction between these cases is made in the numerators of Eq. (19-20) below. For example, the first summand in the numerator of Eq. (19) gives the joint probability $P(\mathrm{NR}(\tau),\mathrm{C}(\tau)=2\,|\,\tau,R\leq\tau,T_1\leq R,\boldsymbol{\theta})$ that a pipe has not been rehabilitated before reaching age τ, and $C(\tau)=2$ given that the first transition occurred when the pipe reached age R or before. Similarly, the second summand gives the same probability given that the first transition occurred when the pipe section reached age R or later.

$$P(\mathrm{NR}(\tau)\,|\,C(\tau)=2,R\leq\tau,\boldsymbol{\theta})=\frac{\int_0^R\int_\tau^\infty e^{-\lambda_2(\tau-R)}p(t_1,t_2\,|\,\boldsymbol{\theta})dt_2dt_1+\int_R^\tau\int_\tau^\infty e^{-\lambda_1(t_1-R)-\lambda_2(\tau-t_1)}p(t_1,t_2\,|\,\boldsymbol{\theta})dt_2dt_1}{\int_0^\tau\int_\tau^\infty p(t_1,t_2\,|\,\boldsymbol{\theta})dt_2dt_1} \quad (19)$$

$$P(\mathrm{NR}(\tau)\,|\,C(\tau)=3,R\leq\tau,\boldsymbol{\theta}) \quad (20)$$
$$=\frac{e^{-\lambda_3(\tau-R)}\int_0^R\int_{t_1}^R p(t_1,t_2\,|\,\boldsymbol{\theta})dt_2dt_1+\int_0^R\int_R^\tau e^{-\lambda_2(t_2-R)-\lambda_3(\tau-t_2)}p(t_1,t_2\,|\,\boldsymbol{\theta})dt_2dt_1+\int_R^\tau\int_{t_1}^\tau e^{-\lambda_1(t_1-R)-\lambda_2(t_2-t_1)-\lambda_3(\tau-t_2)}p(t_1,t_2\,|\,\boldsymbol{\theta})dt_2dt_1}{\int_0^\tau\int_{t_1}^\tau p(t_1,t_2\,|\,\boldsymbol{\theta})dt_2dt_1}$$

So far, we have described the probability that a pipe was not rehabilitated before age τ given the observed CS at that age. However, replacement of pipes and hence discarding of further pipe records may continue beyond pipe age τ until the pipes reach age D. Thus, we need to consider the probability $P(\mathrm{NR}(D)\,|\,C(\tau)=c_\tau,\boldsymbol{\theta})$ that a pipe was not rehabilitated before age D given that $C(\tau)=c_\tau$ and $\boldsymbol{\theta}$. To accommodate this fact, we assume that $\lambda(t)=\lambda(\tau),t>\tau$. This assumption implies that decisions on pipe rehabilitation are made on the basis of the observed condition state c_τ. Thus, possible changes in condition states taking place at ages t, $\tau<t\leq D$ do not affect the probability of pipe rehabilitation. This assumption allows us to calculate the probability that a pipe was not rehabilitated in the interval between age τ and D by the following

expression (The event that a pipe has not been rehabilitated in the interval between age τ and D is abbreviated by $\mathrm{NRI}(\tau, D)$):

$$P(\mathrm{NRI}(\tau,\mathrm{D}) \mid C(\tau)=c_\tau,\tau,R,D,\boldsymbol{\theta}) = \begin{cases} 1 & R \geq D \\ e^{-\int_R^D \lambda_i(t)dt} = e^{-\lambda_i(D-R)} & \tau \leq R \leq D \\ e^{-\int_\tau^D \lambda_i(t)dt} = e^{-\lambda_i(D-\tau)} & R \leq \tau \end{cases} \quad (21)$$

Finally, we calculate the probability $P(\mathrm{NR}(D) \mid C(\tau)=c_\tau,\tau,R,D,\boldsymbol{\theta})$ by multiplying Eqs. (17-20) by Eq. (21) respectively, which leads to the following general expression:

$$\begin{aligned} &P(\mathrm{NR}(D) \mid C(\tau)=c_\tau,\tau,R,D,\boldsymbol{\theta}) = \\ &P(\mathrm{NR}(\tau) \mid C(\tau)=c_\tau,\tau,R,\boldsymbol{\theta}) \cdot P(\mathrm{NRI}(\tau,D) \mid C(\tau)=c_\tau,\tau,R,D,\boldsymbol{\theta}) \end{aligned} \quad (22)$$

Further formulations of Eq. (22) for $R \leq \tau$ and $c_\tau = \{2,3\}$ are given by Eqs. (A.1–A.2) in the Appendix.

3.4 Model calibration

To estimate the model parameters, we use Bayesian inference (Bolstad, 2007; Congdon, 2006; Gelman et al., 2004). This enables us to include additional (prior) knowledge and therefore handle datasets of limited size and strength that lead to poor model identifiability with frequentist inference methods. Prior knowledge may be obtained by eliciting experts or from the results of previous studies. As Bayesian inference allows sequential updating of the posterior when new data becomes available, posteriors from precedent inferences are an optimal choice for the prior.

Prior knowledge is described by a probability density function of the parameters $\boldsymbol{\theta}$. Generally, we assume that the prior for these parameters is distributed independently.

As there is no analytical form of the posterior distribution, numerical Monte Carlo Markov Chain techniques are applied for inference. These techniques enable us to take samples from the posterior. Statistical properties of the posterior distribution are then approximated from these samples. We used the algorithm of Vihola (2012) and the respective implementation by Scheidegger (2012) in R (R Core Team, 2012).

4 Model behavior analysis using synthetic data

In this section we analyze the model behavior to gain an understanding of the identifiability of the model parameters using synthetic data from the network condition simulator (NetCoS) (Scheidegger et al., 2011) (Section 4.1). We further address the sensitivity of the model with respect to changes in the specification of the prior (Section 4.2).

4.1 Model test using NetCoS

NetCoS can be used to benchmark different deterioration models under specific data management strategies that result in different data availabilities. We did this with the proposed model and considered replacement as the exclusive rehabilitation measure. This does not provide a ‘proof’ of model goodness for real case applications which may involve highly variable deterioration and rehabilitation processes. However, it enables us to analyze the identifiability of the model.

A synthetic dataset of a sewer network is generated by NetCoS using the parameters listed in Table 2. The first sewer pipes were installed 100 years ago and the network has been extended by $N^{\exp} = 20$ pipes annually up to the present. The simulation resulted in 2000 ‘active’ sewer pipes at the end of the simulation period and 1253 replaced pipes within the simulated period. Pipe replacement was introduced 73 years after the first pipes were laid.

Only data of the 'active' sewer pipes are used for the inference, i.e. all replaced pipes are discarded from the dataset. The priors of the parameters $\boldsymbol{\theta}$ are independently log-normally distributed with means $\mathbf{M}$ and standard deviations $\boldsymbol{\Sigma}$, see Table 2. The derivation of the prior of $\boldsymbol{\theta}_{1:3}$ is outlined in the first paragraph of Section 5.2. The prior of $\boldsymbol{\theta}_{4:6}$ is derived from similar data of a real sewer network as outlined in the second paragraph of Section 5.2.

Table 2. Predefined parameter set used by NetCoS for data generation, as well as means $\mathbf{M}$ and standard deviations $\boldsymbol{\Sigma}$ of the prior of the log-normally distributed model parameters $\boldsymbol{\theta} = (\alpha, \beta, \mu, \lambda_1, \lambda_2, \lambda_3)^T$. The predefined values used for data generation correspond to the mode of the prior.

Parameter	Predefined values used for data generation	j	M_j	Σ_j
α	3.1	1	3.69	1.31
β	56.8	2	60.3	12.2
μ	15.6	3	23.6	13.3
λ_1	0.011	4	0.016	0.008
λ_2	0.068	5	0.095	0.048
λ_3	0.160	6	0.224	0.112
$N^{\exp}$	20		-	-

The results of the inferences are shown in Figure 2. The reduction of the variance of the parameter distribution between the prior and posterior reflects the knowledge gained from the inference. The poor identifiability of λ_1 is reflected by the almost identical shapes of the prior and posterior marginals. This can be explained by the low importance of the parameter, as discussed in Section 4.2.

Figure 3 shows distinct correlations between the model parameters β, λ_2; β, λ_3 and μ, λ_3. The correlations suggest that faster deterioration of the pipes can be compensated by higher rehabilitation activity with regard to pipes in CS 2 and 3. Further insight into the importance and correlation of the parameters is gained by the sensitivity analysis discussed in Section 4.2.

In Figure 4, the results of the inference are showed as survival functions expressing the probability of a pipe being in a certain condition depending on its age. In the case of m condition states, the probability that a pipe is in CS $C(t) \leq i$ is described by the survival function $S_i(t), i = 1...(m-1)$, see Eqs. A.3 and A.4 in the Appendix. On average, the model can identify the survival function parameters, as indicated by the almost identical predefined and estimated mean survival functions shown in Figure 4. The results also indicate that considerably larger uncertainties are associated with $S_2(t)$ compared to $S_1(t)$. This fact can be explained by (i) the rather uncertain prior of μ,(ii) the great importance of λ_3 and (iii) the correlation between λ_3 and μ.

In the supplementary material, results are provided from simulations using the same data but the unconditioned likelihood according to Eq. (1). These results illustrate the underestimation of pipe deterioration if the rehabilitation process is neglected.

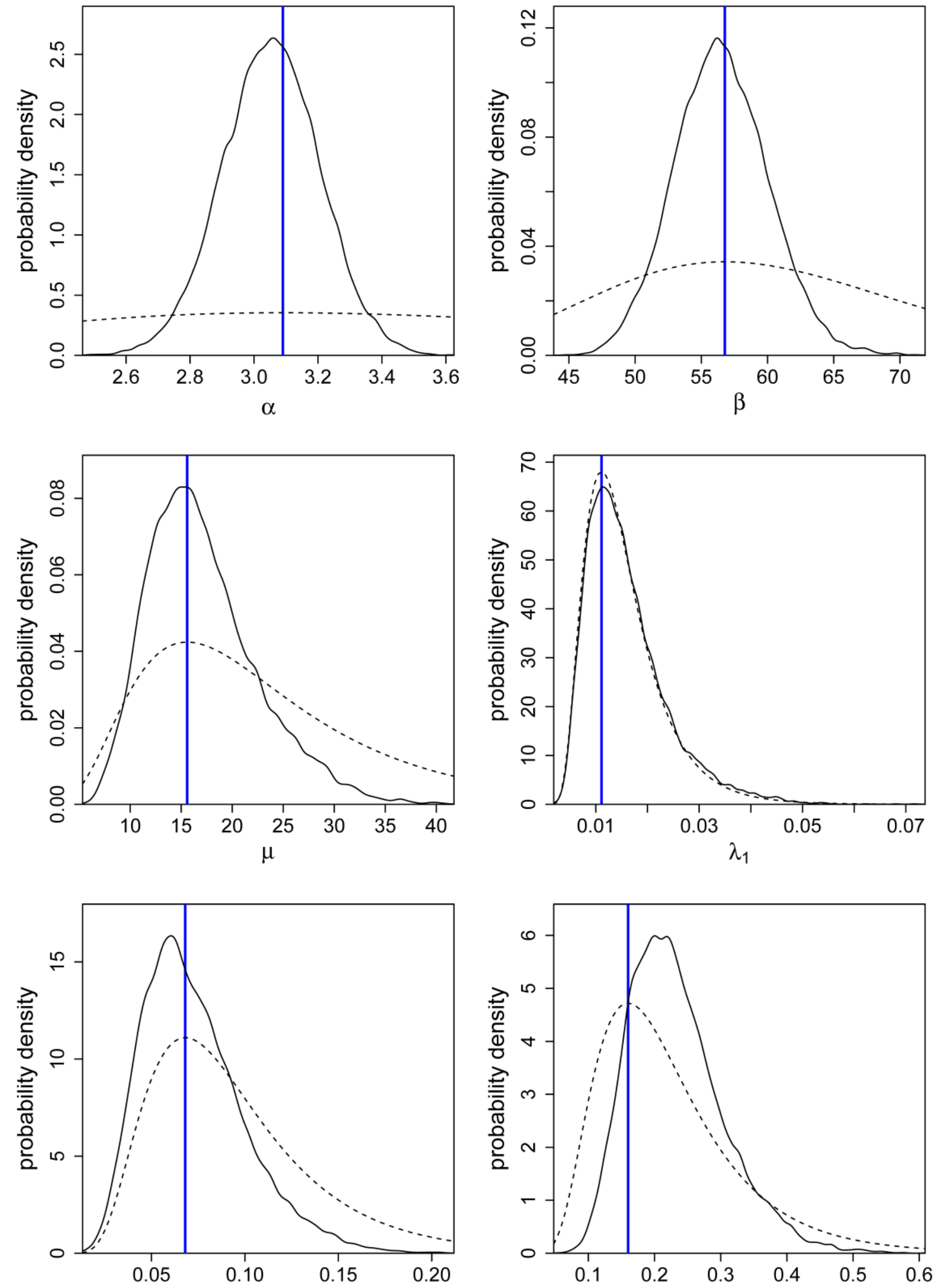

Figure 2. Prior (dashed lines) and posterior (solid lines) marginal distributions of the model parameters $\boldsymbol{\theta}$. The vertical lines indicate the predefined parameter values used for data generation with NetCoS.

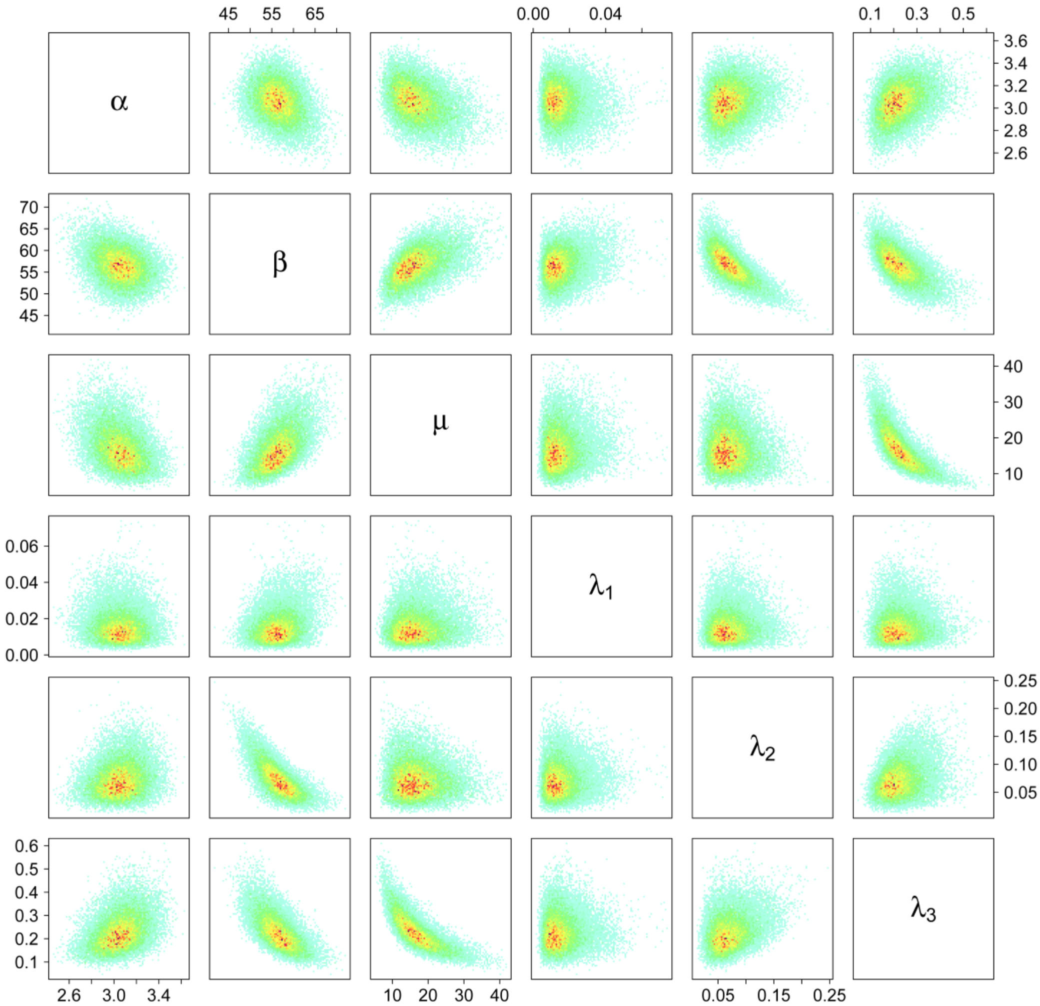

Figure 3. Scatter plot matrix of parameters sampled from the posterior by MCMC. All combinations of two-dimensional marginal distributions are given illustrating the correlations between the parameters. Warm colors denote regions of high probability density.

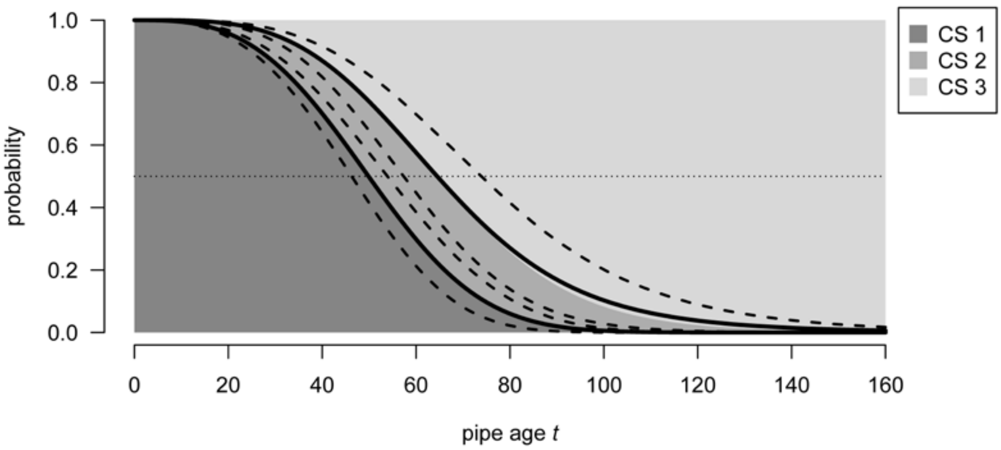

Figure 4. Predefined and estimated survival functions. The conditioned likelihood was used for inference. The gray shaded areas indicate the predefined survival functions based on the parameters in Table 2 used for data generation in NetCoS. The solid lines are the means of the estimated survival functions, and the dashed lines are the 10 % and 90 % quantiles based on the posterior distribution of $\boldsymbol{\theta}$. Good convergence is obtained if the conditioned likelihood is used.

4.2 Model sensitivity to the prior

A common way of obtaining priors is to elicit them from experts (O'Hagan et al., 2006). Eliciting probability distributions is demanding and may be biased for a range of reasons (Tversky and Kahneman, 1974). There is also concern about the problem of specifying probability distributions precisely based on subjective beliefs (Rinderknecht et al., 2012). Given the evidence of the uncertainty of our prior and its insufficient description, we are concerned about the sensitivity of model outputs to the specification of prior probability distributions. Specifically, we are interested in identifying the most influential parameters specifying the location and variances of the prior distributions on model predictions. That indicates the parameters for which prior elicitation is critical.

4.2.1 Methods

Prior knowledge of the model parameters $\boldsymbol{\theta}$ is described by (independent) probability distributions with mean $\mathbf{M}$ and standard deviation $\boldsymbol{\Sigma}$. The goal is to analyze the change in model output resulting from a change in the hyperparameters $\mathbf{M}$ and $\boldsymbol{\Sigma}$. Having specified adequate ranges for $\mathbf{M}$ and $\boldsymbol{\Sigma}$, we draw a sample of them, assuming that they are independent and uniformly distributed. Each sample represents one possible prior. We perform inferences with each of the generated priors in combination with one specific dataset and the likelihood as specified by Eq. (6), resulting in posterior distributions each associated with one prior. We calculate the specified model outputs from each of the posterior distributions. This yields samples of influencing parameters and model outputs.

We use variance-based techniques for regional sensitivity analysis (Saltelli et al., 2000) to explore the impact of changes in $\mathbf{M}$ and $\boldsymbol{\Sigma}$ on the model output derived from the properties of the posteriors. Different smoothing algorithms exist, allowing variance-based sensitivity coefficients to be estimated on the basis of samples of influencing parameters and corresponding model outputs (Gasser et al., 1991; Seifert and Gasser, 1996, 2000). We used Kernel Regression Smoothing with an Adaptive Plug-in Bandwidth algorithm implemented by Herrmann and Maechler (2011) in the statistics and graphics language and environment R (R Core Team, 2012).

Using MCMC for inference is computationally demanding even for small datasets. To overcome this limitation, we apply importance sampling to extend the MCMC-based results for one prior to the others (Robert and Casella, 2010). This permits us to efficiently approximate posterior distributions for extensive realizations of priors on the basis of one or a few samples drawn from posteriors by MCMC with different priors. We calculate the *effective sample size* (ESS) for every posterior distribution generated by importance sampling (Robert and Casella, 2010). The ESS is a useful measure for examining the worth of the samples generated by this technique. In cases of unacceptably low ESS, the respective samples were substituted by samples generated by MCMC.

As the model output, we focus on the ages at which 50% of the pipes are transferred from CS 1 to 2 and 2 to 3, respectively. Another relevant property is the standard deviation of the pipe ages at these transitions. As the inference yields the distribution of $\boldsymbol{\theta}$, the model output also has a distribution. Therefore, we consider the mean and the standard deviation of

(i) the age at which 50% of the pipes pass from CS 1 to CS 2 (median pipe age when CS changes from 1 to 2)
(ii) the age at which 50% of the pipes pass from CS 2 to CS 3 (median pipe age when CS changes from 2 to 3).

We further consider the mean of

(iii) the standard deviation of the pipe age when CS changes from 1 to 2
(iv) the standard deviation of the pipe age when CS changes from 2 to 3.

4.2.2 Results

The sensitivity analysis is performed by using the same synthetic dataset used in the analysis discussed in Section 4.1. We refer our analysis to the prior specified in Table 2.

The possible variations of the hyperparameters $\mathbf{M}$ and $\boldsymbol{\Sigma}$ are defined by the ranges shown in Table 3, which correspond to a deviation of +/- 50 % from the hyperparameters specifying the given prior. Our analysis is based on 10,000 randomly sampled priors, given that $\mathbf{M}$ and $\boldsymbol{\Sigma}$ are independently and uniformly distributed within the intervals.

Table 3. Lower and upper limits a, b of the hyperparameters $\mathbf{M}$ and $\mathbf{\Sigma}$. See also Table 2.

Parameter	Hyperparameter	Lower limit a	Upper limit b
α	M_1	1.84	5.53
	Σ_1	0.65	1.96
β	M_2	30.2	90.5
	Σ_2	6.12	18.4
μ	M_3	11.8	35.4
	Σ_3	6.66	20.0
λ_1	M_4	0.008	0.023
	Σ_4	0.004	0.012
λ_2	M_5	0.048	0.142
	Σ_5	0.024	0.071
λ_3	M_6	0.111	0.335
	Σ_6	0.056	0.168

The results in terms of relative sensitivity coefficients relating to the model outcomes specified above are summarized in Table 4. In general, high sensitivity coefficients associated with specific hyperparameters reflect either (i) high importance of the corresponding parameters, (ii) low identifiability of the corresponding parameters, or (iii) a combination of both. In turn, low sensitivity coefficients indicate low importance and/ or good identifiability. From the results shown in Table 4 we can see that the hyperparameters defining the locations of the model parameters β, μ, λ_2 and λ_3 have relative sensitivities higher than 0.1 and can be labeled as important. Furthermore, the standard deviation of the median pipe ages when CS changes from 1 to 2 is also sensitive to the standard deviation of the prior of λ_2 (S_5), and similarly, the standard deviation of the median pipe age when CS changes from 2 to 3 is sensitive to the standard deviation of the prior of λ_3(S_6). From this sensitivity analysis we can conclude that prior knowledge of the means of the model parameters β, μ, λ_2 and λ_3 as well as the standard deviations of λ_2 and λ_3 have a decisive influence on the outcome of the parameter inference. All other hyperparameters are of minor sensitivity and importance.

Table 4. Relative sensitivities of model results to the hyperparameters $\mathbf{M}$ and $\mathbf{\Sigma}$ of the prior distributions of the model parameters $\boldsymbol{\theta} = (\alpha, \beta, \mu, \lambda_1, \lambda_2, \lambda_3)^T$. Relative sensitivities >0.1 are highlighted.

Model result	Relative sensitivities											
	α		β		μ		λ_1		λ_2		λ_3	
	M_1	Σ_1	M_2	Σ_2	M_3	Σ_3	M_4	Σ_4	M_5	Σ_5	M_6	Σ_6
Mean of the median pipe age when CS changes from 1 to 2	0.003	0.001	0.362	0.040	0.067	0.010	0.004	0.002	0.247	0.029	0.049	0.001
Standard deviation of the median pipe age when CS changes from 1 to 2	0.017	0.003	0.087	0.001	0.002	0.008	0.006	0.056	0.138	0.155	0.073	0.065
Mean standard deviation of the pipe age when CS changes from 1 to 2	0.027	0.005	0.333	0.045	0.085	0.013	0.003	0.002	0.194	0.024	0.060	0.001
Mean of the median pipe age when CS changes from 2 to 3	0.008	0.001	0.315	0.035	0.216	0.034	0.002	0.001	0.096	0.011	0.113	0.001
Standard deviation of the median pipe age when CS changes from 2 to 3	0.019	0.004	0.110	0.002	0.012	0.042	0.004	0.018	0.052	0.037	0.350	0.132
Mean standard deviation of the pipe age when CS changes from 2 to 3	0.017	0.003	0.245	0.029	0.312	0.050	0.001	0.001	0.029	0.003	0.147	0.001

5 Model application

5.1 Data

The data for the practical application discussed in the present chapter derives from a utility in which systematic, extensive rehabilitation of the sewer network was introduced in the mid-eighties and has continued to the present. We used a subset of the data comprising more than 6700 pipes made of spun concrete with diameters of 800 mm or less. For this group of pipes, only replacement was applied as a rehabilitation measure. The utility aims to replace pipes which are in CS 2 and 3 due to their structural deficits within few years. Condition records of sewer pipes replaced in the past are no longer available. Pipe conditions are rated according to VSA (2007) which is based on DIN EN 752 (2008). The rating system comprises five condition levels assessed on the basis of CCTV records (DIN EN 13508-1, 2013) as specified in DIN EN 13508-2 (2011). We aggregated pipes in the two best and two worst condition classes to one condition class respectively. This was done (i) to avoid identifiability problems, and (ii) due to the intricate prior elicitation which becomes more demanding as more condition states are considered. The age and condition distributions shown in Figure 5 indicate a very good overall condition of the sewer network, including older parts. This is due to the extensive rehabilitation.

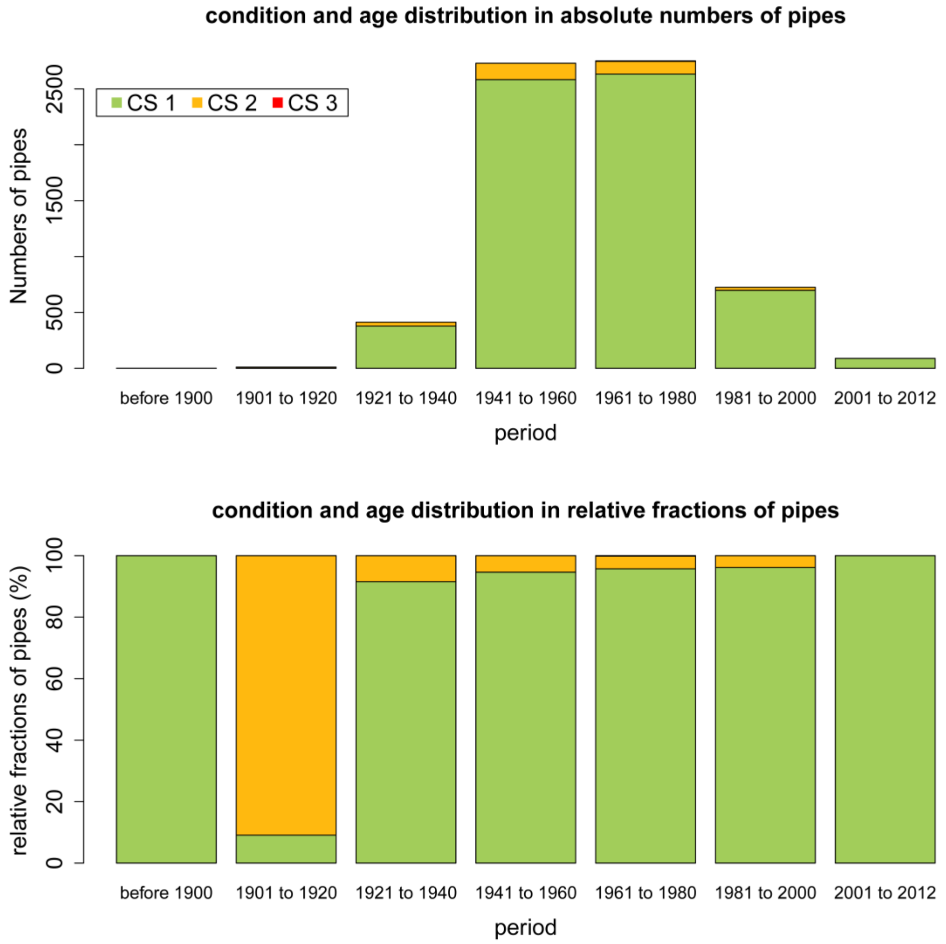

Figure 5. Condition and age distributions in absolute and relative fractions of the real case dataset used for inference. Pipe conditions are rated according to VSA (2007), which comprises five condition classes. The two best and two worst condition classes are aggregated to CS 1 and CS 3, respectively.

5.2 Prior elicitation of model parameter distributions

Prior knowledge of the parameters $\boldsymbol{\theta}_{1:3} = (\alpha, \beta, \mu)^T$ defining the aging behavior were elicited from seven engineers with expertise in the assessment of sewer conditions and rehabilitation (Arreaza Bauer, 2011). The methodologies for expert elicitation and aggregation of several expert opinions to one (inter-subjective) prior were used as applied by Scholten et al. (2013) for water supply mains. Specifically, partial pooling

(Gelman and Hill, 2009) was used for aggregation. The prior distributions of $\boldsymbol{\theta}_{1:3}$ are based on elicited 5, 25, 50, 75 and 95 % quantiles of T_1 and T_2 of concrete sewer pipes irrespective of any further pipe characteristics and influencing factors such as construction period, diameter, traffic load, etc. Results from the individual interviews are shown in Figure B.1 in the Appendix and further described by Arreaza Bauer (2011). The prior parameters are assumed to be independently log-normal distributed with mean $\mathbf{M}_{1:3}$ and standard deviations $\boldsymbol{\Sigma}_{1:3}$. We selected log-normal distributions to describe the priors as the parameters $\boldsymbol{\theta}_{1:3}$ cannot be negative. The values for $\mathbf{M}_{1:3}$ and $\boldsymbol{\Sigma}_{1:3}$ gained by elicitation and subsequent aggregation of the individual expert estimates correspond to those used in the simulations discussed in Section 4.1, see Table 2.

While using rather generic (inter-subjective) prior knowledge about sewer pipe deterioration, we formulated priors for the parameters $\boldsymbol{\theta}_{4:6} = (\lambda_1, \lambda_2, \lambda_3)^T$ based on information from the utility of the sewer network considered here. According to statements by employees of the utility, the rehabilitation activity was approximately constant in the period from the mid-eighties until the present. Given this evidence, we used data from current rehabilitation planning indicating which sewer pipes will be replaced within a planning horizon of five years. Table 5 shows the numbers and percentages of sewer pipes in CS i which will be or have been replaced in this five-year planning period. The respective averaged percentages can be formulated as rehabilitation rates λ_i using Eq. (14) and setting the percentages equal to the probability P_i^{reh} that a pipe in CS i is rehabilitated within one year. Since $\boldsymbol{\theta}_{4:6}$ are zero or positive, we assume the parameters to be independently log-normally distributed with means $\mathbf{M}_{4:6}$ and standard deviations $\boldsymbol{\Sigma}_{4:6}$. Since we have no reliable evidence for the uncertainty of $\boldsymbol{\theta}_{4:6}$, we further assume that $\boldsymbol{\Sigma}_{4:6} = \mathbf{M}_{4:6}/2$. The values obtained for $\boldsymbol{\theta}_{4:6}$ based on the numbers given in Table 5 can be considered as our best knowledge and hence as the most probable values. We therefore set these values equal to the modes of the log-normally distributed priors of the parameters $\boldsymbol{\theta}_{4:6}$ and derive from these the means $\mathbf{M}_{4:6}$ and standard deviations $\boldsymbol{\Sigma}_{4:6}$. The derived values for $\mathbf{M}_{4:6}$ and $\boldsymbol{\Sigma}_{4:6}$ are included in Table 5.

Table 5. Total numbers of pipes in CS $i = 1,2,3$, percentages of pipes in CS $i = 1,2,3$ to be replaced within the planning period from 2011 to 2015 and hyperparameters $\mathbf{M}_{4:6}$ and $\boldsymbol{\Sigma}_{4:6}$.

CS	total number of pipes	Percentages of pipes to be replaced (%)						Hyperparameters		
		2011	2012	2013	2014	2015	Mean	j	M_j	Σ_j
1	6383	0.3	0.4	0.5	0.6	0.3	0.4	4	$5.70 \cdot 10^{-3}$	$2.85 \cdot 10^{-3}$
2	334	3.0	0.9	3.0	4.2	0.3	2.3	5	$3.22 \cdot 10^{-2}$	$1.61 \cdot 10^{-2}$
3	4	50.0	0.0	0.0	0.0	0.0	10.0	6	0.15	$7.36 \cdot 10^{-2}$

5.3 Results of the inference

Figure 6 shows both inferred survival functions and the mean of the survival functions as suggested by the prior. The estimated survival functions suggest a shorter residence time in CS 2 but a longer total physical lifespan (defined here as the age a pipe at transition to CS 3) compared to the prior. The resulting median physical lifespan of approximately 95 years appears to be realistic, knowing that strict quality control procedures are in place in this utility. Again, $S_2(t)$ is much more uncertain than $S_1(t)$ for the possible reasons already discussed in Section 4.1. The figure illustrates further the estimated mean probability that a pipe with age t is not replaced. This survival function represents the functional survival of the pipes.

Figure 7 shows prior and posterior marginal distributions of the model parameters. The low identifiability of λ_1 is also apparent here. The locations of both posterior marginal distributions of λ_2 and λ_3 are clearly shifted towards larger values. As it reveals from Table 5, only 5.0 % of the pipes are currently in CS 2 and 0.06 % are in CS 3. Thus, rather high replacement rates λ_2 and λ_3 do not necessarily imply that an unrealistically large number of pipes has been replaced. We admit that the prior of λ_3 is derived on the basis of very few records. It can be expected that more pipes, particular pipes in CS 3, are replaced until 2015 than indicated by Table 5 if further pipes are observed to be in CS 3 in this period. This would explain that the posteriors suggest higher rehabilitation rates than the priors.

To show the relevance of considering pipe rehabilitation, we also performed an inference with the unconditioned likelihood function according to Eq. (1). Figure 8 shows the corresponding estimated survival functions together with the mean of the survival functions as described by the prior. The effect of ignoring pipe rehabilitation is evident, as we obtain a completely unrealistic median physical lifespan of approximately 440 years. The difference between Figure 6 and Figure 8 reflects the substantial rehabilitation carried out by the utility in the past. As a consequence only a relatively small number of pipes is in CS 2 and even less in CS 3. Rehabilitation leads to a selection effect on pipes (the worse the condition of a pipe the more likely is its replacement) so that slow-aging pipes are over-represented in the data. The fact that rehabilitation depends on the CS is further reflected by a more distinct bias of $S_2(t)$ compared to $S_1(t)$. This is confirmed by the results obtained from the analysis with NetCoS, see supplementary material.

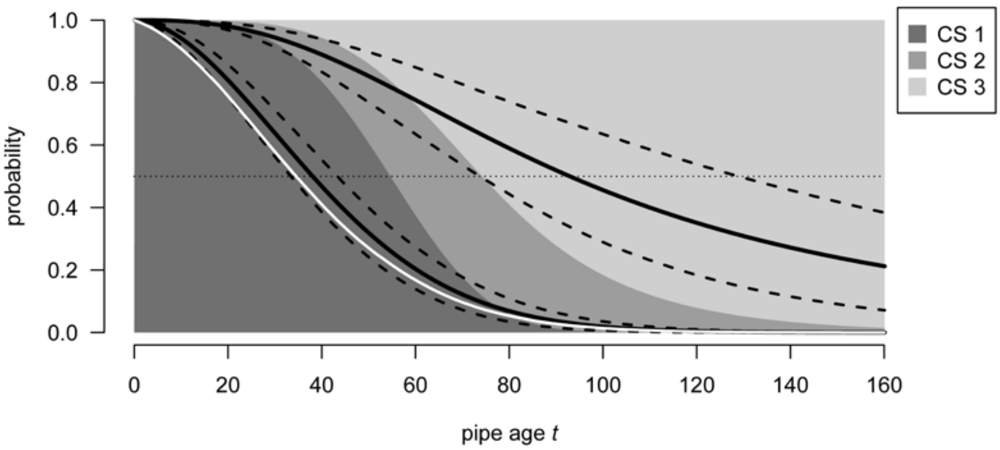

Figure 6. Mean of the prior and estimated survival functions. The conditioned likelihood was used for inference. The gray shaded areas indicate the mean of the survival functions described by the prior. The black solid lines are the means of the estimated survival functions and the dashed lines are the 10% and 90% quantiles based on the posterior distribution of $\boldsymbol{\theta}$. The white line describes the mean probability that a pipe with certain age is not rehabilitated based on the posterior distribution of $\boldsymbol{\theta}$.

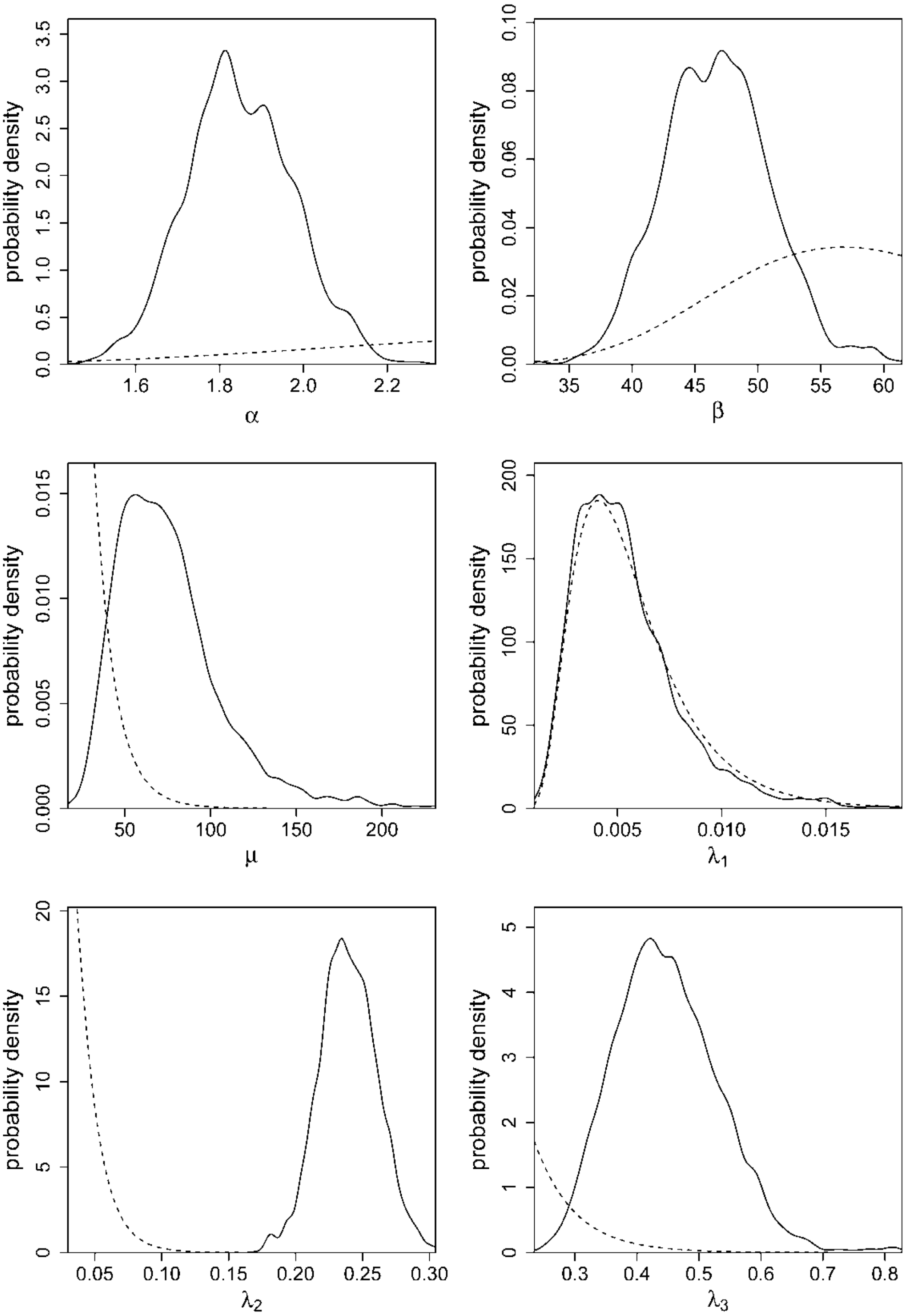

Figure 7. Prior (dashed lines) and posterior (solid lines) marginal distributions of the model parameters $\boldsymbol{\theta}$.

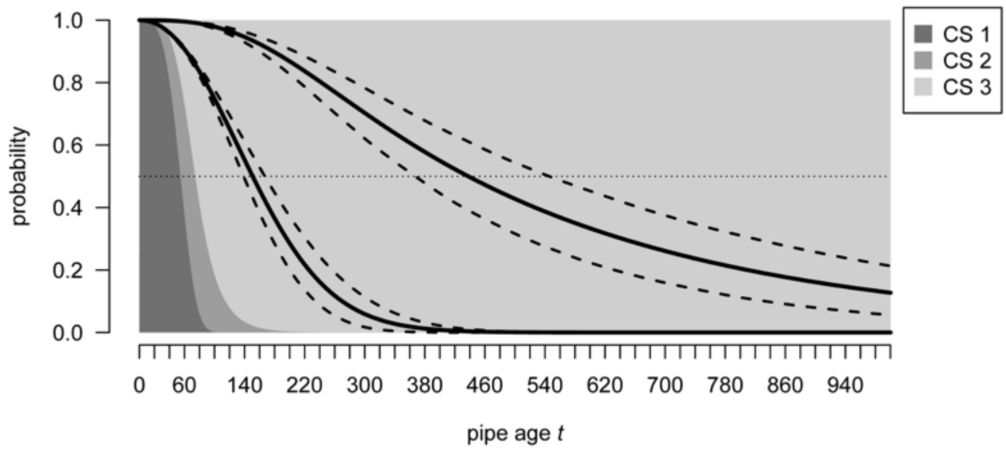

Figure 8. Mean of the prior and estimated survival functions. The unconditioned likelihood was used for inference. The gray shaded areas indicate the means of the survival functions described by the prior. The solid lines are the means of the estimated survival functions and the dashed lines are the 10% and 90% quantiles based on the posterior distribution of $\boldsymbol{\theta}$. The estimated physical lifespan is unrealistically high.

6 Discussion

We introduced a sewer deterioration model to deal with missing historical records of sewer conditions. We approached the problem by conditioning the likelihood on the fact that we only use condition data from pipes that have *not* been rehabilitated. A rehabilitation model was needed to calculate the conditioned likelihood. We applied Bayesian inference to identify the model. Our results show that the proposed deterioration model copes satisfactorily with a lack of historical records of sewer conditions. We will discuss the results and the limitations of our model in more detail below.

6.1 Explicit consideration of past rehabilitation

In practice, the availability of asset data is often less than optimal, and historical records of maintenance and rehabilitation are very often missing. We consequently developed our model to deal with two important shortcomings: (i) lack of historical data and (ii) small datasets. When rehabilitation is not considered adequately, the lack of historical data leads to a systematic overestimation of sewer life spans, as previously reported (Scheidegger et al., 2011; Schmidt, 2009). In this article, we show that this bias can be removed by combining the deterioration model with a rehabilitation model, and that parameters can be estimated when combining prior information with data via Bayesian inference. Two examples are used to demonstrate these points.

In the first example, we applied the proposed model to a well-defined synthetic dataset generated on the basis of the same underlying models for deterioration and rehabilitation of the sewer network as were used for the inference. Very good compliance is obtained between the estimated survival functions and the predefined ones used for data generation. We would stress that other models that do not consider rehabilitation failed to reproduce the original parameter values for this idealized data generated by NetCoS as revealed by Scheidegger et al. (2011).

We further applied the model to data of a real sewer network which underwent extensive rehabilitation in recent decades. Without considering these rehabilitations in the model, the data suggests extremely long and unrealistic physical life spans. However, the proposed model effectively compensates for the bias, resulting in realistically estimated life spans.

6.2 Model identifiability and limitations

We did not succeed in estimating the model parameters by frequentist inference, e.g. by maximizing the likelihood. The main reason is that the available datasets do not contain enough information to estimate the rehabilitation rates independently of the parameters defining the deterioration. Thus, a Bayesian approach to include prior knowledge is needed for parameter inference. The use of expert knowledge on pipe deterioration has already been proposed by Herz (1995) and Kleiner (2001) in the case of scarce data and information availability. In this sense, we used an approach which allows us to exploit the best available (expert) information and to update this knowledge by inferences from data.

In order to estimate the quantitative influence of the prior on the parameter inference, we performed a sensitivity analysis and identified the most influential hyperparameters. Knowledge about the importance of the hyperparameters may be useful when elaborating a concept for eliciting prior knowledge. Elicitation and quantification of prior knowledge was outside the scope of this paper and may be found in O'Hagan et al. (2006), Rinderknecht et al. (2011, 2012) and Scholten et al. (2013). We assume that the deterioration of sewer pipes does not differ fundamentally between similar sewer networks in similar regions. Thus, prior knowledge based on different expert opinions or datasets appears meaningful. However, we have experienced that the rehabilitation strategy may differ substantially between different utilities. Prior knowledge of rehabilitation thus needs to be acquired carefully for each individual case.

By analyzing the model with the aid of synthetic data, we gained important insights into its behavior. However, the synthetic data probably do not reflect the variability of real data. So the results do not necessarily imply that the model will perform well in real life. Nevertheless, the model shows promising performance when applied to real data lacking historical records. Even though we recognized a considerable shift in the location of the posteriors of model parameters defining the rehabilitation rates in relation to the priors, the available data gives no indication of possible deficits in the model structure.

We implemented a very simple rehabilitation model, assuming age-invariant and exclusively condition-dependent rehabilitation rates. In reality, it is probable that the rehabilitation strategy and hence the rehabilitation rates vary over time. It is important to keep in mind that the model does not aim to identify past rehabilitation but to determine deterioration as accurately as possible from the available information. However, the rehabilitation model could be substituted by a more complex model if useful. Further insight into the deficits of the model structure may be gained by using NetCoS and introducing variability to the user-defined processes deterioration and rehabilitation driving the data generator. This would allow the supposed variability of real data to be emulated. Similarly, exceptional real cases comprising both extensive rehabilitation in the past and historical data may extend our knowledge of the model behavior. In this case, the results could be compared by using (i) the unconditioned likelihood (of the deterioration model alone) in combination with the dataset including the historical records, and (ii) the conditioned likelihood (of the combined deterioration-rehabilitation model) and the data without historical records.

The proposed deterioration model may also be substituted by another one or extended by a range of additional features. These could include the incorporation of additional factors influencing deterioration, and consideration of more than one observed condition state per sewer line, allowing more accurate predictions and considerably extending its application.

7 Conclusions

- If datasets lack historical records, sewer life spans are overestimated if the applied model does not account for the combined effect of deterioration and rehabilitation. The proposed combined deterioration and rehabilitation model effectively compensates for the bias in estimated life span.
- The inclusion of prior knowledge is a necessity due to the limited information in the data and therefore the limited identifiability of the model parameters. Since at least some prior knowledge is available, the strength of Bayesian inference is obvious, in particular in the case of small datasets.
- The analysis of model sensitivity to the prior revealed that the inference results are mainly influenced by the means and only partly by the standard deviations of the priors of four out of six model parameters. This result can facilitate the knowledge elicitation process from experts, since the elicitation of parameter uncertainty is more challenging than merely eliciting its mean.
- The applied importance sampling technique for sensitivity analysis permitted an efficient implementation of regional sensitivity analysis with reasonable computational demand.
- The approach presented here is flexible and allows individual aspects to be substituted and extended. Consideration of (i) more than three condition classes, (ii) two or more subsequently observed condition states per sewer pipe, and (iii) additional factors influencing pipe deterioration such as pipe material and diameter may be relevant to a broader range of applications.

8 Acknowledgements

This research was funded within the National Research Programme NRP 61 on Sustainable Water Management by the Swiss National Science Foundation (SNF). We thank the seven experts who participated in interviews for eliciting their knowledge in sewer deterioration. We further thank Lisa Scholten, Adelaida Arreaza and Simon Lukas Rinderknecht who prepared and conducted the expert interviews.

9 References

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

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

Arreaza Bauer, A.S., 2011. Robust elicitation of imprecise expert knowledge for water infrastructure management. Master thesis, Swiss Federal Institute of Technology Zurich, Zurich, Switzerland.

Baur, R. and Herz, R., 2002. Selective inspection planning with ageing forecast for sewer types. Water Science and Technology 46 (6-7), 389-396.

Bolstad, M.B., 2007. Introduction to Bayesian statistics, John Wiley & Sons Hoboken, New Jersey.

Congdon, P., 2006. Bayesian statistical modelling, John Wiley & Sons, Chichester, England.

DIN EN 752, 2008. Drain and sewer systems outside buildings, English version of DIN EN 752:2008-04, DIN, Germany

DIN EN 13508-1, 2013. Investigation and assessment of drain and sewer systems outside buildings - Part 1: General requirements; English version EN 13508-1: 2012, English translation of DIN EN 13508-1:2013 01, DIN, Germany

DIN EN 13508-2, 2011. Investigation and assessment of drain and sewer systems outside buildings - Part 2: Visual inspection coding system, English translation of DIN EN 13508-2:2011-08, DIN, Germany

Gasser, T., Kneip, A. and Köhler, W., 1991. A flexible and fast method for automatic smoothing. Journal of the American Statistical Association 86 (415), 643-652.

Gelman, A., Carlin, J.B., Stern, H.S. and Rubin, D.B., 2004. Bayesian data analysis, Chapman & Hall/CRC, London.

Gelman, A. and Hill, J., 2009. Data analysis using regression and multilevel/ hierarchical models, Cambridge University Press, New York.

Herrmann, E. and Maechler, M., 2011. lokern: Kernel Regression Smoothing with Local or Global Plug-in Bandwitdth. R package version 1.1-3, http://CRAN.R-project.org/package=lokern.

Herz, R., 1995. Aging and renewal of infrastructure assets - a cohort survival model (Alterung und Erneuerung von Infrastrukturbeständen – ein Kohortenüberlebensmodell). Jahrbuch für Regionalwissenschaft. 14/15, 5-29 (in German).

Kleiner, Y., 2001. Scheduling inspection and renewal of large infrastructure assets. Journal of Infrastructure Systems 7 (4), 136-143.

Kleiner, Y. and Rajani, B., 2001. Comprehensive review of structural deterioration of water mains: statistical models. Urban Water 3 (3), 131-150.

Le Gat, Y., 2009. An extension of the Yule process for stochastic modeling of recurrent events. Applications on failures of pressurized water supply pipes (Une extension du processes de Yule pour la modélisation stochastique des événements récurrents. Application aux défaillances de canalisations d`eau sous pression.). Ph.D. thesis, Agro Paris Tech, Cemagref de Bordeaux, Paris (in French).

Mailhot, A., Pelletier, G., Noel, J.F. and Villeneuve, J.P., 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.

Micevski, T., Kuczera, G. and Coombes, P., 2002. Markov model for storm water pipe deterioration. Journal of Infrastructure Systems 8 (2), 49-56.

Mishalani, R.G. and Madanat, S.M., 2002. Computation of infrastructure transition probabilities using stochastic duration models. Journal of Infrastructure Systems 8 (4), 139-148.

Müller, K., 2002. Development of a generally applicable methodology for selective primary inspection of sewer pipes and house connecting pipes. Part A: scientific investigations. Final report (Entwicklung eines allgemein anwendbaren Verfahrens zur selektiven Erstinspektion von Abwasserkanälen und Anschlussleitungen. Teil A: wissenschaftliche Untersuchungen. Abschlussbericht). Chair and Institute of Environmental Engineering, RWTH Aachen, Aachen, Germany (in German).

O'Hagan, A., Buck, C.E., Daneshkhah, A., Eiser, J.R., Garwite, P.H., Jenkinson, D.J., Oakley, J.E. and Rakow, T., 2006. Uncertain Judgements: Eliciting Experts' Probabilities, John Wiley & Sons, Chichester, London.

R Core Team, 2012. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, http://www.R-project.org/.

Rajani, B. and Kleiner, Y., 2001. Comprehensive review of structural deterioration of water mains: physically based models. Urban Water 3 (3), 151-164.

Rinderknecht, S.L., Borsuk, M.E. and Reichert, P., 2011. Eliciting density ratio classes. International Journal of Approximate Reasoning 52 (6), 792-804.

Rinderknecht, S.L., Borsuk, M.E. and Reichert, P., 2012. Bridging uncertain and ambiguous knowledge with imprecise probabilities. Environmental Modelling & Software 36, 122-130.

Robert, C.P. and Casella, G., 2010. Introducing Monte Carlo methods with R, Springer, New York.

Rutsch, M., Rieckermann, J., Cullmann, J., Ellis, J.B., Vollertsen, J. and Krebs, P., 2008. Towards a better understanding of sewer exfiltration. Water Research 42 (10-11), 2385-2394.

Saegrov, S., 2005. CARE-S, Computer Aided Rehabilitation of Sewer and Storm Water Networks, IWA Publishing, London.

Saltelli, A., Chan, K. and Scott, E.M., 2000. Sensitivity analysis, John Wiley & Sons, Chichester, England.

Scheidegger, A., 2012. adaptMCMC: Implementation of a generic adaptive Monte Carlo Markov Chain sampler. R package version 1.1, http://cran.r-project.org/web/packages/adaptMCMC/.

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

Scheidegger, A. and Maurer, M., 2012. Identifying biases in deterioration models using synthetic sewer data. Water Science and Technology 66 (11), 2363-2369.

Scheidegger, A., Scholten, L., Mauer, M. and Reichert, P., 2013. Extension of pipe failure models to consider the absence of data from replaced pipes. Water Research 47 (11), 3696-3705.

Schmidt, T., 2009. Modeling of sewer deterioration processes on the basis of condition records (Modellierung von Kanalalterungsprozessen auf der Basis von Zustandsdaten). Ph.D. thesis, TU Dresden, Dresden, Germany (in German).

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

Seifert, B. and Gasser, T., 1996. Finite-sample variance of local polynomials: Analysis and solutions. Journal of the American Statistical Association 91 (433), 267-275.

Seifert, B. and Gasser, T., 2000. Data adaptive ridging in local polynomial regression. Journal of Computational and Graphical Statistics 9 (2), 338-360.Tran, H.D., 2007. Investigation of deterioration models for stormwater pipe systems. Ph.D. thesis, Victoria University, Melbourne, Australia.

Trujillo Alvarez, R., 1995. Prediction of the rehabilitation demand and development of rehabilitation strategies for urban water pipe networks.(Bedarfsprognose und Strategieentwicklung für die Rehabilitation städtischer Wasserrohrnetze). Schriftenreihe des Instituts für Städtebau und Landesplanung. Universität Karlsruhe, (in German).

Tversky, A. and Kahneman, D., 1974. Judgment under uncertainty: heuristics and biases. Biases in judgments reveal some heuristics of thinking under uncertainty. Science 185 (4157), 1124-1131.

Vihola, M., 2012. Robust adaptive Metropolis algorithm with coerced acceptance rate. Statistics and Computing 22 (5), 997-1008.

VSA, 2007. Maintenance of sewer systems, condition assessment of urban drainage assets (Erhaltung von Kanalisationen, Zustandsbeurteilung von Entwässerungsanlagen). Verband Schweizer Abwasser- und Gewässerschutzfachleute, Zurich, Switzerland (in German).

Appendix A: Equations

$$P(\mathrm{NR}(D) \mid C(\tau)=2, R \le \tau, D, \boldsymbol{\theta}) \tag{A.1}$$

$$= \frac{\int_0^R \int_\tau^\infty e^{-\lambda_2(\tau-R)} p(t_1,t_2 \mid \boldsymbol{\theta}) dt_2 dt_1 + \int_R^\tau \int_\tau^\infty e^{-\lambda_1(t_1-R)-\lambda_2(\tau-t_1)} p(t_1,t_2 \mid \boldsymbol{\theta}) dt_1 dt_2}{\int_0^\tau \int_\tau^\infty p(t_1,t_2 \mid \boldsymbol{\theta}) dt_2 dt_1} \cdot e^{-\lambda_2(D-\tau)}$$

$$= \frac{e^{-\lambda_2(\tau-R)} \int_0^R t_1^{(\alpha-1)} e^{-(t_1/\beta)^\alpha - (\tau-t_1)/\mu} dt_1 + \int_R^\tau t_1^{(\alpha-1)} e^{-\lambda_1(t_1-R)-\lambda_2(\tau-t_1)-(t_1/\beta)^\alpha-(\tau-t_1)/\mu} dt_1}{\int_0^\tau t_1^{(\alpha-1)} e^{-(t_1/\beta)^\alpha-(\tau-t_1)/\mu} dt_1} \cdot e^{-\lambda_2(D-\tau)}$$

$$P(\mathrm{NR}(D)\mid C(\tau)=3, R\le\tau, D, \boldsymbol{\theta}) \tag{A.2}$$

$$= \frac{e^{-\lambda_3(\tau-R)}\int_0^R\int_{t_1}^R p(t_1,t_2\mid\boldsymbol{\theta})dt_2dt_1 + \int_0^R\int_R^\tau e^{-\lambda_2(t_2-R)-\lambda_3(\tau-t_2)}p(t_1,t_2\mid\boldsymbol{\theta})dt_2dt_1 + \int_R^\tau\int_{t_1}^\tau e^{-\lambda_1(t_1-R)-\lambda_2(t_2-t_1)-\lambda_3(\tau-t_2)}p(t_1,t_2\mid\boldsymbol{\theta})dt_2dt_1}{\int_0^\tau\int_{t_1}^\tau p(t_1,t_2\mid\boldsymbol{\theta})dt_2dt_1}\cdot e^{-\lambda_3(D-\tau)}$$

$$= \left(\frac{e^{-\lambda_3(\tau-R)}\int_0^R t_1^{(\alpha-1)}e^{-(t_1/\beta)^{\theta_1}}\left(1-e^{-(R-t_1)/\mu}\right)dt_1}{\int_0^\tau t_1^{\alpha-1}e^{-(t_1/\beta)^\alpha}\left(1-e^{-(\tau-t_1)/\mu)}\right)dt_1} + \frac{\int_0^R t_1^{(\alpha-1)}e^{-(t_1/\beta)^\alpha}\left(e^{-\lambda_2(\tau-R)-(\tau-t_1)/\mu}-e^{-(R-t_1)/\mu-\lambda_3(\tau-R)}\right)dt_1 + \int_R^\tau t_1^{(\alpha-1)}e^{-(t_1/\beta)^\alpha-\lambda_1(t_1-R)}\left(e^{(-1/\mu-\lambda_2)(\tau-t_1)}-e^{-\lambda_3(\tau-t_1)}\right)dt_1}{\left(\mu(\lambda_3-\lambda_2)-1\right)\cdot\int_0^\tau t_1^{\alpha-1}e^{-(t_1/\beta)^\alpha}\left(1-e^{-(\tau-t_1)/\mu)}\right)dt_1}\right)\cdot e^{-\lambda_3(D-\tau)}$$

$$S_1(t\mid\boldsymbol{\theta}) = P(T_1>t\mid\boldsymbol{\theta}) = \int_t^\infty p_1(t_1\mid\boldsymbol{\theta})dt_1 = e^{-(t/\beta)^\alpha} \tag{A.3}$$

$$S_2(t\mid\boldsymbol{\theta}) = P(T_2>t\mid\boldsymbol{\theta}) = \int_0^\infty p(t_1\mid\boldsymbol{\theta})\int_t^\infty p(t_2\mid t_1,\boldsymbol{\theta})dt_2dt_1 \tag{A.4}$$

$$= e^{-(t/\beta)^\alpha} + \int_0^t \frac{\alpha}{\beta}\left(\frac{t_1}{\beta}\right)^{\alpha-1} e^{-\left(\frac{t_1}{\beta}\right)^\alpha-\frac{t-t_1}{\mu}}dt_1$$

Appendix B: Figures

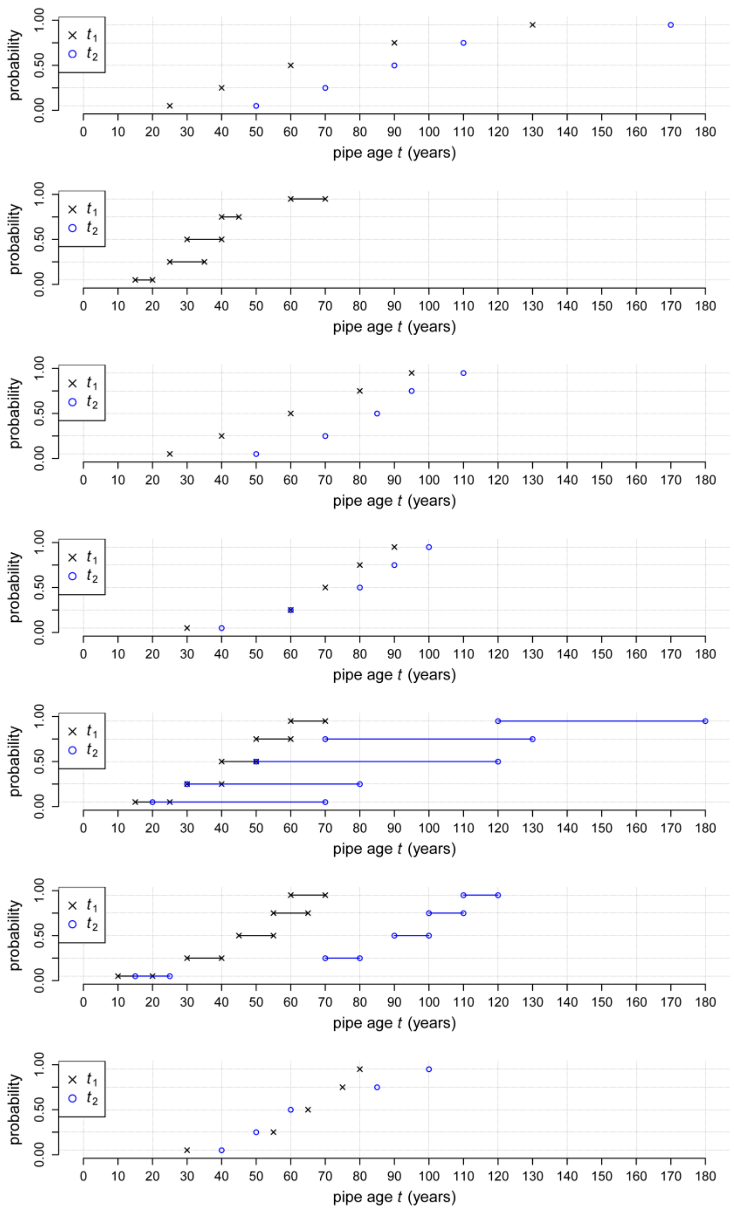

Figure B.1. Quantiles of pipe ages T_1 and T_2 for concrete pipes elicited from seven experts. The quantiles were partly elicited as single values and partly as ranges, see Arreaza Bauer (2011) and Scholten et al. (2013). In one case, only T_1 was elicited (second graph). In some cases, T_2 is smaller or equal to T_1, $T_2 \leq T_1$. This is due to the difficulty of expressing quantities in form of quantiles and to neglecting a consistency check in the interview protocol to ensure that $T_2 > T_1$.

Supplementary material - additional results

The figure below shows supplemental results of the example discussed in section 4.1 in form of survival functions. The same data was used but using the unconditioned likelihood according to Eq. (1). The figure illustrates the bias obtained if the pure deterioration model is used in combination with data affected by both deterioration and rehabilitation.

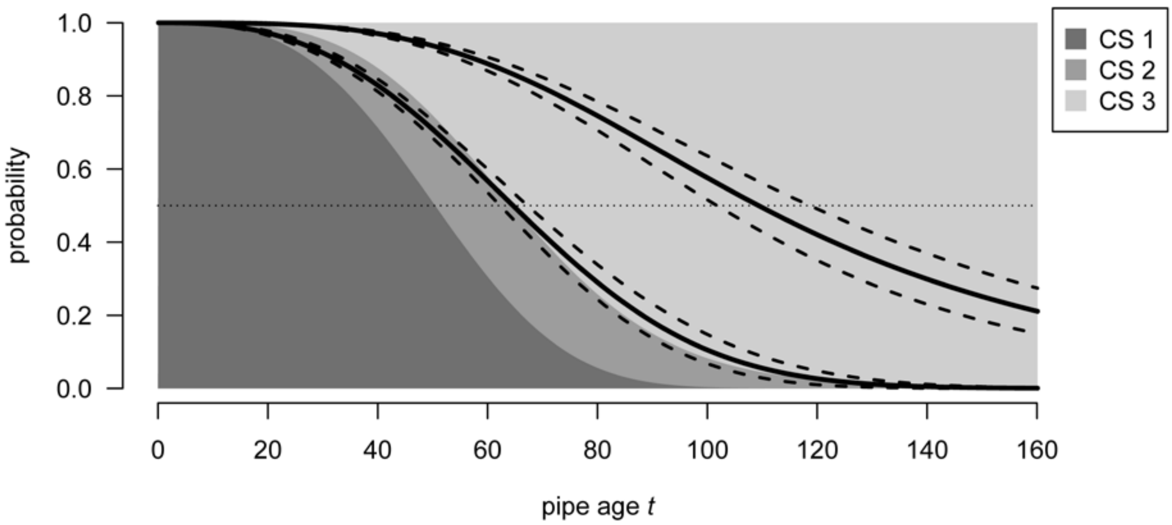

Figure 1. Predefined and estimated survival functions. The unconditioned likelihood was used for inference. The gray shaded areas indicate the predefined survival functions based on the parameters in Table 2 used for data generation in NetCos. The solid lines are the means of the estimated survival functions, and the dashed lines are the 10 % and 90 % quantiles based on the posterior distribution of $\boldsymbol{\theta}$. The estimated survival functions are biased suggesting considerably longer sojourn times in CS 1 and 2.