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. It was funded within the National Research Programme 61 on "Sustainable Water Management" by the Swiss National Science Foundation (http://www.nrp61.ch).

Combining expert knowledge and local data for improved service life modeling of water supply networks

Lisa Scholten[a,b*], Andreas Scheidegger[a], Peter Reichert[a,b] and Max Maurer[a,c]

[a] *Eawag, Swiss Federal Institute of Aquatic Science and Technology,* U*eberlandstrasse 133, P.O. Box 611, CH-8600 Duebendorf, Switzerland*

[b] *ETH Zürich, Department of Environmental Systems Science, Institute of* 20 *Biogeochemistry and Pollutant Dynamics, CH-8092 Zurich, Switzerland*

[c] *ETH Zürich, Department of Civil, Environmental and Geomatic Engineering, Institute of Environmental Engineering, Wolfgang-Pauli-Strasse 15, CH-8093* 18 *Zurich, Switzerland* 19

** Corresponding author. E-Mail: lisa.scholten@eawag.ch*

Published in Environmental Modelling & Software (42): 1-16.

View at publisher: *http://www.sciencedirect.com/science/article/pii/S1364815212002848*

PUBLISHED MANUSCRIPT: Scholten et al. 2013. *Combining expert knowledge and local data for improved service life modeling of water supply netwoks*. Environmental Modelling & Software 42: 1-16.

Combining expert knowledge and local data for improved service life modeling of water supply networks

Lisa Scholten[a,b*], Andreas Scheidegger[a], Peter Reichert[a,b], Max Maurer[a]

[a] Eawag: Swiss Federal Institute of Aquatic Science and Technology, Überlandstrasse 133, P.O.Box 611, CH-8600 Dübendorf, Switzerland
[b] ETH Zürich, Department of Environmental Sciences, Institute of Biogeochemistry and Pollutant Dynamics (IBP), CH-8092 Zürich, Switzerland
* Corresponding author. Tel.: +41 58 765 5590; Fax +41 58 765 5028; E-mail address : lisa.scholten@eawag.ch

Abstract

The presented approach aims to overcome the scarce data problem in service life modeling of water networks by combining subjective expert knowledge and local replacement data. A procedure to elicit imprecise quantile estimates of survival functions from experts, considering common cognitive biases, was developed and applied. The individual expert priors of the parameters of the service life distribution are obtained by regression over the stated distribution quantiles and aggregated into a single prior distribution. Furthermore, a likelihood function for the commonly encountered censored and truncated pipe replacement data is formulated. The suitability of the suggested Bayesian approach based on elicitation data from eight experts and real network data is demonstrated. Robust parameter estimates could be derived in data situations where frequentist maximum likelihood estimation is unsatisfactory, and to show how the consideration of imprecision and in-between-variance of experts improves posterior inference.

Keywords

scarce data, expert knowledge elicitation, expert aggregation, Bayesian inference, water supply network, service life modeling

1 Introduction

1.1 Challenge

The coming of age of the water infrastructure poses an increasing challenge for utility managers. One of the key issues is to assess the long-term development of network rehabilitation demand. The motivation is to ensure that sufficient funding is raised and appropriately allocated to achieve the foreseen level of service. As a result, the last decade of water infrastructure management has seen increased development, testing, and application of mathematical models in rehabilitation planning and network failure estimation (Alvisi and Franchini, 2010; Dridi et al., 2009; Eisenbeis et al., 1999; Fuchs-Hanusch et al., 2008; Kleiner and Rajani, 2001; Pelletier et al., 2003; Rajani and Kleiner, 2001).

1.2 Network rehabilitation and survival modeling with scarce data

Within these models, the expected service life of water supply pipes (also referred to as "pipe lifetime" or "pipe survival"), is inferred from historic failure or replacement data. A shortcoming of these models is that they are only applicable for rather well-kept and extensive data sets, which are not ubiquitous in many utilities, as for example in Switzerland (>50 % of the population served by utilities with < 10.000 customers, (SVGW, 2009)) where even the best documentation does not help to overcome the prevalence of short network length and thus small sample size.

Different strategies have been proposed to handle scarce data situations (i.e. situations in which model parameters cannot be identified or are too uncertain to be of use for practical rehabilitation planning):

Purely data-based methods. Renaud, De Massiac et al. (2009) tried to overcome the scarce data difficulty by amalgamating the data from a number of French water utilities to calibrate the model, but found that this did not result in models that were more effective.

Purely expertise-based methods. The survival model for rehabilitation prediction and its parameters or quantiles are directly elicited from experts, for example based on cohort survival (Herz, 1995, 1998). Even though the value of subjective expert judgment is largely unquestioned in practice, only few have proposed its use in water infrastructure engineering (Dridi et al., 2009; Korving and van Noortwijk, 2008).

Bayesian combination of subjective expert knowledge with data, e.g. (Dridi et al., 2009). Especially for small data sets, Bayesian inference might be advantageous over frequentist (purely data driven) inference. While the likelihood function is often discussed in detail, e.g. (Mailhot et al., 2000), the elicitation and influence of the prior probability distribution are rarely explored. This work does not only discuss the derived likelihood function for left-truncated and right-censored data, but also shows a meaningful procedure for quantitative expert knowledge elicitation and combination with locally available data. The performance of the Bayesian approach is compared to using a purely data-driven frequentist estimation, on the basis of different amounts of data.

1.3 Background on expert knowledge elicitation and aggregation

Regarding the elicitation of expert knowledge (referred to as "expert elicitation"), a wealth of publications and guidelines exist. Recent reviews on the role of expert knowledge in modeling and important elicitation aspects are the works of Krueger et al. (2012), Kynn (2008) and Low Choy et al. (2009). The former not only provides an overview about the formal use of expert opinion in modeling practice, but also a discussion about common critiques such as the definition of expertise, and the representativeness of experts. Kynn (2008) offers a critical review of the past decades of psychological research on expert elicitation, as well as relevant work from other fields. She concludes that over a decade of research into heuristics and biases has been almost completely ignored by the statistical literature on expert elicitation. (Interestingly, in the literature regarding elicitation of Herz's cohort survival model for pipe survival estimation, the wealth of publications on expert elicitation seems to have gone similarly unnoticed.) The latter work of Low Choy et al. (2009) comprises a review of applications of expert elicitation throughout the ecological literature. Apart from these, the reader is referred to the works of (Ayyub, 2001; Cooke, 1991; Cooke and Goossens, 2008; O'Hagan et al., 2006) for more in-depth information on the historical and theoretical background. Even though the effect of imprecision in the elicited data itself is sometimes discussed (O'Hagan, 2012; O'Hagan et al., 2006; Oakley and O'Hagan, 2007), it seems that the possibility of explicit elicitation of such imprecision has been overlooked in the past. The elicitation guideline developed hereafter includes the elicitation of imprecise estimates.

Consulting multiple experts can be interpreted as an artificial increase in sample size of the experiment (Clemen and Winkler, 1999) with the objective of getting an approximation to the intersubjective knowledge of the expert community rather than the subjective knowledge of a single expert (Gillies, 1991; Rinderknecht et al., 2012). A key decision is the way to aggregate this information into one single distribution, which is also reflected in numerous publications, e.g. (Ayyub, 2001; Clemen and Winkler, 1999; Cooke, 1991; Genest and Zidek, 1986; Jouini and Clemen, 1996; Kuhnert et al., 2010; O'Hagan, 2012; O'Hagan et al., 2006; O'Leary et al., 2009).

Following the categorization of Clemen and Winkler (1999), aggregation can be achieved by mathematical and behavioral combination. Unless a mutual consensus of the experts is envisaged, elicitation is performed on an individual basis and later mathematically aggregated. Mathematical combination approaches are often further subdivided into axiomatic (also named classical or pooling) approaches and Bayesian approaches (Clemen and Winkler, 1999; Cooke, 1991).

Many of the axiomatic approaches consist of linear pooling (e.g. simple weighted averaging) and differ only in the weighting of the elicited probabilities. Weights can be equal, or different for individual experts, , e.g.assigned according to confidence levels or calibration. In comparative aggregation studies equal weighting performed reasonably well, though it was outperformed by more complex weighting rules in specific situations (Cooke, 1991; Cooke and Goossens, 2008). Clemen and Winkler (1999) conclude that simpler aggregation methods such as the simple equally-weighted arithmetic average perform just as well as more complex methods, a notion widely supported by others (Larrick and Soll, 2006; O'Hagan et al., 2006).

Therefore, two axiomatic aggregation methods were chosen for comparison. In approach A the differences between experts are considered to stem from the variability between different networks, whereas in approach B it is assumed that all experts refer to the same network (but the expert statements are uncertain and therefore different). The experts were assumed to be equally qualified, thus assigning equal weights.

1.4 Elicitation of the parameters from a multivariate survival function

When it comes to practical elicitation, it is often not possible to elicit the unknown probability distribution directly, but only the observable quantities (Lele and Allen, 2006; O'Hagan et al., 2006). This is because the experts can neither be expected to define a specific distributional form nor to estimate distributional parameters of possible functional models directly unless specifically trained. In the case of multivariate survival models, correlation between model parameters makes direct elicitation of parameters even more unreliable. To overcome this limitation, an approach to elicit the model parameters indirectly from experts' judgment on selected quantiles of the survival distribution was developed.

1.5 Goal and structure of the paper

The objective of this paper is to present an approach to overcome the scarce data problem in water pipe lifetime modeling by combining expert knowledge and local data.

The methodic contribution of this approach consists of

1. nomination of considered survival models (section 2.1),
2. a specifically developed elicitation guideline to obtain (imprecise) survival function quantiles from experts (2.2),
3. inference of a bivariate prior distribution of the survival function parameters from the stated quantiles,
4. mathematical aggregation of the experts statements into a single prior distribution for the survival model (2.3),
5. formulation of a novel likelihood function of the survival model for censored and truncated data (2.4.1),
6. frequentist parameter estimation under varying amounts of data (2.4.2), and
7. Bayesian updating of the expert prior using different amounts of data (2.4.3).

The utility data considered are summarized in section 2.5. In section 2.6, it is presented how possible sources of uncertainty were dealt with. Results consequently cover the elicited expert knowledge (3.1), parametric model identification (3.2), most appropriate aggregation method (3.3), as well as the performance of maximum likelihood estimation (3.4) and Bayesian inference (3.5) in light of scarce data. Conclusions for its use in water main survival modeling are drawn in section 4.

As this work is interdisciplinary, the aim is to address readers from different professional backgrounds. Please bear with us for making aspects explicit which an expressed specialist might judge as banality. For linguistic convenience, experts will be male, the interviewer and the analyst female.

2 Methods

2.1 Choice of the survival model

The age t_i is the age of pipe i at the end of its service lifetime. As the service lifetime is different for every pipe it seems natural to model t_i with a random variable T. Because negative lifetimes are impossible, the distribution of T should support only positive values. T can be described by its probability density function p(t) or its survival function $S(t)$=P(T>t).

In practice $S(t)$ is described by a parametric function that ideally has a small number of parameters while being flexible enough to fit the data. Three parametric models that satisfy these requirements, and allow for a great variety of shapes, are the Weibull, lognormal and gamma distribution (Wayne, 2004).

These three models are used to infer the survival function from (a) the elicited expert knowledge expressed in the form of stated quantiles, (b) the available utility data with frequentist inference, and (c) a combination of both by Bayesian inference. The Weibull distribution is parameterized with $\boldsymbol{\theta} = (\alpha, \beta)^T$ such that $E(T) = \beta\Gamma(1 + 1/\alpha)$ and $\mathrm{Var}(T) = \beta^2\Gamma(1 + 2/\alpha) - E(T)^2$, the lognormal with $\boldsymbol{\theta} = (\mu, \sigma)^T$, whereas E($T$) =$\mu$ and sd(T)=σ, and the gamma distribution with $\boldsymbol{\theta} = (k, s)^T$, so that E($T$) =$ks$ and sd(T)=ks^2.

For Bayesian inference a prior distribution for $\boldsymbol{\theta}$ is required. However, as experts cannot be expected to make reliable statements about the distribution of $\boldsymbol{\theta}$, an approach to elicit the distribution indirectly has been developed.

2.2 Expert Elicitation

A generic elicitation guideline, developed within the "Sheffield Elicitation Framework"(Oakley and O'Hagan, 2010), has been a major guidance for the design and adaptation of the elicitation procedure described below. Further details can be found in (Arreaza, 2011).

2.2.1 Minimizing cognitive biases

The elicitation guideline was developed keeping minimization of cognitive biases in mind. These are attributable to misunderstanding or discrepancies between the experts' responses and an accurate description of their knowledge (Spetzler and Stael Von Holstein, 1975). The underlying research is comprehensively reviewed in (Ayyub, 2001; Cooke, 1991; Eisenführ et al., 2010; Kynn, 2008; Low Choy et al., 2009; O'Hagan et al., 2006). According to Kynn (2008), following the development of cognitive models to describe the encountered cognitive biases, three dimensions are categorized:

1. Internal consistency (and coherence) is mostly concerned with how well the experts' statements fulfill or contradict the laws of probability.
2. External consistency deals mostly with the ability of a person to control overconfidence in giving probability statements (calibration).
3. Self-consistency (reliability) deals with the variation in between statements when performing repetitive tests.

Often, not the biases or bias categories themselves, but rather more prominent heuristics leading to such biases are cited or even intermixed, e.g.(Kuhnert et al., 2010). Such heuristics are availability, adjustment and anchoring, and representativeness, originally reported and explained by Tversky and Kahneman (1974).
Several measures to avoid distortions to the elicitation data are described in (Cooke, 1991;

Kynn, 2008; Low Choy et al., 2009), among others. From these measures the guidelines to minimize biases (bias category in parenthesis) were compiled:

a) making the desired mathematical implication of questions explicit (internal consistency, especially general additivity),
b) naming frequencies along with probabilities (internal consistency, especially conditional probability, and general additivity),
c) using tools (also visual) or checks during the elicitation procedure, and discussing possible incoherence with the expert to ensure the laws of probability are not violated (internal and external consistency),
d) ordering the questions in such a way that anchoring is avoided, e.g. non-sequentially as in bi-section method (external consistency),
e) calibration of the expert based on training questions which are related to the test questions (internal and external consistency),
f) using different trials, duplicated assessment and different encoding of questions (self-consistency, i.e. reliability of the results), and
g) assessing only non-tacit assumptions, i.e. not asking for extreme probabilities of distributions (internal consistency).

An accurate elicitation procedure including checks and repetition can also lead to the reduction of uncertainty in the stated quantities, whereas adequate preparation enhances consistency and reliability (Low Choy et al., 2009). Kynn (2008), Low Choy, O'Leary et al. (2009) and Oakley and O'Hagan(2010) furthermore emphasize the importance of motivating the experts to participate with diligence.

2.2.2 Quantile elicitation method

When it comes to practical elicitation, it is often not possible to elicit the unknown probability distribution directly, but only observable quantities (Lele and Allen, 2006; O'Hagan et al., 2006). Given that experts could not be expected to estimate the correlated parameters of a bivariate survival model directly, experts were asked to give estimates of the age until which a certain proportion of a pipe cohort is expected to last, i.e. the quantiles of the age distribution (e.g. *"How long does it take until 50 % of the members of this pipe group have been taken out of service?"*). This *quantile elicitation method* is equivalent to the analyst stating probabilities or relative frequencies of a cumulative distribution and the expert estimating the expected age at replacement, see (Rinderknecht et al., 2011) for more details. This is different from the typical application of the quantile elicitation method, because the expert does not state the quantiles of a distribution describing the expert's uncertainty about a quantity to be elicited, but the expert states point estimates of the quantiles of a pipe survival distribution. The elicitation of marginal distributions characterizing the expert's knowledge of all quantiles of the elicited age distributions would be much too demanding and would still leave the problem of missing information about the dependence structure of these marginal distributions.

The quantiles selected for the interview are, in this sequence, the 95 %, 5 %, 50 %, 75% and 25% quantiles, characterizing the tails, the position, and two easily interpretable quantiles in between, which allow for adjustment of the shape of the curve, respectively. Winkler (1967) also suggests this sequence in order to avoid anchoring and adjustment effects. The motivation for 95 % and 5 % quantiles as outer ranges is based on the reported limited ability of experts to correctly express the extreme tails of distributions, e.g. when asking for 99 %

and 1 % quantiles (Alpert and Raiffa, 1982 in (Oakley and O'Hagan 2007)). Experts might find it easier not to specify their opinion with absolute precision. Thus, the elicitation guideline was developed for both precise (point estimates) and imprecise values (stated intervals).

2.2.3 Selection of experts

To ensure enough diversity of opinion and expertise while at the same time avoiding redundancy of information (Ayyub, 2001), eight individual expert interviews, at a duration of approximately two hours each, were performed. The experts were selected following suggestions from the Swiss Gas and Water Association (SVGW/SSIGE). People from different parts of Switzerland with major experience in the fields of planning, construction, operation and maintenance of water supply networks were chosen. All of them carry a higher education degree. An overview of the experts and their specific qualification is given in Table 9, Appendix A.

2.2.4 Choice of pipe groups

It is effective to differentiate the pipe network by material and laying period for pipe survival analysis (Fuchs, 2001; Kleiner and Rajani, 1999; Roscher et al., 2008). Stratification based on other criteria such as diameter, pressure zone, soil conditions etc. is possible, but was not done because this might likely overtax the abilities of the experts and make the elicitation overly complicated. Though diameter can be a useful grouping criterion (Carrión et al., 2010) it was neglected in this study because the focus is on small networks in which diameter differences are small and diameters larger than 300 mm are generally rare. Additionally, the more stratification of data, the smaller the sample sizes for parameter inference. Out of thirteen possible pipe groups formed from material and laying period, five were chosen based on their frequency of occurrence in Swiss water supply networks, familiarity of experts with them, and the time these pipe groups had been in service. They are: grey cast iron (3rd generation only, GI3), ductile iron (1st generation only, DI1), asbestos cement (AC), steel (ST), and polyethylene (PE).

2.2.5 Pre-elicitation information

As preparation for the interviews, all experts were supplied with pre-elicitation information material at least three weeks prior to elicitation. Therein, the purpose of the study and background were stated, along with further information. The information covered the five pipe groups to be elicited, a rough scheme of the elicitation procedure, and suggestions on how to prepare for the interview. The experts were asked to provide feedback on the selected pipe groups. Furthermore, they were requested to thoroughly read through an elicitation example on the service life of an imaginary pump group with formulated questions and potential answers. The reason for pumps instead of pipes was to stay within the domain of the expert, while at the same time avoiding anchoring effects.

2.2.6 Elicitation procedure

First, an elicitation briefing is done. It includes *setting the scene* (purpose and procedure of the interview, expert's expertise, clarification of questions, selection of the four most familiar out of the five proposed pipe groups), *focusing* (characteristics of the pipe groups, motivation), and *training*. Goal of the training is to familiarize the expert with the question layout and to sensitize the experts to possible biases. The training example is the survival of women born in Switzerland in the year 1940 (for the reasoning behind this see Appendix A). Cross-checks with real data (Cordazzo, 2006) help to highlight specific features potentially leading to biases during pipe survival elicitation. Using a different domain for training avoids anchoring of the interviewees.

After this follows the main elicitation, for each pipe group separately. In the beginning, the experience of the expert with the specific pipe group is explored. Then the quantiles are elicited.

Quantities are roughly visualized using 100 paper clips (representing 100 % of the pipe group) and a paper sheet with a time bar. Experts are requested to disregard replacement because of initial laying failures (e.g. within the first year after laying), and replacements following managerial or other considerations not related to age or condition, such as coordinative ground works with other infrastructure providers. This helps to focus on technical lifespans and not on effects of different management decisions.

For a second round, the 75 %, 50 %, and 25 %- quantiles are re-elicited using bets, adjusting the stated ages until the expert is indifferent between the bets. This technique is used to confirm the statements by making the experts think differently about the quantities. After the bets, the elicited values are read out and confirmed with the expert. These checks and repetitions ensure that experts' statements are reliable, consistent and correctly documented. Lastly, the experts are asked for a qualitative description of the imaginary density curve (if possible), which in turn reveals whether it can be assumed unimodal.

Experts are requested to assess half of the pipe groups with imprecise estimates and the other half using precise estimates. At the end, experts are asked for feedback on the difficulty of the interview, and their preference regarding precise and imprecise estimates.

A more detailed description of the entire elicitation procedure is given in Appendix A and in (Arreaza, 2011).

2.3 Derivation of experts' priors of survival function parameters and aggregation

To combine the experts' statements and to construct an intersubjective prior distribution for the survival model parameters, axiomatic, equally-weighted pooling was chosen. Therefore, the elicited quantiles are considered as data to which the survival functions are fitted using nonlinear least squares regression. The resulting estimates and variance-covariance matrix of the survival function parameters were used to parameterize a bivariate lognormal distribution then called *expert prior*. The ages t_k that the expert assigns to the cumulative probabilities π_k are treated as dependent variables. The goodness of fit of the Weibull, lognormal, and gamma distributions can be directly compared based on the residual sum of squares (RSS) because they have the same number of parameters.

At least two aggregation options for combining the different experts into one intersubjective, general prior for Swiss water networks arise (terms are according to the classification of Gelman and Hill (2009) for hierarchical model regression):

A) *Partial pooling:* Fitting a distribution to each expert's estimates separately, and subsequently aggregating these distributions into one prior distribution.
B) *Complete pooling* of all experts' estimates before fitting a distribution to the stated quantiles.

With regard to pipe service life estimation, all experts are considered to be equally credible. They should thus receive equal weights. The possibility of correlation among experts caused by similar training or exchange of experience cannot be excluded (Jouini and Clemen, 1996) and is thus to be accommodated in the aggregated prior.

It is expected that for this example, partial pooling will be more appropriate than complete pooling, because the individual expert priors take better into account the expectedly different underlying environmental conditions in the experts domains throughout Switzerland.

Aggregation option A: partial pooling

The experts gained their knowledge on different water distribution networks whose deterioration is determined by diverse local conditions. Thus, dissimilar distributions can be expected to result from elicitation and fitting. To ensure these different conditions are accordingly reflected in the prior, not only the individual imprecision, but also the in-between-variance of the single experts' fitted distributions shall be considered. The procedure to obtain this prior is:

1. The inverse S^{-1} of the parametric survival function S is fitted to the ages quantified by the expert with a non-linear least squares regression for each expert separately:

$$t_{e,k} \sim S^{-1}\left(\pi_{e,k} | \exp(\theta_e^*)\right) + \varepsilon_{e,k};\ \varepsilon_{e,k} \sim N(0, \sigma_e) \tag{1}$$

Thereof for each expert e, $e = 1 \ldots E$, an approximate multivariate normal distribution $\mathrm{p}_e(\boldsymbol{\theta}^* | \mu_e, \Sigma_e)$ for $\boldsymbol{\theta}_e^* = \ln(\boldsymbol{\theta}_e)$ is obtained with normal distributed error $\varepsilon_{e,k}$. Accordingly, the parameters of the survival distribution $\boldsymbol{\theta}_e$ (Weibull, lognormal or gamma) are lognormal distributed: $\mathrm{p}_e(\boldsymbol{\theta} | \mu_e, \Sigma_e)$. If the expert stated intervals, both endpoints of the intervals are used for the regression.

2. The mixture of all E distributions $\mathrm{p}_e(\boldsymbol{\theta} | \mu_e, \Sigma_e)$ can then be used as prior distribution:

$$p(\boldsymbol{\theta} | \mu_1, \ldots, \mu_E, \Sigma_1, \ldots, \Sigma_E) = \sum_{e=1}^{E} w_e p_e(\boldsymbol{\theta} | \mu_e, \Sigma_e) \tag{2}$$

where w_e is the weight of expert e and $\sum_{e=1}^{E} w_e = 1$.

This model is a mixture of the fitted individual prior distributions of the experts. It is sometimes referred to as the *density version of the linear opinion pool* (Genest and Zidek, 1986).

3. Because the mixture of the experts priors is likely to be multimodal, it is approximated (or rather smoothed) with a two-dimensional lognormal distribution $\tilde{p}(\Theta | \mu, \Sigma)$ that has the mean μ and covariance Σ, calculated as follows:

The (raw) moments of a mixture are the weighted average of the same moments of the component distributions (Frühwirth-Schnatter, 2006). Therefore the first moment (the expected value) of the mixture is

$$E(\boldsymbol{\theta}) = \sum_{e=1}^{E} w_e E_e(\boldsymbol{\theta}) = \sum_{e=1}^{E} w_e \mu_e = \mu \tag{3}$$

where $E_e(\boldsymbol{\theta})$ is the expected value of the distribution of expert e. The second moment is derived from the covariance of each component distribution:

$$E_e(\boldsymbol{\theta}\boldsymbol{\theta}^T) = \Sigma_e + \mu_e \mu_e^T \tag{4}$$

The second moment of the mixture is the weighted average of the second moments of the component distributions:

$$E(\boldsymbol{\theta}\boldsymbol{\theta}^T) = \sum_{e=1}^{E} w_e E_e(\boldsymbol{\theta}\boldsymbol{\theta}^T) \tag{5}$$

Thereof, the covariance of $\tilde{p}(\boldsymbol{\theta}|\mu, \Sigma)$ is calculated:

$$\Sigma = E(\boldsymbol{\theta}\boldsymbol{\theta}^T) - E(\boldsymbol{\theta})E(\boldsymbol{\theta})^T \tag{6}$$

This prior does not necessarily become narrower when more experts are considered. It might even become wider if new experts have gained their knowledge from different systems with other conditions.

Aggregation option B: complete pooling

Another approach is to pool the data beforehand to perform one single regression over all the data at once. This only makes sense if experts are considered as independent measurement devices and if they assess values based on experience from the same or very similar systems. In this case, the variance between experts is interpreted as measurement imprecision. Practically, aggregation consists of one single weighted non-linear least squares regression for all experts together:

$$t_k \sim S^{-1}\left(\pi_k | exp(\boldsymbol{\theta}^*)\right) + \varepsilon_k, \quad \varepsilon_k \sim N(0, \sigma) \tag{7}$$

Quantiles from experts who stated intervals estimates (two measurements per distribution quantile, endpoints of the intervals used) received half the weight compared to quantiles from experts who stated precise estimates (one measurement per quantile). From the weighted non-linear least squares regression a multivariate normal distribution $\mathrm{p}(\Theta^*|\mu_e, \Sigma_e)$ of the estimated parameters $\boldsymbol{\theta}^* = \ln(\boldsymbol{\theta})$ is obtained. Therefrom a log-normal distributed $\mathrm{p}(\boldsymbol{\theta}|\mu_e, \Sigma_e)$ is derived for the parameters $\boldsymbol{\theta}$.

Complete pooling has the effect that the more experts are asked, the smaller the uncertainty of the prior becomes. This is because the number of measurements (experts) increases while the number of parameters to be inferred remains the same.

2.4 Model parameter estimation

2.4.1 Likelihood function for left-truncated and right-censored data

The likelihood function of a model is required for frequentist and Bayesian parameter estimation. The likelihood function expresses the probability density to observe life-spans $t = \{t_1, \dots, t_N\}$ given the parameters $\boldsymbol{\theta}$.

As in most utilities, historic failure and replacement data have not been systematically documented in the studied utility until rather recently. This causes a left-truncation-right-censoring (LTRC) data scheme (Figure 1).

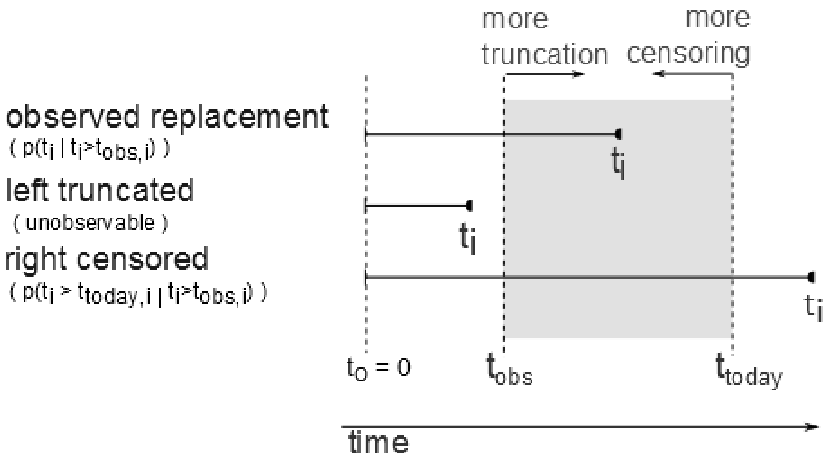

Figure 1: Left truncation and right censoring of a pipe group. The shaded area is the observation window between start t_{obs} and end t_{today} of observations.

Klein and Moeschberger (2003) describe right-censoring as an event which is only observed if it occurs before some pre-specified time, e.g. the end of a study. Consequently, pipes that are still in service at the end of the observation interval t_{today} are *right-censored observations*. *Left truncation* describes a situation where only subjects that have not yet experienced the event enter the study at a particular age and that are followed from this delayed entry time until the event occurs (or until the subject is censored). That means that data of pipes replaced before the start of observations are not available to the analyst, leading to only the more resistant of the pipes being observed.

The probability density to observe an uncensored age t_i of pipe i is written as:

$$p(t_i|T > t_{obs,i}, \boldsymbol{\theta}) = \frac{p(t_i|\boldsymbol{\theta})}{S(t_{obs,i}|\boldsymbol{\theta})} \quad (8)$$

where $t_{obs,i}$ is the age of the pipe i at the beginning of the observation period.

In situations where the end of lifetime could not be observed the likelihood for a single pipe becomes:

$$P(t_i > t_{today,i}|t_i > t_{obs,i}, \boldsymbol{\theta}) = \frac{S(t_{today,i}|\boldsymbol{\theta})}{S(t_{obs,i}|\boldsymbol{\theta})} \quad (9)$$

where $t_{today,i}$ denotes the age of pipe i at the end of the observation period.

A censoring indicator δ_i allows for a short notation for the likelihood for all N pipes. δ_i equals zero if the datum is censored and one if uncensored. With this and the assumption that the pipes are independent the joint likelihood function for all pipes is:

$$\mathrm{p}(\mathbf{t}, \boldsymbol{\delta}|\, \boldsymbol{\theta}) = \prod_{i=1}^{N} \left(\frac{p(t_i|\boldsymbol{\theta})}{S(t_{obs}|\boldsymbol{\theta})}\right)^{\delta_i} \left(\frac{S(t_{today}|\boldsymbol{\theta})}{S(t_{obs}|\boldsymbol{\theta})}\right)^{1-\delta_i} \quad (10)$$

This likelihood function is used for frequentist and Bayesian parameter inferences.

2.4.2 Frequentist parameter inference

Maximum likelihood estimation (MLE) is a common method to infer model parameters. The parameters that maximize the likelihood function for given data are used as best estimate. Large sample size properties of MLE allow the estimation of the variance-covariance matrix of the parameters from the inverse expected Fisher information matrix (Harrell, 2001), see Appendix B for more details.

Practically, this is done by a search through the parameter space by different optimization algorithms implemented in the R package *optimx* (Nash and Varadhan, 2011). The parametric models fitted are Weibull, lognormal, and gamma, as described in section 2.1. Multiple runs with different initial parameter values were performed to ensure stable estimates.

2.4.3 Bayesian inference

The aim of Bayesian inference is to update the prior probability distribution $p(\boldsymbol{\theta})$ with observed data $\{\mathbf{t}, \boldsymbol{\delta}\}$. The resulting posterior probability distribution is calculated with the Bayes theorem:

$$\mathrm{p}(\boldsymbol{\theta}|\mathbf{t},\boldsymbol{\delta}) = \frac{\mathrm{p}(\mathbf{t},\boldsymbol{\delta}|\,\boldsymbol{\theta})\mathrm{p}(\boldsymbol{\theta})}{\int \mathrm{p}(\mathbf{t},\boldsymbol{\delta}|\,\boldsymbol{\theta}')\mathrm{p}(\boldsymbol{\theta}')\,d\boldsymbol{\theta}'} \tag{11}$$

More in-depth information on Bayesian inference can be found in (Gelman et al., 2004). In this study, informative priors were derived based on expert elicitations and two different aggregation options (see sections 2.2 and 2.3). It can be shown that the choice of the prior distribution strongly influences the posterior result and is thus to be carefully chosen (Berger, 1990; Gelman et al., 2004).

The posterior distribution is derived by means of iterative Markov-Chain Monte-Carlo sampling (MCMC) with 6000 draws. The first 1000 draws are discarded as burn-in period and the acceptance rate is kept between 0.3 and 0.4 with the help of an adaptive sampler (Scheidegger, 2011; Vihola, 2011).

2.4.4 Graphical validation with a non-parametric survival estimate

Wayne (2004) suggests the use of a non-parametric survival estimate to visualize the fit of the parametric model. A Nelson-Aalen estimator adapted for LTRC data (also referred to as *extended Nelson estimator*) is used as described in Pan and Chappell (1998) and applied to pipe survival in Carrión et al. (2010). More details are given in the Appendix B.

The Nelson-Aalen estimator is capable of dealing with small sample sizes and can handle both censored and incomplete data as in our case of LTRC pipe survival (Klein and Moeschberger, 2003).

2.5 Utility data

The data used in this study consists of replacement records from a large Swiss water utility. Only pipe groups that were used in the prior elicitations were extracted from the provided pipe inventory. Reliable recording of pipe replacement started in this utility in 2000, so that only replacement entries between 2001-01-01 and 2010-12-31 were used for inference. The characteristics of the pipe groups are summarized in Table 1.

The effect of a decreasing amount of data (sample size) on parameter estimation is simulated by randomly reducing the available data to 500, 300, 150, and 50 pipes. These numbers correspond to the amount of data expected in mid-size or small water utilities for which Bayesian combination of expert opinion with local data is proposed. The ratio of replaced pipes to the overall number of pipes is kept constant. Shorter observation periods (more truncation, Figure 1) are also studied. They are accounted for by shifting the start of observation to 2003, 2005, 2007, 2008, and 2009 and removing replacement entries before this date, respectively.

Table 1: Summary characteristics of the examined pipe groups. Legend: GI3- 3rd generation grey cast iron (1930-1965), ST- steel, DI1- 1st generation ductile iron(1965-1980), PE- polyethylene, AC- asbestos cement incl. Eternit.

group	pipes (no.)	total length (km)	laying date (min-**med.**-max)	removal year (min-**med.**-max)	inner diameter # no. in mm	of which replaced
GI3	1295	104.5	1930-**1951**-1965	2001-**2005**-2010	[0- **100**[: 181 [100-300[: 854 [>300): 260	571 (**44.1 %**)
DI1	1009	87.45	1968-**1976**-1980	2001-**2007**-2010	[0- 100[:12 [100-300[: 865 [>300): 132	134 (**13.3 %**)
AC	153	22.45	1900-**1958**-1977	2001-**2006**-2010	[0- 100[: 31 [100-300[: 117 [>300): 5	38 (**24.8 %**)
ST	991	89.83	1875-**1965**-2009	2001-**2004**-2010	[0- 100[: 39 [100-300[: 594 [>300): 358	318 (**32.1 %**)
PE	195	18.02	1972-**2004**-2010	2006-**2008**-2009	[0- 100[: 25 [100-300[: 140 [>300): 30	6 (**3.1 %**)

3 Results and Discussion

3.1 Expert elicitation

The developed guideline was used for interviews with eight experts, numbered E1… E8. They provided estimates for grey and ductile cast iron (GG3 and DI1), whereas for steel (ST), polyethylene (PE), and asbestos cement (AC) only three, five, and six experts, respectively, provided their opinion. E4 estimated the service life of three pipe groups only, because in his network domain the two other materials were too rarely used. The obtained estimates are given in detail in Table 8 (see Appendix A) and summarized in Table 2. Two estimates were not considered in the later analysis: E1 for polyethylene, because the expert was highly uncomfortable about giving estimates for this material and could not imagine in which way it might deteriorate; E4 for steel pipes, because they were declaredly based on a past decision of this expert's utility to replace all steel pipes within only five years.

During the feedback at the end of the interview, five out of eight experts said they favored giving intervals instead of point estimates (Arreaza, 2011).

Table 2: Summary statistics for stated quantile values (ages) from experts for given cumulative probabilities, mean = arithmetic mean of stated ages, sd= standard deviation. Pipe groups are explained in 2.2.4.

group	**GI3**		**DI1**		**AC**		**ST**		**PE**	
probability	mean	sd	mean	sd	mean	sd	mean	sd	mean	sd
0.05	38.3	15.3	22.9	8.5	36.3	14.1	32.0	8.4	38.8	22.5
0.25	54.2	16.4	37.9	12.0	64.4	19.0	48.2	8.9	68.1	38.0
0.5	78.3	15.3	55.0	9.8	81.3	22.5	60.4	11.4	86.9	35.1
0.75	90.0	16.4	68.6	9.7	98.1	26.7	74.6	19.5	98.8	30.8
0.95	105.4	18.3	81.1	11.5	115.6	30.6	87.0	24.9	117.5	44.0

The summary statistics in Table 2 show that the quantile estimates between pipe groups, visible from the quantile means, are clearly different. With regard to the quantile standard deviations, not only a pronounced difference between materials is visible, but also an increasing uncertainty towards the upper quantiles. E4 and E5 gave distinctively lower estimates than other experts (Table 8 in Appendix A). Contrarily, estimates from E8 were consequently larger for all pipe groups. These visible differences between material groups and single expert values indicate the experts' awareness and ability to differentiate the ageing behavior of the selected pipe groups. E4 and E5 named specific influences, such as strong deficits in laying or bedding quality, or difficult environmental conditions that could explain lower estimates (Table 9 in Appendix A). The longer lifetime suggested by E8 might also stem from anchoring to rather high values established by a former study this water supplier had commissioned. Other than this, the additional information given by the experts roughly allows us to explain differences between experts' statements and is thus considered as reflection of the encountered variability of conditions in the utility networks.

The usefulness of an expert is usually judged upon his contribution to an increase in knowledge. Measuring this usefulness based on the precision of statements or contribution to noise reduction, e.g. (Lele and Allen, 2006; Runge et al., 2011) is not appropriate in a case like ours. Rich knowledge is not necessarily equivalent to a high density of the mean and little spread of the fitted expert distribution. If the expert bases his knowledge on a variety of different water networks (or other objects of study), he might well accommodate this in more imprecise statements. Also, personal confidence and interrogation layout may play a role. An overconfident expert is likely to state shorter intervals than necessary to reflect his confidence levels (Speirs-Bridge et al., 2010). In this study, experts were encouraged to adequately consider their uncertainty in giving interval statements. The more useful expert is thus the expert stating wide enough intervals that contain his uncertainty about the quantity.

3.2 Parametric model identification

Non-linear least squares regressions were performed with the Weibull, lognormal, and gamma parametric models over the individual experts' statements. The goodness of fit measures were calculated (see residual sum of squares (RSS) in Table 11, Appendix B). The Weibull distribution provides the best fit for 23 out of the 28 single expert assessments. In one case, the RSS of the Weibull distribution is equal to the lognormal (E2 for polyethylene), and only inferior to the gamma or lognormal distribution in four single regressions (E8 for grey cast iron and polyethylene, as well as E5 in the case of ductile cast iron).

Similarly, maximum likelihood estimations (MLE) were done with the available pipe replacement data (see description in 2.4) for all three distribution models. Table 3 shows that the Weibull distribution for DI1, ST, and PE leads to smaller likelihoods than the lognormal or gamma distributions. In the case of GI3 and AC, the lognormal model fits the data slightly better.

Table 3: Obtained parameters and likelihood values of pipe group data from a large Swiss water supplier. Bold numbers indicate the model that maximizes the likelihood value log $\mathbf{p}(\mathbf{t},\boldsymbol{\delta}|\,\boldsymbol{\theta})$. Pipe groups are explained in 2.2.4.

	Weibull distribution			lognormal distribution			gamma distribution		
	α	β	$\ln p(\mathbf{t},\boldsymbol{\delta}\|\,\boldsymbol{\theta})$	μ	σ	$\ln p(\mathbf{t},\boldsymbol{\delta}\|\,\boldsymbol{\theta})$	k	s	$\ln p(\mathbf{t},\boldsymbol{\delta}\|\,\boldsymbol{\theta})$
GI3	2.07	45.54	-2206.70	3.88	0.33	**-2204.74**	6.96	6.98	-2514.00
DI1	5.48	47.45	**-686.89**	3.86	0.31	-687.48	13.23	3.78	-687.13
AC	2.46	60.90	-172.59	4.07	0.29	**-170.50**	12.34	6.26	-171.27
ST	2.46	48.77	**-1259.96**	3.73	0.40	-1275.74	6.38	7.08	-1263.49
PE	1.81	65.72	**-36.29**	4.65	1.45	-36.82	1.85	58.73	-36.45

Though the Weibull likelihoods for GI3 and AC are only slightly larger than the lognormal, important deviations between the two models exist. This can be visualized by graphically comparing the nonparametric extended Nelson-Aalen estimation (see section 2.4.4) with the parametric models (Figure 2).

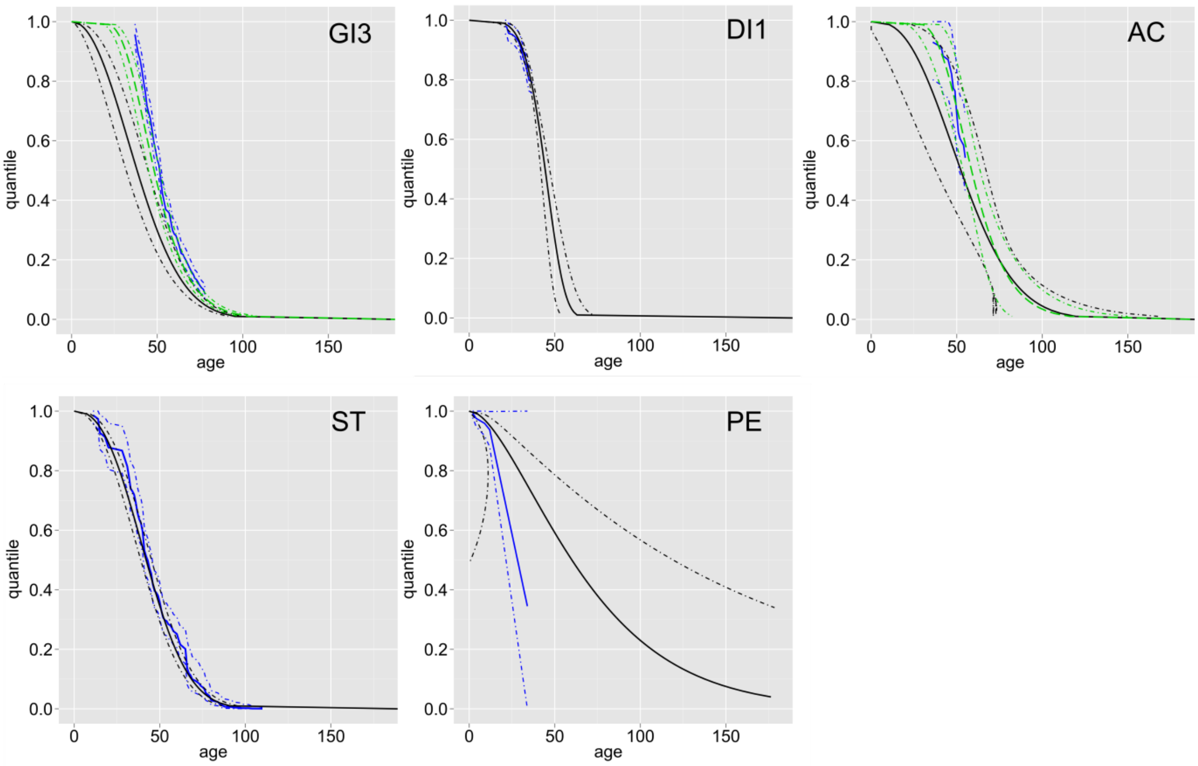

Figure 2: Comparison of a Weibull fit (black solid lines) and a nonparametric fit (blue step lines) for five pipe groups. The lognormal fit (green dashed lines) was added to the plot where it provided the best fit. The dash-dotted lines correspond to the 95 % confidence intervals of the mean (solid line).

The advantage of a nonparametric estimation is that it reflects the lifetime distribution more realistically within the observed time interval and can be used as a control for the possible parametric models. If both parametric and nonparametric models are overlaid, the parameter estimates satisfactorily approximate the „true" parameter values. Nevertheless, the nonparametric estimation model cannot replace parametric models because the outcome is a discrete estimate of the survival function (Coolen, 1996). It cannot be extended to unobserved intervals in time, something especially important for materials with a rather short history (for example for newer materials such as ductile cast iron or polyethylene). This is inconvenient for forecasting and especially for error propagation. Theoretical parametric models can be found that allow for extrapolation outside the data range and error propagation in combination with coupled predictive models. The structural deficiencies of the model are reflected in larger parametric uncertainty.

The Weibull model is best at approximating the nonparametric survival curves of DI1 and ST, but inferior to a lognormal distribution for GI3 and AC. Therein, the tail behavior of the lognormal distribution allows for an overall steeper survival curve whereas the Weibull model leads to underestimation of about 40 % to 50 % of the studied GI3 and AC cohort survival. The graph for polyethylene shows that the available data are clearly not sufficient to infer a trustworthy predictive distribution by means of MLE. This is also reflected in the uncertainty of the estimated parameters as specified in Table 4 (see MLE for all data), given a Weibull survival model for all materials. Despite the better fit of a lognormal model to GI3 and AC data, in the following sections regression and inference of parameters from the utility data is done for the Weibull model, unless otherwise stated.

3.3 Expert prior aggregation

Table 4 shows the mean parameter estimates of the Weibull shape ($\hat{\alpha}$) and scale ($\hat{\beta}$) parameters and corresponding uncertainty measures using a lognormal error distribution, $sd(\hat{\alpha})$ and $sd(\hat{\beta})$ respectively, for individual and pooled experts (see section 2.3). The parameters reflect the differences between pipe materials and experts as described for the elicitation results in 3.1. The parameter estimation of the complete and partial pooling prior results in similar means. Judging from a comparison of the aggregated scale parameters, the 63.2 % quantile, AC pipes are believed to have the longest service life, followed by GI3, then PE, ST, and DI1 pipes. These observations are in line with survival estimates from the literature for German and Austrian water utilities (Fuchs, 2001; Roscher et al., 2005; Trujillo Alvarez, 1995), where AC and GI3 are usually judged most durable and DI1 least durable of the five considered pipe groups (Table 10, Appendix A). The impact of smoothing the experts' mixture on the prior used for inference is visible in Figure 3. The simple mixture according to equation (3) has a multi-modal density whereas the smoothed mixture, equation (4), is unimodal.

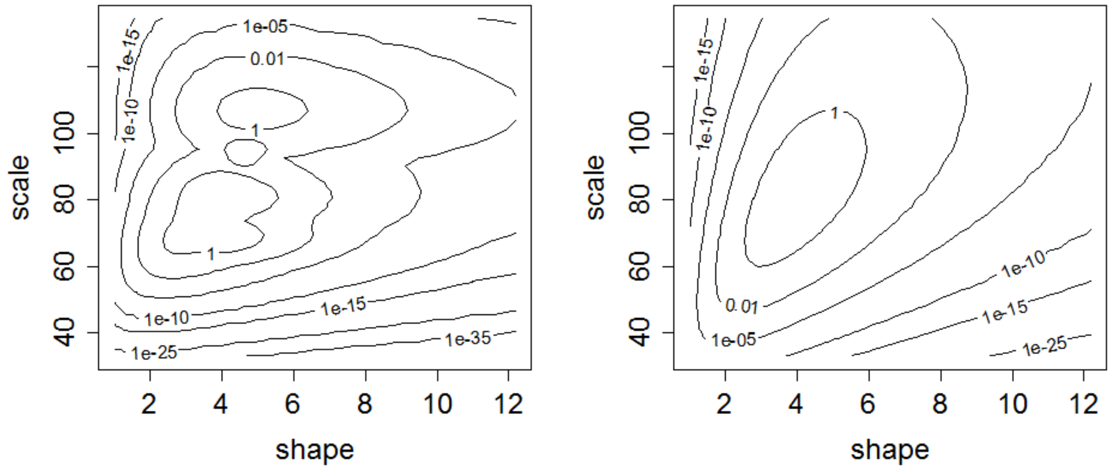

Figure 3: Bivariate probability density distribution of the aggregated prior (partial pooling) before smoothing (left, multimodal) and after smoothing (right, unimodal) for GI3.

Figure 4 shows the expert statements and mean survival function including 95 % confidence intervals for partial pooling as compared to the confidence intervals for complete pooling (exemplary for GI3). Unsurprisingly, the partial pooling yields larger standard deviations. This is because partial pooling incorporates the in-between variances, thus allowing for a better representation of the underlying differences in the experts' domains. The variance is not simply attributable to the experts' measurement error. As discussed in section 3.1 and 3.2, not only did the experts state diverse reasons for different ageing behaviors in their utilities, these differences are also reflected in their statements. It is important to make clear that the experts do not have to agree in this context. If the expert judgments represent different distributions, of which each describes the underlying pipe survival in the expert's water utility, both in-between-differences and individual uncertainty are important sources of information. As opposed to this, complete pooling does not accommodate the in-between variance. It considers all judgments as measurements resulting from assessment of the same underlying distribution and its parametric uncertainty reproduces the error attributable to lack of fit from the model.

Although presumably obvious in a Bayesian learning framework, the aggregation approach chosen is not Bayesian. This is mainly owing to the need to formulate an unbiased prior to be combined with the individual priors, and that furthermore accurately considers the dependence structure between experts (Clemen and Winkler, 1999; O'Hagan et al., 2006). The definition of such a supra-prior is not only demanding, but also the analyst's independence from the experience gained during elicitation and data analysis cannot be expected.

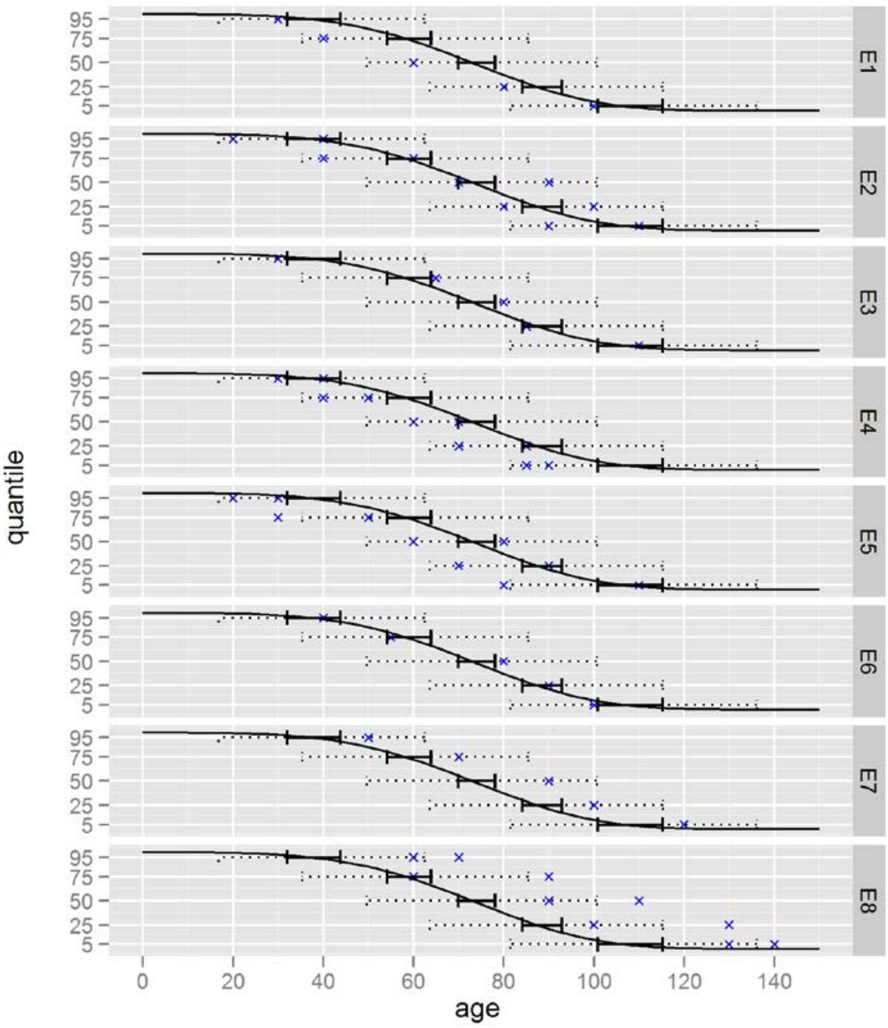

Figure 4: Comparison of GI3 priors and estimates from experts. Blue crosses represent quantile values as stated by the expert indicated on the right edge (E1…E8). Solid error bars give the 95 % confidence intervals for complete pooling, dotted error bars for partial pooling. The survival curve is calculated from the mean parameters ($\hat{\alpha}$= 3.97; $\hat{\beta}$= 81.22) of the partial pooling prior, see Table 4.

Table 4: Results from 1) non-linear least squares regression over experts E1…E8 ("Single experts"), 2) parameters obtained for the two aggregation methods complete pooling and partial pooling ("Aggregation"), and 3) maximum likelihood inference for all data, shortened observation windows, and artificial data reductions ("MLE"). The survival model is a Weibull distribution with parameters S(θ)= (α,β)^T.

		Grey cast iron (1930-64)				Ductile iron (1965-80)				Asbestos cement				Steel				Polyethylene			
		$\hat{\alpha}$	$\hat{\beta}$	sd($\hat{\alpha}$)	sd($\hat{\beta}$)	$\hat{\alpha}$	$\hat{\beta}$	sd($\hat{\alpha}$)	sd($\hat{\beta}$)	$\hat{\alpha}$	$\hat{\beta}$	sd($\hat{\alpha}$)	sd($\hat{\beta}$)	$\hat{\alpha}$	$\hat{\beta}$	sd($\hat{\alpha}$)	sd($\hat{\beta}$)	$\hat{\alpha}$	$\hat{\beta}$	sd($\hat{\alpha}$)	sd($\hat{\beta}$)
Single experts	**E 1**	2.96	69.16	0.28	2.18	2.04	57.15	0.19	2.40	3.49	99.24	0.51	4.27	-	-	-	-	-	-	-	-
	E 2	3.68	78.20	0.73	4.45	2.53	51.51	0.41	3.21	4.52	99.48	0.84	4.56	-	-	-	-	4.00	85.21	0.23	1.30
	E 3	3.77	82.40	0.53	3.26	4.05	66.02	0.53	2.30	3.87	110.9	0.42	3.27	-	-	-	-	-	-	-	-
	E 4	4.13	68.52	0.54	2.33	3.00	47.91	0.51	2.72	-	-	-	-	-	-	-	-	4.54	54.91	0.24	0.70
	E 5	3.18	69.62	0.64	4.50	2.97	58.00	0.40	2.60	2.93	55.99	0.14	0.93	-	-	-	-	2.38	44.62	0.22	1.67
	E 6	4.38	80.52	0.56	2.59	4.08	58.73	0.54	2.03	-	-	-	-	-	-	-	-	4.71	97.63	1.02	5.06
	E 7	4.61	94.43	0.21	1.03	3.44	60.42	0.41	2.17	3.95	63.90	0.64	2.80	3.67	73.25	0.50	2.87	-	-	-	-
	E 8	5.05	107.2	0.93	4.46	3.88	73.17	0.51	2.64	-	-	-	-	3.88	73.17	0.51	2.64	4.85	110.5	0.42	2.24
Aggre-gation	**complete pool**	3.92	81.31	0.36	2.05	3.16	59.34	0.22	1.30	3.79	86.01	0.60	3.83	3.75	73.19	0.34	1.84	4.22	78.78	1.02	4.90
	partial pool	3.97	81.22	0.91	12.70	3.25	59.11	0.88	7.83	3.75	86.05	0.74	24.14	3.77	73.21	0.52	2.76	4.11	74.40	1.21	26.73
MLE	**All data**	2.07	45.54	0.20	3.14	5.48	47.45	0.74	1.82	2.46	60.90	0.85	6.21	2.46	48.77	0.13	1.58	1.81	57.38	0.58	33.60
	≥ 2003	1.79	38.88	0.24	5.19	5.85	46.60	0.99	1.87	2.17	59.79	1.41	15.71	2.92	53.48	0.18	1.73	-	-	-	-
	≥ 2005	2.37	49.81	0.32	4.23	5.20	47.85	1.27	2.52	3.21	2866	0.52	2272279	2.97	55.45	0.24	2.13	-	-	-	-
	≥ 2007	2.49	49.81	0.42	5.27	3.98	51.10	1.96	4.34	-	-	-	-	2.79	51.95	0.30	2.84	1.21	145.06	0.64	359.62
	≥ 2008	2.37	52.85	0.59	8.03	4.68	50.45	2.64	4.89	-	-	-	-	2.96	53.52	0.39	3.40	-	-	-	-
	≥ 2009	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	500 pipes	1.95	43.81	0.33	5.88	3.66	53.94	1.08	5.24	-	-	-	-	2.42	47.68	0.18	2.21	-	-	-	-
	300 pipes	2.22	48.04	0.41	5.97	5.51	47.67	1.40	3.40	-	-	-	-	2.61	52.01	0.23	2.83	-	-	-	-
	150 pipes	2.91	55.32	0.65	5.53	3.24	60.22	2.44	13.82	1.81	56.25	0.99	16.51	2.37	46.70	0.30	4.07	2.39	39.29	0.70	13.71
	50 pipes	2.39	49.72	1.30	18.52	5.64	51.45	5.13	12.36	6.25	63.05	3.01	3.61	3.42	54.59	0.61	5.04	-	-	-	-

3.4 Maximum likelihood estimation from data

A maximum likelihood estimation (MLE) of the parameters of the survival function was done with the available pipe replacement data. The underlying model used is Weibull, the same model as used for describing the prior distributions of the experts. However, it must be noted that MLE of the parameters is independent of these priors.

For GI3, DI1, and ST, reasonably certain parameter estimates were derived (Table 4). Regarding AC and PE, a set of parameters was obtained as well, but with larger uncertainty due to the smaller number of data. The parameters from MLE show substantial differences as compared to the aggregated experts. Taking the Weibull scale parameter ($\hat{\beta}$) as representative for the age reached by 63.2 % of pipes in this group, basically the same ranking as given by the experts is observed: AC reaches higher ages than (in this order) PE, ST, DI1, and GI3. The exception is GI3, which is second most durable according to the aggregated experts, but least durable if only inferring from the data. Compared to the characteristic lifetime inferred by MLE from data, the aggregated experts estimates are approximately 11 (DI1) to 35 (GI3) years longer.

The results from randomly reducing the data to 500, 300, 150, and 50 pipes, while keeping the ratio of replaced to in-service pipes constant, demonstrate that the fewer data are available, the more uncertain the parameter estimates get. Analogously, increasing truncation when reducing the observation period to seven, five, three, two, and one year(s) leads to increasingly uncertain parameter estimates. This truncation is mirrored in the diminishing ratio of documented pipe replacements to pipes in service, given in Table 5.

Table 5: Effect of truncation on the ratio of replaced pipes to pipes in service. Ratios for which no MLE parameter estimates were obtained are highlighted.

	GI3	DI1	AC	ST	PE
2000	0.44	0.13	0.25	0.32	0.03
2003	0.36	0.12	0.18	0.23	**0.03**
2005	0.26	0.10	0.15	0.16	**0.03**
2007	0.18	0.06	**0.08**	0.11	0.05
2008	0.10	0.04	**0.07**	0.07	**0.00**
2009	**0.04**	**0.02**	**0.04**	**0.03**	**0.00**

For less than two years of observations, no parameters could be identified. As visible in Table 5, this can be attributed to ratios of less than ca. 10 % (GG3), 4 % (DI1), and 7 % (ST). Records of less than five years for AC (ratio of approx. 15 %), did not allow for any reliable parameter estimates, possibly explicable by the small sample size (153, Table 1). The scale parameters for PE and AC ($\hat{\beta}$) can only be estimated with very large uncertainty sd($\hat{\beta}$), an effect possibly caused by the small sample size and increasing truncation leading to a flat likelihood surface, such that the parameters are difficult to estimate.

Consequently, if a utility has only recently started reliably recording pipe replacement (high truncation), or if the number of pipes in the network is small, no reliable parameters can be found with MLE.

Furthermore, the considerable differences between the distribution parameters inferred from expert statements and utility data cannot merely be attributed to more extreme local conditions, but rather are an effect of local management strategies on pipe survival. For example DG1 was often referred to by the interviewed experts (Table 9) as the “problem child”. Lacking corrosion protection and aggravating exposure to electrical currents from households grounding electric appliances on the water pipelines has led to major *pro-active* replacement campaigns. It could furthermore explain the much steeper survival curve of this material in the investigated data set. From consultation with a local expert from whose utility the data was taken, a substantial fraction of pipes is usually replaced before the end of its technical service life, owing to coordination efforts by different network utilities. For instance, the rehabilitation of the sewer system often requires the removal of the above lying water supply pipes. If a substantial part of the replacement is for reasons other than technical end of life, the consequence are considerably shorter observed lifetimes.

The available data neither indicate the reason nor the condition of the replaced pipes. The available survival data does not allow for the description of ageing-induced technical (or structural) service life, as it is managerial replacement which is recorded. This means that the expert prior and the data to be combined by Bayesian inference describe two unalike phenomena: the prior describes technical ageing and replacement according to the experts’ experience, whereas the data represent the ageing and replacement process distorted by managerial replacement strategies.

The easiest way to avoid this discrepancy is to solely use survival data of pipes that were replaced due to technical end-of-life, thereby creating congruent information pools.

Nevertheless, this problem easily develops into a philosophical one. Someone may perhaps anyway distrust experts’ capacity of differentiation between observed managerial replacement and the replacement caused by structural ageing processes. This would mean that the obtained prior and the recorded data do not contradict each other. But who can be trusted if not the interviewed experts who show themselves very able to give such estimates?

Otherwise, the analyst could try to rectify the recorded managerial replacement with a correction factor, thus making it comparable to the experts prior and suitable to describe technical replacement needs. Additional prior estimates for all pipe groups, however, would be needed; something which is already hard to expect from the current managerial strategy. But how to quantify the impact of former management? A way out seems to be basing models on events more capable of describing technical/structural ageing, e.g. failure instead of replacement records. Such records are increasingly available and used in different modeling approaches. Instead of priors for technical replacement, priors about the time of a failure or between failures of different orders could be elicited from experts, if not from other utility data. This does not mean, however, that problems induced by former management such as replacement biases in the data can be completely avoided, so that adaptations to the existing models might be necessary.

Although the obtained prior and data might not entirely describe the same phenomenon resulting in pipe replacement, it is nonetheless important to show the performance of Bayesian inference and the prior impact on the posterior. Even if expert and data can be reconciled, prior data conflicts may arise, for example owing to especially unfavorable conditions inducing faster replacement in the study utility than predicted by the experts.

Table 6: Resulting posterior parameters from Bayesian inference with data from a large water utility in Switzerland. The survival model is a Weibull model with parameters $\theta = (\alpha, \beta)^T$. Parameters are given for the case of inference with all data, and artificial data reductions for a prior either derived from complete pooling or partial pooling. MLE gives the parameters obtained by frequentist MLE from all data.

	Grey cast iron (1930-64)				**Ductile iron (1965-80)**				**Asbestos cement**				**Steel**				**Polyethylene**			
	$\hat{\alpha}$	$\hat{\beta}$	sd($\hat{\alpha}$)	sd($\hat{\beta}$)	$\hat{\alpha}$	$\hat{\beta}$	sd($\hat{\alpha}$)	sd($\hat{\beta}$)	$\hat{\alpha}$	$\hat{\beta}$	sd($\hat{\alpha}$)	sd($\hat{\beta}$)	$\hat{\alpha}$	$\hat{\beta}$	sd($\hat{\alpha}$)	sd($\hat{\beta}$)	$\hat{\alpha}$	$\hat{\beta}$	sd($\hat{\alpha}$)	sd($\hat{\beta}$)
a) With complete pooling prior																				
Prior	*3.92*	*81.31*	*0.36*	*2.05*	*3.16*	*59.34*	*0.22*	*1.30*	*3.79*	*86.01*	*0.60*	*3.83*	*3.75*	*73.19*	*0.34*	*1.84*	*4.22*	*78.78*	*0.34*	*0.27*
All data	3.36	62.10	0.13	0.84	3.06	58.10	0.16	1.01	2.93	77.36	0.34	3.04	2.97	59.98	0.12	0.98	1.95	74.94	0.24	4.77
500 pipes	3.50	67.48	0.20	1.20	3.04	58.55	0.16	1.12	-	-	-	-	3.03	63.44	0.16	1.29	-	-	-	-
300 pipes	3.61	70.91	0.24	1.46	3.08	58.76	0.18	1.17	-	-	-	-	3.25	67.01	0.19	1.43	-	-	-	-
150 pipes	3.70	74.66	0.28	1.66	3.09	59.01	0.19	1.17	2.83	77.70	0.33	2.93	3.22	68.91	0.22	1.56	2.16	74.82	0.28	4.43
50 pipes	3.71	78.54	0.31	1.83	3.14	59.26	0.19	1.31	3.32	81.90	0.45	3.53	3.66	71.50	0.29	1.75	3.48	76.11	0.68	4.41
0 pipes	3.90	81.32	0.36	2.11	3.16	59.36	0.22	1.29	3.81	86.03	0.59	3.72	3.74	73.26	0.33	1.84	4.22	78.85	1.02	5.03
b) With partial pooling prior																				
Prior	*3.97*	*81.22*	*0.91*	*12.70*	*3.25*	*59.11*	*0.88*	*7.83*	*3.75*	*86.05*	*0.74*	*24.14*	*3.77*	*73.21*	*0.52*	*2.76*	*4.11*	*74.40*	*0.40*	*1.55*
All data	2.26	48.33	0.17	2.34	3.88	52.61	0.52	2.31	2.94	63.51	0.40	3.77	2.84	56.06	0.12	1.14	2.26	43.31	0.28	7.18
500 pipes	2.38	50.55	0.23	3.09	3.01	56.85	0.45	3.19	-	-	-	-	2.93	58.85	0.16	1.51	-	-	-	-
300 pipes	2.62	53.70	0.29	3.25	3.27	55.64	0.52	3.36	-	-	-	-	3.11	63.43	0.19	1.67	-	-	-	-
150 pipes	3.11	58.12	0.43	3.47	2.87	56.67	0.50	3.73	2.74	64.10	0.36	4.30	3.05	65.62	0.25	2.15	2.46	42.40	0.33	6.90
50 pipes	3.07	60.40	0.49	4.82	2.96	56.80	0.57	4.85	3.50	65.90	0.60	5.77	3.63	69.66	0.40	2.53	3.35	48.47	0.65	10.66
0 pipes	3.95	80.78	0.95	12.82	3.26	59.10	0.86	7.79	3.74	86.21	0.73	23.80	3.79	73.15	0.52	2.74	4.08	73.84	1.17	25.87
MLE	2.07	45.54	0.20	3.14	5.48	47.45	0.74	1.82	2.46	60.90	0.85	6.21	2.46	48.77	0.13	1.58	1.81	57.38	0.18	2.37

3.5 Bayesian inference

Bayesian inference was performed using the pooled results from expert elicitation as prior and the pipe replacement data. This was also done for different amounts of data. The resulting posterior mean parameters and standard deviations are given in Table 6. Important remark: Following the discussion in section 3.4, the survival functions from Bayesian inference described below are not valid for the prediction of technical rehabilitation demand in the studied utility, as the data is about actual replacement that includes replacement for reasons other than technical aging. Nevertheless, the discussion regarding the suitability of the MLE and Bayesian approaches for scarce data situations is valid, as are the observations regarding different prior aggregations.

The most important observation from Table 6 is that in contrast to MLE, reasonably certain parameter estimates could be determined even for small numbers of pipes. This applies especially to pipe groups with few records where MLE returns parameter estimates with high uncertainty (e.g. AC and PE with 150 or less records, or 50 pipes for GI3 and DI1). There, the posterior distributions are more informative than any of the obtainable distributions from MLE alone. For larger data sets however, MLE does lead to reliable parameter estimates which are closer to the mean parameters obtained from MLE of the full data set, making Bayesian combination of data and expert knowledge unnecessary.

Furthermore, as typical for Bayesian inference, the posterior mean values lie between the prior and the MLE (utility data) mean. Also, the standard deviations of the inferred posteriors are smaller than the prior and MLE standard deviations (except for DI1 and PE, see Table 4), meaning that something could be learned from the data. Analogous to MLE, the uncertainty of the parameter estimates increases with decreasing number of pipes used for inference. The smaller the number of pipes, the more the posterior approximates the expert prior (see Figure 5).

In Table 6 and Figure 5, the influence of the prior distribution on the posterior parameter estimates is clearly visible. The wider, partial pooling prior naturally causes wider posterior distributions than the posterior calculated from the more precise complete pooling prior. It also approximates the mean parameters obtained from MLE more closely than the posterior obtained with the complete pooling prior (compare the vertical lines in Figure 5 representing the means of MLE and posterior). In some cases however, the complete pooling posterior coincidentally performs slightly better (being nearer to the MLE estimate).

Regarding the influence of prior aggregation on the posterior, the effect of prior choice is exemplarily shown for 3rd generation grey cast iron in Figure 5. The posterior obtained from inference with the more uncertain partial pooling posterior is notably closer to the MLE than the posterior obtained from inference with the complete pooling prior (Table 6). This effect is attributable to arising prior-data conflicts resulting from discordant information from the observed data and the prior (Bousquet, 2008, among others). Even though the parameter means ($\hat{\alpha}$, $\hat{\beta}$) of the complete and partial pooling prior are nearly identical, the larger standard deviations ($sd(\hat{\alpha})$, $sd(\hat{\beta})$) of the partial pooling prior reduce the conflict with the data. No satisfactory approximation towards the MLE shape parameter ($\hat{\alpha}$) of DI1, however, was achieved with any of the two priors. The conceptually more appropriate partial pooling prior (see 3.3) leads to a compromise between the prior and the data (posterior $\hat{\alpha} \approx 3.88$ as opposed to prior 3.25 and MLE 5.48) and the complete pooling to hardly any change (posterior $\hat{\alpha} \approx 3.06$, prior ≈ 3.16). Checking for conflicts in the distributions in Figure 6 (dotted and dashed lines) as opposed to the distribution obtained from MLE (shaded), it is visible that partial prior and data distributions only partly overlap with approximately similar densities, but the complete pooling prior and the data widely disagree.

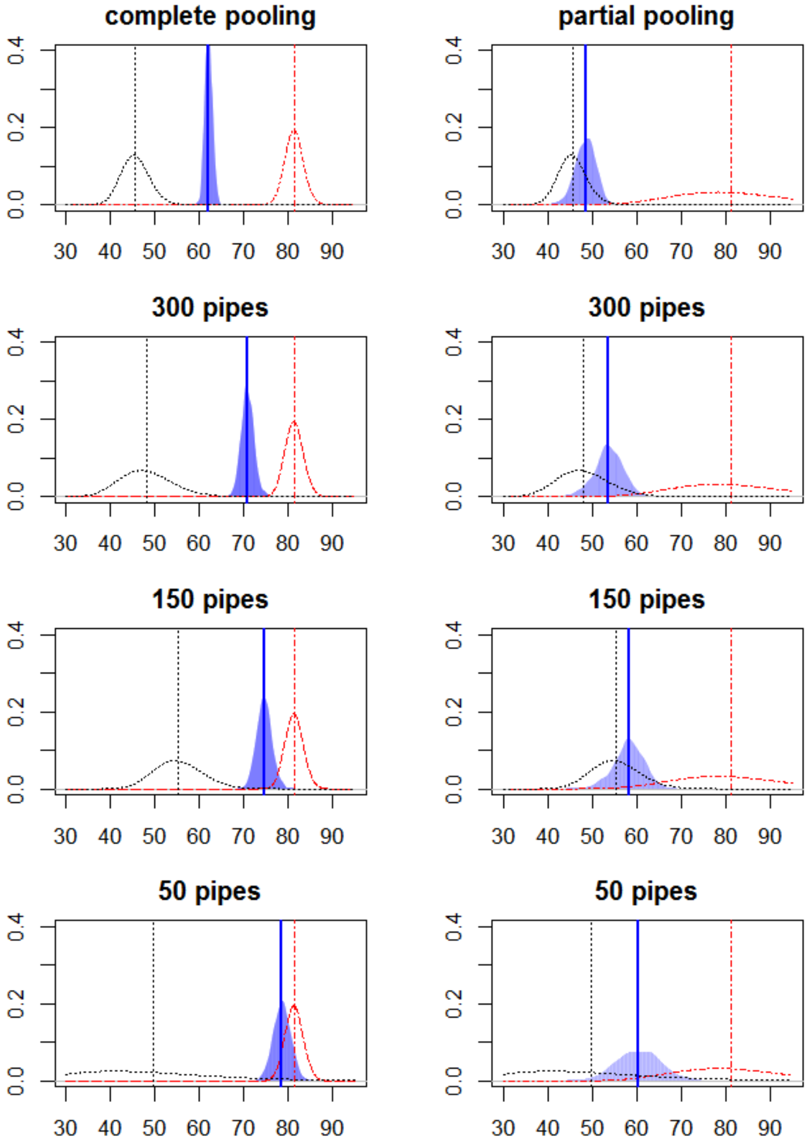

Figure 5: Bayesian inference with a complete (left) and partial (right) pooling prior for GI3. Prior (red dash-dotted), posterior (blue filled), and MLE (black dotted) marginal density distributions of the Weibull scale parameter μ_β are shown. Vertical lines indicate the position of the corresponding means. The top row shows the inference results using all data. Note that for no utility data (0 pipes, not shown), the posterior and prior marginal distributions are coincident.

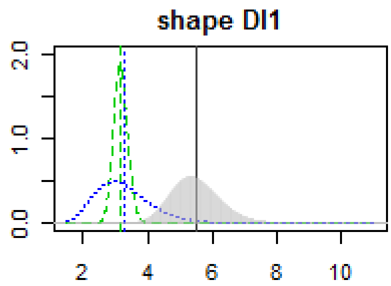

Figure 6: Shape parameter μ_α distributions of the two priors compared with the MLE (grey- shaded, solid) for DI1. Partial pool: blue dotted lines; complete pool: green dashed lines. Vertical lines indicate the position of the corresponding means.

4 Summary and Conclusions

4.1 Improved service life modeling under scarce data

We suggest a systematic approach to water pipe service life modeling that uses expert information from several utilities as prior which is then updated with local utility data. For this purpose available methods of expert elicitation of an unknown probability distribution were adapted and furthermore extended to imprecise quantile estimation of the pipe survival function. Contrary to currently existing approaches, encouraging experts to state interval estimates for the quantiles, leads to the imprecision or uncertainty of the expert being explicitly included in the analysis. From these statements, a bivariate expert prior for the survival function parameters is inferred, thus overcoming the difficulties confronted in elicitation of multivariate (i.e. correlated) distributions. The resulting expert priors can be aggregated with a linear pooling approach to get an intersubjective prior approximating the state of knowledge across experts and environmental conditions. For both Bayesian and frequentist inference of the parameters of the survival function from utility data, a likelihood function for the commonly encountered left-truncated right-censored pipe network data is derived.

The results from section 3.5 testify that the proposed approach improves estimation of the expected service life of water networks under scarce data, leading to the ability to identify parameters where otherwise not possible or to derive more informative estimates than with using MLE alone. This is a key improvement for more effective rehabilitation planning and water distribution network management.

4.2 Expert elicitation and prior aggregation

Priors for the parameters of the survival function characterizing technical service life of five pipe groups were obtained from interviews with eight water utility experts, ordered by perceived durability, most durable to least: asbestos cement, grey cast iron (1930-1965), polyethylene, steel, and ductile cast iron (1965-1980). This durability order is in agreement with literature estimates from Germany and Austria; the specific lifetime may vary depending upon local conditions.

An important aspect of our approach is the incorporation of the between-experts-variance into the aggregated prior distribution from individually fitted expert distributions. It is shown that not only the uncertainty of the single-expert, but also the deviation between experts is a valuable source of knowledge in itself, as it covers the different network conditions from various utilities. The resulting partial pooling prior of the experts' distributions leads to an intersubjective prior that covers these different water distribution network conditions as well as the different opinions of the experts. Only if the aim was a specific prior for a single utility and if more than one expert representing this utility was available, the complete pooling prior would be more appropriate because differences between experts are interpreted as measurement errors.

4.3 Frequentist and Bayesian inference

The results reviewed in section 3 suggest that Bayesian inference of survival function parameters by considering expert knowledge is a suitable approach to bridge the scarce data situation encountered in many water utilities. Frequentist estimation remains the less demanding approach if sufficient data is available (roughly more than 150 pipes with at least 5 years of data for this utility). To avoid prior-data conflicts, the validity of prior and posterior distributions for the locally encountered conditions can be discussed with a local expert.

4.4 Ambiguity of model selection

The problem of ambiguity arising in model selection is addressed by fitting three standard parametric survival models to the experts' data and choosing the best fitting for all experts. The Weibull model overall provided the best fit to experts statements, but did not prove the best choice for all pipe groups and experts.

Following the discussion in section 3.4, the use of pipe replacement data for prediction of a network's technical rehabilitation demand needs to be revised. Replacement data alone are not suitable to predict structural ageing only, but also reflect managerial decisions. In order to adopt more efficient rehabilitation approaches, a predictive model needs to be able to quantitatively describe these two factors separately. As the presented model might be prone to systematic biases induced by these limitations, it is not meaningful to discuss its predictive capability. The model's goodness of fit is nevertheless demonstrated by the close overlay of the non-parametric and parametric model shown in section 3.2.

4.5 Consideration of uncertainty

Tackling uncertainty on different levels and during different steps is useful to get an overall assessment of uncertainty the assessed quantities. This involves the use of examples and visual support tools to help the experts correctly express their belief in a quantitative manner. Cognitive biases can be reduced through training and control during elicitation, besides double-checking and repetition of the stated values. Influence of model selection can be elucidated by fitting several models and choosing the one that minimizes the deviation between measurements and modeled data. Lack of fit to the stated expert quantiles is made explicit by measurement of the error in the parameters.

5 Acknowledgements

This research was funded within the National Framework Programme 61 on Sustainable Water Management by the Swiss National Science Foundation (SNF). We thank the eight anonymous experts, the Swiss Gas and Water Association (SVGW), as well as Sébastien Apothéloz and Eauservice Lausanne (the large water utility providing replacement data), for their cooperation and contribution in this study. Special thanks owe to Adelaida Arreaza, Christoph Egger, and Simon Rinderknecht for their important contributions in preparing and performing the interviews with the experts, as well as colleagues for the valuable discussions.

6 References

Alpert, M., Raiffa, H., 1982. A progress report on the training of probability assessors, In: Kahneman, D., Slovic, S., Tversky, A. (Eds.), Judgment and Uncertainty: Heuristics and Biases. Cambridge University Press, pp. 294-305.

Alvisi, S., Franchini, M., 2010. Comparative analysis of two probabilistic pipe breakage models applied to a real water distribution system. Civil Engineering and Environmental Systems 27(1) 1-22.

Arreaza, A.S., 2011, Robust Elicitation of Imprecise Expert Knowledge for Water Infrastructure Management, Institute of Environmental Engineering, MSc thesis

Ayyub, B.M., 2001. Elicitation of Expert Opinions for Uncertainty and Risks, 1st ed. CRC Press LLC.

Berger, J.O., 1990. Robust Bayesian Analysis- Sensitivity to the prior. Journal of Statistical Planning and Inference 25(3) 303-328.

Bousquet, N., 2008. Diagnostics of prior-data agreement in applied Bayesian analysis. Journal of Applied Statistics 35(9) 1011-1029.

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.

Clemen, R.T., Winkler, R.L., 1999. Combining Probability Distributions From Experts in Risk Analysis. Risk Analysis 19(2) 187-203.

Cooke, R.M., 1991. Experts in uncertainty : opinion and subjective probability in science. Oxford University Press, New York.

Cooke, R.M., Goossens, L.L.H.J., 2008. TU Delft expert judgment data base. Reliability Engineering & System Safety 93(5) 657-674.

Coolen, F.P.A., 1996. On Bayesian reliability analysis with informative priors and censoring. Reliability Engineering and System Safety 53(1) 91-98.

Cordazzo, V., 2006. Die Sterblickkeit der Schweizer Geburtenjahrgänge 1900 bis 2030, Demos. Informationen aus der Demografie. Federal Statistical Office (BFS): Neuchâtel, p. 32.

Dridi, L., Mailhot, A., Parizeau, M., Villeneuve, J.-P., 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.

Eisenbeis, P., Rostum, J., Le Gat, Y., 1999. Statistical models for assessing the technical state of water networks : some european experiences, AWWA American Water Works Association Annual Conference. AWWA: Chicago, Illinois.

Eisenführ, F., Weber, M., Langer, T., 2010. Rational Decision Making, 1st ed. Springer.

Frühwirth-Schnatter, S., 2006. Finite mixture and markov switching models. Springer, New York.

Fuchs-Hanusch, D., Gangl, G., Kornberger, B., Kölbl, J., Hofrichter, J., Kainz, H., 2008. PiReM- Pipe Rehabilitation Management Developing A Decision Support System for Rehabilitation Planning of Water Mains. Water Practice & Technology 3(1) 1-9.

Fuchs, D., 2001, Decision Support Systeme für die Rehabilitationsplanung von Wasserrohrnetzen, Faculty of Civil Engineering, Institute of Urban Water Management and Landscape Water Engineering, PhD thesis

Gelman, A., Carlin, J.B., Stern, H.S., Rubin, D.B., 2004. Bayesian data analysis, 2nd ed. Chapmann & Hall/CRC.

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

Genest, C., Zidek, J.V., 1986. Combining Probability Distributions: A Critique and an Annotated Bibliography. Statistical Science 1(1) 114-135.

Gillies, D., 1991. Intersubjective probability and confirmation theory. British Journal for the

Philosophy of Science 42(4) 513-533.
Harrell, J., Frank E., 2001. Regression modeling strategies: with applications to linear models, logistic regression, and survival analysis. Springer- Verlag.
Herz, R.K., 1995. Alterung und Erneuerung von Infrastrukturbeständen- Ein Kohortenüberlebensmodell, Jahrbuch für Regionalwissenschaft.
Herz, R.K., 1998. Exploring rehabilitation needs and strategies for water distribution networks. Journal of Water Supply Research and Technology AQUA 47(6) 275.
Jouini, M.N., Clemen, R.T., 1996. Copula models for aggregating expert opinions. Operations Research 44(3) 444-457.
Klein, J.P., Moeschberger, M.L., 2003. Survival analysis: techniques for censored and truncated data, 2nd ed. New York : Springer, New York.
Kleiner, Y., Rajani, B., 1999. Using Limited Data to Assess Future Needs. Journal AWWA 91(7) 47-61.
Kleiner, Y., Rajani, B., 2001. Comprehensive Review of structural deterioration of water mains: statistical models. Urban Water(3) 131-150.
Korving, H., van Noortwijk, J.M., 2008. Bayesian updating of a prediction model for sewer degradation. Urban Water Journal 5(1) 51-57.
Krueger, T., Page, T., Hubacek, K., Smith, L., Hiscock, K., 2012. The role of expert opinion in environmental modelling. Environmental Modelling & Software 36(0) 4-18.
Kuhnert, P.M., Martin, T.G., Griffiths, S.P., 2010. A guide to eliciting and using expert knowledge in Bayesian ecological models. Ecology Letters 13(7) 900-914.
Kynn, M., 2008. The 'heuristics and biases' bias in expert elicitation. Journal of the Royal Statistical Society: Series A (Statistics in Society) 171(1) 239-264.
Larrick, R.P., Soll, J.B., 2006. Intuitions about combining opinions: Misappreciation of the averaging principle. Management Science 52(1) 111-127.
Lele, S.R., Allen, K.L., 2006. On using expert opinion in ecological analyses: a frequentist approach. Environmetrics 17(7) 683-704.
Low Choy, S., O'Leary, R., Mengersen, K., 2009. Elicitation by design in ecology: using expert opinion to inform priors for Bayesian statistical models. Ecology 90(1) 265-277.
Mailhot, A., Pelletier, G., Noel, J.F., 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.
Nash, J.C., Varadhan, R., 2011. Unifying Optimization Algorithms to Aid Software System Users: optimx for R. Journal of Statistical Software 43(9) 1-14.
O'Hagan, A., 2012. Probabilistic uncertainty specification: Overview, elaboration techniques and their application to a mechanistic model of carbon flux. Environmental Modelling & Software 36(0) 35-48.
O'Hagan, A., Buck, C.E., Daneshkhah, A., Eiser, J.R., Garthwaite, P.H., Jenkinson, D.J., Oakley, J.E., Rakow, T., 2006. Uncertain Judgements: Eliciting Experts' Probabilities. John Wiley & Sons.
O'Leary, R.A., Choy, S.L., Murray, J.V., Kynn, M., Denham, R., Martin, T.G., Mengersen, K., 2009. Comparison of three expert elicitation methods for logistic regression on predicting the presence of the threatened brush-tailed rock-wallaby Petrogale penicillata. Environmetrics 20(4) 379-398.
Oakley, J.E., O'Hagan, A., 2007. Uncertainty in prior elicitations: a nonparametric approach. Biometrika 94(2) 427-441.
Oakley, J.E., O'Hagan, A., 2010. SHELF: the Sheffield Elicitation Framework (version 2.0): School of Mathematics and Statistics, University of Sheffield.
Pan, W., Chappell, R., 1998. A Nonparametric Estimator of Survival Functions for Arbitrarily Truncated and Censored Data. Lifetime Data Analysis 4(2) 187-202.

Pelletier, G., Mailhot, A., Villeneuve, J.P., 2003. Modeling water pipe breaks - three case studies. Journal of Water Resources Planning and Management 129(2) 115-123.
Rajani, B., Kleiner, Y., 2001. Comprehensive review of structural deterioration of water mains: physically based models. Urban Water 3 151-164.
Renaud, E., De Massiac, J.C., Brémond, B., Laplaud, C., 2009. SIROCO, a decision support system for rehabilitation adapted for small and medium size water distribution companies, In: Alegre, H., do Céu Almeida, M. (Eds.), Strategic Asset Management of Water Supply and Wastewater Infrastructures. IWA Publishing: London, pp. 329-344.
Rinderknecht, S.L., Borsuk, M.E., Reichert, P., 2011. Eliciting density ratio classes. International Journal of Approximate Reasoning 52(6) 792-804.
Rinderknecht, S.L., Borsuk, M.E., Reichert, P., 2012. Bridging uncertain and ambiguous knowledge with imprecise probabilities. Environmental Modelling & Software 36(0) 122-130.
Roscher, H., Ahrens, J., Rödiger, S., Sorge, H.-C., 2008. Sanierung und Erneuerung von Wasserrohrnetzen, In: Weimar, B.-U. (Ed.), Rehabilitation von Rohrleitungen-Sanierung und Erneuerung von Ver- und Entsorgungsnetzen, pp. 63-204.
Roscher, H., Mischner, J., Sorge, H.-C., 2005, Vergleich von Bewertungssystemen für Rohrleitungen: Wasser- und Gasrohrnetze, Erfurt,
Runge, M.C., Converse, S.J., Lyons, J.E., 2011. Which uncertainty? Using expert elicitation and expected value of information to design an adaptive program. Biological Conservation.
Scheidegger, A., 2011. adaptMCMC: Implementation of a generic adaptive Monte Carlo Markov Chain sampler, 1.0 ed. CRAN R project.
Speirs-Bridge, A., Fidler, F., McBride, M., Flander, L., Cumming, G., Burgman, M., 2010. Reducing Overconfidence in the Interval Judgements of Experts. Risk Analysis 30(3) 512-523.
Spetzler, C.S., Stael Von Holstein, C.-A., S. , 1975. Probability Encoding in Decision Analysis. Management Science 22(3) 340-358.
SVGW, 2009, Statistische Erhebungen der Wasserversorgungen in der Schweiz Betriebsjahr 2008, Zürich, Schweizer Verein des Gas- und Wasserfaches,
Trujillo Alvarez, R., 1995. Bedarfsprognose und Strategieentwicklung für die Rehabilitation städtischer Wasserrohrnetze. Schriftenreihe des Instituts für Städtebau und Landesplanung, Universität Fridericiana zu Karlsruhe.
Tversky, A., 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., 2011. Robust adaptive Metropolis algorithm with coerced acceptance rate. Statistics and Computing 1-12.
Wayne, N.B., 2004. Applied Life Data Analysis. John Wiley & Sons Inc.
Winkler, R.L., 1967. The assessment of prior distributions in Bayesian analysis. Journal of the American Statistical Association 62 776-880.

7 Appendices

Appendix A- Expert elicitation

Table 7: Description of pipe groups based on common differentiation criteria

pipe group	abbre-viation	differentiation criteria	chosen for elicitation and underlying rationale
asbestos cement incl. Eternit	**AC**	years 1930-1985	Yes, although no longer produced. Has proved to be very resistant if appropriately installed and is still common in Swiss water networks, especially in smaller communities, under the name of "Eternit".
1st generation grey cast iron	**GI1**	from horizontal sand molds, corrosion-resistant, varying wall thickness; before 1880	No, because close to no occurrence in today's Swiss networks.
2nd generation grey cast iron	**GI2**	vertically cast, more corrosion resistance, ca. 1880-1930	No, mostly in bigger cities only; building of networks in smaller communities rather later.
3rd generation grey cast iron	**GI3**	centrifugal casting, susceptible to corrosion, 1930-1965	Yes, slowly being replaced, but still common.
1st generation ductile cast iron	**DI1**	centrifugal casting, lacking external corrosion protection, high tensile strength, 1964-1980	Yes, slowly being replaced, but still common.
2nd generation ductile cast iron	**DI2**	similar to DI1, but improved external corrosion protection, after 1980	No, because probably only little replacement up to today and time constraints. Interesting in hindsight because of the rather large proportion in today's networks.
1st generation steel	**ST1**	welded or seamless, lacking external corrosion protection, before 1930	Yes, but without differentiation into generations because occurrence of this material varied in the networks the experts were familiar with and they were not confident in making further differentiations.
2nd generation steel	**ST2**	insufficient external corrosion protection, ca. 1930 - 1980	
3rd generation steel	**ST3**	enhanced external corrosion protection, after 1980	
PVC polyvinylchloride	**PVC**	approx. 1930-1990	No, only rarely used in Switzerland's water supply systems.
1st generation polyethylene	**PE1**	PE-LD; PE-HD; and PE 63; before 1980	Yes, but also without differentiation because it is a rather new application in Switzerland and because of the high expected service life. Not yet many replacements if installed correctly (leading to a lack of experience of experts with this material).
2nd generation polyethylene	**PE2**	PE 80 and PE 100; after 1980	
3rd generation polyethylene	**PE3**	PE-X, cross-linked PE, high toe-crack resistance; after 1992	

Table 8: Stated quantile values from expert elicitation, l = lower value, u = upper value. If only one value is given, the expert stated precise estimates. The first column indicates material and quantile.

	E1		**E2**		**E3**		**E4**		**E5**		**E6**		**E7**		**E8**	
group	l	u	l	u	l	u	l	u	l	u	l	u	l	u	l	u
GI3																
0.05	30		20	40	30		30	40	20	30	40		50		60	70
0.25	40		40	60	65		40	50	30	50	55		70		60	90
0.5	60		70	90	80		60	70	60	80	80		90		90	110
0.75	80		80	100	85		70	85	70	90	90		100		100	130
0.95	100		90	110	110		85	90	80	110	100		120		130	140

	E1		E2		E3		E4		E5		E6		E7		E8	
group	l	u	l	u	l	u	l	u	l	u	l	u	l	u	l	u
DI1																
0.05	20		10	15	25	30	10	20	20	30	20	30	20		30	40
0.25	30		20	30	40	55	20	30	30	40	40	50	45		40	60
0.5	45		40	60	60	70	45	55	40	60	50	60	55		60	70
0.75	65		60	70	70	80	55	60	60	70	60	70	70		80	90
0.95	100		70	80	80	85	60	70	80	90	70	80	80		90	100
AC																
0.05	30	50	40		40	60	-		20		-		20	30	-	
0.25	60	80	80		75	85	-		35		-		40	60	-	
0.5	80	100	100		90	110	-		50		-		55	65	-	
0.75	100	120	110		115	135	-		65		-		65	75	-	
0.95	120	150	120		140	150	-		80		-		75	90	-	
ST																
0.05	-		-		-		20		-		-		30	40	30	40
0.25	-		-		-		41		-		-		45	55	40	60
0.5	-		-		-		42		-		-		60	70	60	70
0.75	-		-		-		43		-		-		70	90	80	90
0.95	-		-		-		45		-		-		90	110	90	100
PE																
0.05	0	50	40		-		30		15		50	60	-		65	
0.25	50	150	60		-		40		25		60	80	-		80	
0.5	100	150	80		-		50		35		80	100	-		100	
0.75	100	150	95		-		60		55		100	110	-		120	
0.95	100	200	110		-		70		70		100	150	-		140	

Table 9: Description of experts and locally specific influence factors of network deterioration

expert	Position/qualification	mentioned influence factors
E1	Planning & construction engineer; head of local engineering company servicing several small water suppliers in the Zürcher Oberland	- DI1 is "problem child" - GI3 has mostly been removed (only about 5-10 % of today's network) - bedding - AC: connections deteriorate faster than pipes
E2	Planning & construction engineer of the same local engineering company as expert 1; but servicing different communities	- GI3 ca. 30 % of current network - settling and corrosion problems with GI3 - bedding - electrical grounding of house installations on DC1 pipes - PE more or less 20-30 % of network, some problems with earlier PE pipes - AC: connections deteriorate faster than pipes, mostly used for large diameter transport pipes
E3	Operation & maintenance engineer; head of distribution network department of the public water supply for a medium size city in NE Switzerland	- close to no PE and ST in current network - electrical grounding of house installations on DC1 pipes until late 90s - GI3 approx. 18 % of network - DI1 approx. 20 % of network, usually 45 years of service life assumed - AC: laying depth is rather deep, mostly used for larger diameter transport pipes - favorable soil conditions

expert	Position/qualification	mentioned influence factors
E4	Project and construction manager; head of distribution network department of a private water supply company serving a medium size city in central Switzerland	- laying depth of AC is rather deep - mechanical impacts by traffic - DC1 problems with bedding (wood used as support under the pipe) - many different pressure zones - because of large financial losses caused by GI3 failures, this material is replaced earlier based on risk considerations - ST was massively replaced in the 90s because of failures probably caused by inappropriate bedding
E5	Operation & maintenance engineer of a consortium of small municipal water suppliers in central Switzerland	- strongly varying soil conditions from rugged rocks over river gravel to heavy clay and aggressive moor soils - problems with quality of PE installation, especially welding has been an issue; rather early use of PE; PE only used if conditions do not allow for metal pipes (mostly soil) - rather soft water (12-13 °fH eq. to 1,2-1,3 mmol/L) - tank traffic is problematic for GI3 - overall many different pressure zones
E6	Network utility engineer from a public water supplier of a small city in NW Switzerland	- usually assume fixed service life of 60 years for pipes - strongly favoring PE for replacements; rather high percentage of the network are PE pipes (> 20 %) - problems with both bedding (timber, wooden support) and electrical grounding of house installations on DC1 pipes
E7	Facility manager and engineer of a consortium of small water suppliers in NW Switzerland	- strongly favoring cast iron, PE only for household connections - difficult ground because of soil variations from strongly settling to peaty soils - problems with bedding in 60's and 70's when most of their pipes were installed - have experienced many failures in both GI3 and DI1 pipes
E8	Head of distribution network department of the public water supply of a larger city in NE Switzerland	- assume 100-120 years of service life for GI3 - systematic defects in pipes built shortly after Second World War - earlier PE types expected to have much shorter duration than newer PE types; longer service life assumptions supported by a recent material study of DVGW - heterogenic soil and ground properties - electrical grounding of house installations on DC1 pipes

Detailed description of the elicitation procedure

The interviews were attended by at least three persons: the expert, the interviewer and the analyst. If more people were attending the interviews (e.g. assistants), they were not allowed to actively participate or alter the elicitation procedure. The interviewer and analyst completely abstained from any kind of coaching regarding the order of magnitude of the answers.

Setting the scene: At the beginning, the interviewer repeated the purpose of the study, and explained the way she would proceed during the interview. The aim of the elicitation was clearly stated. Then, the quantity to be elicited was clarified, and the five pipe groups were presented. It was reaffirmed that the named pipe groups actually occurred in the expert's domain. The four most familiar (in one case three) were then selected for elicitation. After this was done, the expert was asked several questions regarding his expertise and familiarity with probability.
If the expert had not read or understood the pre-elicitation information, there was time to go through it in detail.

Focusing: Then, the expert was requested to name the most important influencing factors for pipe ageing in his area (see Table 9). This was not only done to learn about special circumstances of the different localities, but also to make him concentrate on the upcoming task. Then, he was motivated by stating that his knowledge was an important source of information for the estimation of pipe service life. It was made explicit that he was not expected to be all-knowing, but that his expertise was crucial for the study. This was to encourage him to answer as best according to his knowledge while adequately stating his uncertainty.

Training: Before elicitation of the quantities of interest started, an elicitation round identical to the assessment to come of pipe service life was done to train the expert. To avoid anchoring, the survival of women born in Switzerland in the year 1940 as provided by the Swiss Federal Statistics Office (Cordazzo, 2006) was used. It was used to familiarize the expert with the elicitation procedure and to avoid possible misunderstandings. It also helped to cross-check for calibration (which was clearly not possible for pipes because of a lack of factual measurement data for the experts utility). This procedure allowed the interviewers to point out features to keep in mind which can lead to biases that might well arise during elicitation of pipe survival.

Elicitation of pipe groups: After this training, the pipe group being addressed was again specified. The expert was asked about his experience with this group in his area of responsibility. Then, the quantiles were elicited in a first round following the sequence:

1. Define the overall interval defined by the age at which most pipes are expected to have been replaced (95 %) and at which close to all pipes of a group/ cohort are still in service (5 %).
2. The age at which half (50 %) of the pipes are replaced and half remain in service is the median. This is the third quantity elicited.
3. Finally, the values for three-quarters (75 %) and a quarter (25 %) are assessed.

At the same time, the addressed quantities were roughly visualized using 100 paper clips

(representing 100 % of the pipe group) and a paper sheet with the time bar. Experts were requested to disregard replacement because of initial laying failures (e.g. within the first year after laying), as well as replacements following managerial or economic considerations, such as coordinative ground works with other infrastructure providers. This helped to focus on technical ageing and not on effects of different management decisions.
Secondly, the quantiles for 75 %, 50 %, and 25 % were re-elicited using bets, adjusting the stated ages until the expert was indifferent between the bets. This technique was used by the interviewer to confirm the statements by making the experts think differently about the quantities. After the bets, she read out the documented values and individually confirmed them with the expert. These checks and repetitions were done to ensure that experts' statements are reliable, consistent and correctly documented. At the end, the experts were asked to provide a qualitative description of the imaginary density curve, if possible. The description should reveal whether it could be assumed unimodal.
This elicitation procedure was repeated for the selected pipe groups. During the interview, the experts were asked to assess half of the pipe groups stating imprecise estimates and the other half using precise estimates.

Feedback: At the end of the interview, the experts had to assess the difficulty of the interview, how realistic they think their stated answers are, and if they preferred stating point estimates or intervals.

Table 10: Estimates of pipe survival in the literature as reported from Austria and Germany

	Roscher et. al (2005)			**Fuchs (2001)***			**Trujillo Alvarez (1995)***		
	0%	**50 %**	**90 %**	**0%**	**50%**	**90%**	**0%**	**50%**	**90%**
AC	-	-	-	20-50	60-90	80-110	5-80	50-90	60-110
GI1	60-70	65-90	80-110	-	-	-	5-70	40-101	80-150
GI2				30-80	100-160	130-190			
GI3	40-60	65-90	80-100	10-30	50-90	70-110			
DI1	20-30	45-65	70-100	5-20	40-70	70-90	10-60	30-90	50-120
DI2	40-50	75-90	100-130	80-120	100-140	120-160	6-100	50-140	90-165
ST1				-	-	-			
ST2	40-50	60-80	80-110	-	-	-	4-60	25-100	55-120
ST3				100	120	140			
PVC	10-30	30-50	50-70	10-30	40-85				
PE1	15-30	50-75	50-70	-	-	-	5-60	40-80	50-100
PE2	40-50	75-90	100-130	-	-	-			
PE3	40-50	75-90	100-130	20-30	50-70	80-90	-	-	-

* Originally, intervals for an optimistic and pessimistic estimate were given which were herein merged together. Thus, the upper bound of the stated interval is the upper bound of the optimistic estimate; the lower bound equals the lower bound of the pessimistic estimate.

Appendix B- Parametric model fit and parameter uncertainty

Table 11: Goodness of fit of Weibull, lognormal, and gamma distribution from non-linear least squares regression over the elicited quantiles. Bold numbers indicate the model which minimizes the residual sum of squares (RSS) and residual standard error (RSE), dependent on available degrees of freedom (doF).

group	**expert**	**doF**	**Weibull**		**lognormal**		**gamma**	
			RSS	RSE	RSS	RSE	RSS	RSE
GI3	E1	3	**59.6**	4.458	95.8	5.65	61.9	4.542
GI3	E2	8	**1380**	13.14	1930	15.53	1740	14.74
GI3	E3	3	**140**	6.837	342	10.68	272	9.515
GI3	E4	8	**384**	6.932	526	8.112	467	7.641
GI3	E5	8	**1370**	13.1	1780	14.91	1610	14.18
GI3	E6	3	**89.5**	5.463	240	8.937	190	7.963
GI3	E7	3	**14.1**	2.169	100	5.781	65.2	4.663
GI3	E8	8	1420	13.32	1390	13.18	**1360**	13.03
DI1	E1	3	62.1	4.548	**11.6**	1.964	15.4	2.267
DI1	E2	8	**650**	9.015	1030	11.37	850	10.31
DI1	E3	8	**375**	6.851	698	9.338	595	8.626
DI1	E4	8	**497**	7.885	780	9.873	664	9.108
DI1	E5	8	452	7.515	462	7.6	**430**	7.328
DI1	E6	8	**292**	6.046	482	7.765	419	7.236
DI1	E7	3	**61.5**	4.526	217	8.506	159	7.285
DI1	E8	8	**492**	7.846	668	9.138	590	8.591
AC	E1	8	**1270**	12.58	1680	14.48	1490	13.67
AC	E2	3	**277**	9.61	694	15.21	582	13.93
AC	E3	8	**754**	9.705	1220	12.35	1030	11.35
AC	E5	3	**10.8**	1.899	96.3	5.665	53.1	4.205
AC	E7	8	**551**	8.302	809	10.06	723	9.506
ST	E7	8	**577**	8.491	620	8.805	577	8.489
ST	E8	8	**492**	7.846	668	9.138	590	8.591
PE	E2	3	**22**	2.709	162	7.343	108	5.998
PE	E4	3	**6.54**	1.476	23.2	2.782	13.5	2.125
PE	E5	3	**32.5**	3.293	68.8	4.789	3.711	41.3
PE	E6	8	**1820**	15.09	1860	15.26	**1820**	15.07
PE	E8	3	67.3	4.736	63.3	4.593	**42.9**	3.782

Estimation of the variance-covariance matrix from the Fisher information matrix

The inverse expected Fisher information matrix is essentially equivalent to the negative of the Hessian matrix of the obtained estimate, i.e. the second order partial derivatives. In the case of a two-parametric model the Hessian is a 2,2-square matrix calculated from:

$$H(\ln L) = \frac{\delta\theta_2{}^2 \delta\theta_1}{\delta\theta_1{}^2} \ln l(\boldsymbol{\theta}) = \begin{bmatrix} \frac{\delta^2 lnL}{\delta\theta_1^2} & \frac{\delta^2 lnL}{\delta\theta_1 \delta\theta_2} \\ \frac{\delta^2 lnL}{\delta\theta_2 \delta\theta_1} & \frac{\delta^2 lnL}{\delta\theta_2^2} \end{bmatrix} = \begin{bmatrix} \delta_{\theta_1\theta_1} & \delta_{\theta_1\theta_2} \\ \delta_{\theta_2\theta_1} & \delta_{\theta_2\theta_2} \end{bmatrix} \quad (12)$$

where θ_1,θ_2 are the parameters of the parametric model. The Hessian (matrix) describes the local curvature of the logL function. This, and the reasoning behind it are well described in Harrell (2001, pp. 180-183).

Description of the extended Nelson-Aalen estimator

A Nelson-Aalen estimator adapted for left-truncated and right-censored data (also referred to as *extended Nelson estimator)* as described in Pan and Chappell (1998) and applied to pipe survival in Carrión, Solano et al. (2010)) is used.

$\hat{H}_a(t) = \sum_{a \le t_i \le t} \left[\frac{d_i}{Y_i}\right], t \ge a,$ (Klein and Moeschberger, 2003)

t … age at replacement or censoring (=end of study)
a_i… age at entering the study
d_i…number of events(=deaths/replacements) at time t_i
Y_i … number of individuals entering the study before t_i and for which $a < t_i \le t$
the survival can then be estimated with:

$$\hat{S}_a(t) = e^{-\hat{H}_a(t)}$$

With this layout, over-estimation through right-censoring and underestimation caused by left-truncation is reduced/avoided. According to (Klein and Moeschberger, 2003), this estimator can be interpreted as an estimator of the probability of survival beyond time t conditional on the smallest of the entry times: Pr[T > t | T > a]. T: time to event.

Calculation of point-wise confidence intervals of nonparametric fit
(Klein and Moeschberger, 2003) from S.105ff:
log-transformed confidence intervals are chosen because of better reported performance in small samples

$$\left[\hat{S}_{t_0}^{1/\theta}, \hat{S}_{t_0}^{\theta}\right];\ \theta = e^{\left\{\frac{Z_{1-\alpha/2} \cdot \sigma_S(t_0)}{\ln[\hat{S}(t_0)]}\right\}}$$

with $\boldsymbol{\sigma_S^2(t)} = \frac{\hat{V}[\hat{S}(t)]}{\hat{S}^2(t)}$

The resulting confidence interval is not necessarily symmetric about the estimate of the survival function.

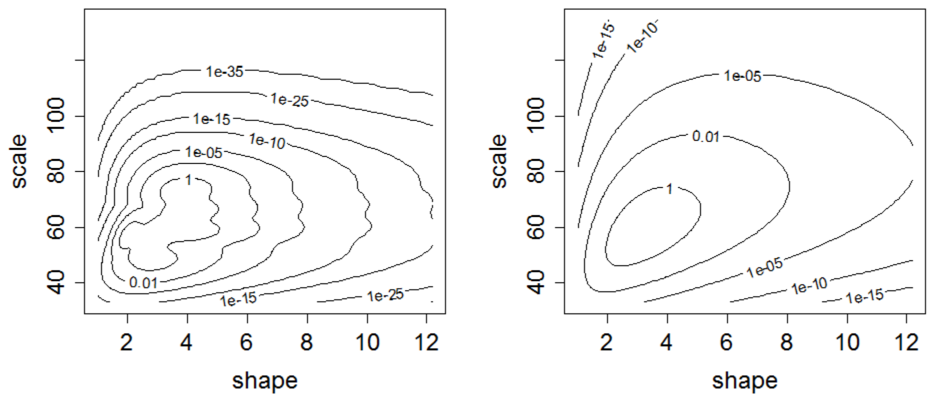

Figure 1: Bivariate probability density distribution of the aggregated prior before smoothing (left, multimodal) and after smoothing (right, unimodal) for ductile cast iron (1965-1980).

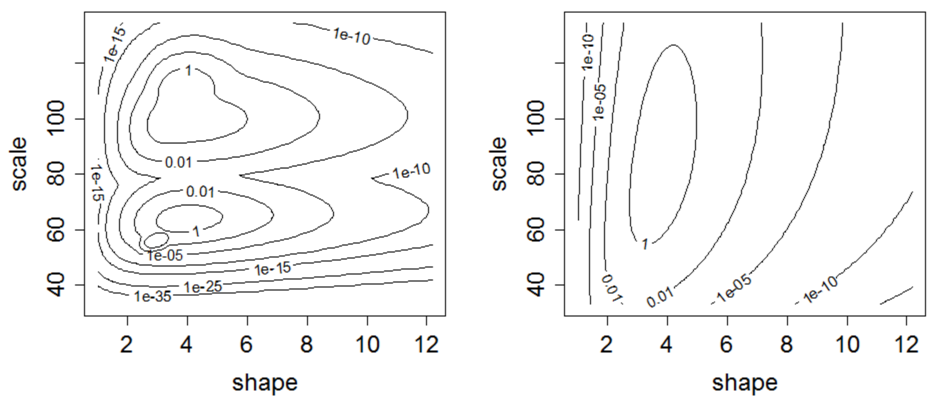

Figure 2: Bivariate probability density distribution of the aggregated prior before smoothing (left, multimodal) and after smoothing (right, unimodal) for asbestos cement.

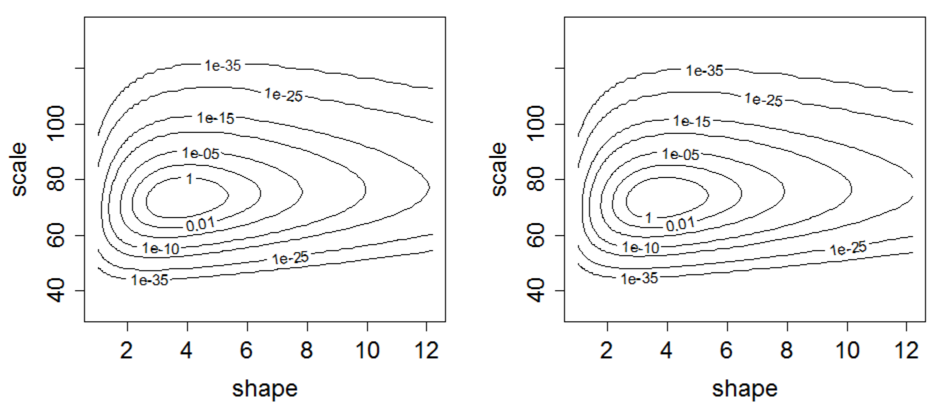

Figure 3: Bivariate probability density distribution of the aggregated prior before smoothing (left, multimodal) and after smoothing (right, unimodal) for steel.

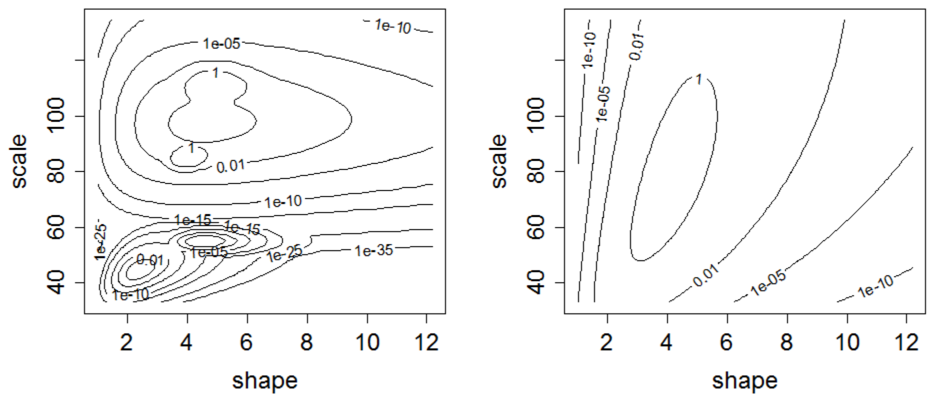

Figure 4: Bivariate probability density distribution of the aggregated prior before smoothing (left, multimodal) and after smoothing (right, unimodal) for polyethylene.

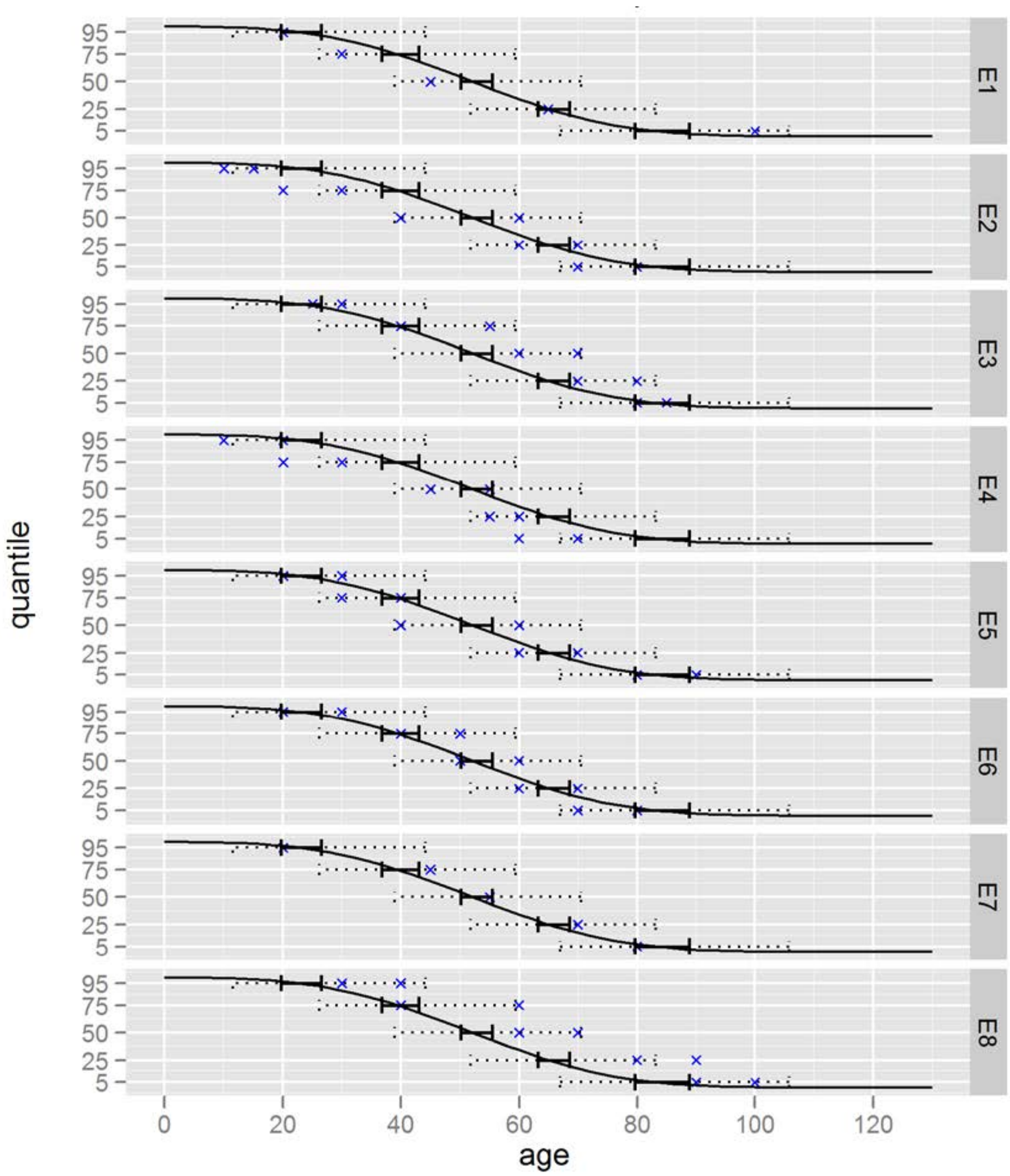

Figure 5: Comparison of DI1 priors and estimates from experts. Blue crosses represent quantile values as stated by the expert indicated on the right edge (E1…E8). Solid error bars give the 95 % confidence intervals for complete pooling, dotted error bars for partial pooling. The survival curve is calculated from the mean parameters $(\hat{\alpha}, \hat{\beta})$ of the partial pooling prior, see Table 4.

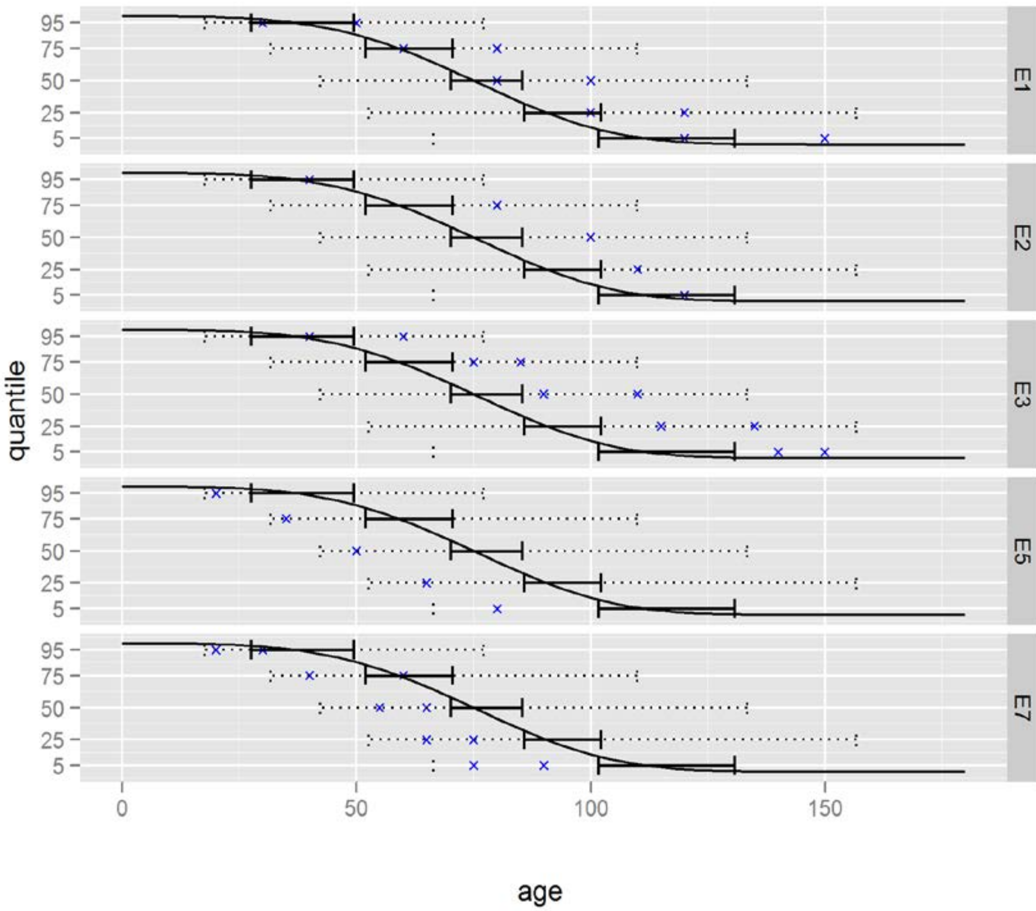

Figure 6: Comparison of AC priors and estimates from experts. Blue crosses represent quantile values as stated by the expert indicated on the right edge (E1…E8). Solid error bars give the 95 % confidence intervals for complete pooling, dotted error bars for partial pooling. The survival curve is calculated from the mean parameters ($\hat{\alpha},\hat{\beta}$) of the partial pooling prior, see Table 4.

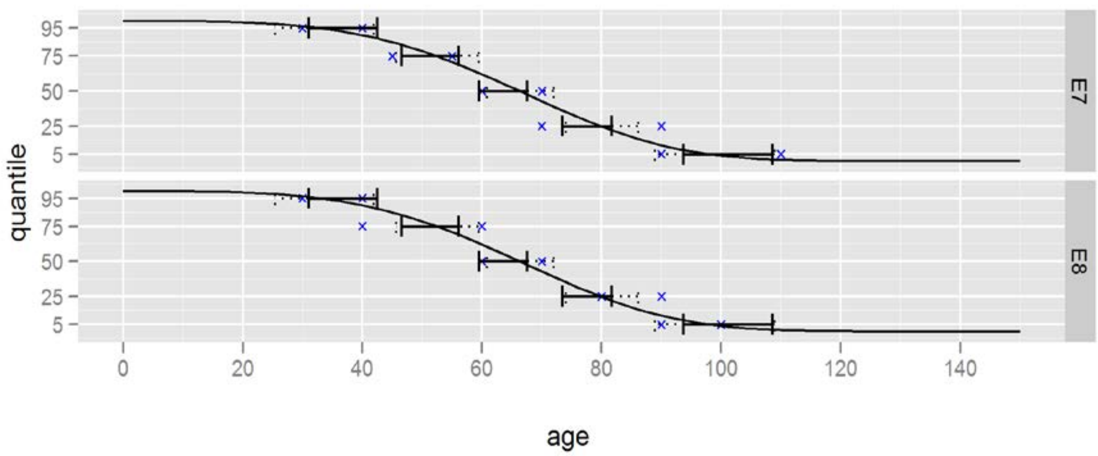

Figure 7: Comparison of ST priors and estimates from experts. Blue crosses represent quantile values as stated by the expert indicated on the right edge (E1…E8). Solid error bars give the 95 % confidence intervals for complete pooling, dotted error bars for partial pooling. The survival curve is calculated from the mean parameters ($\hat{\alpha},\hat{\beta}$) of the partial pooling prior, see Table 4.

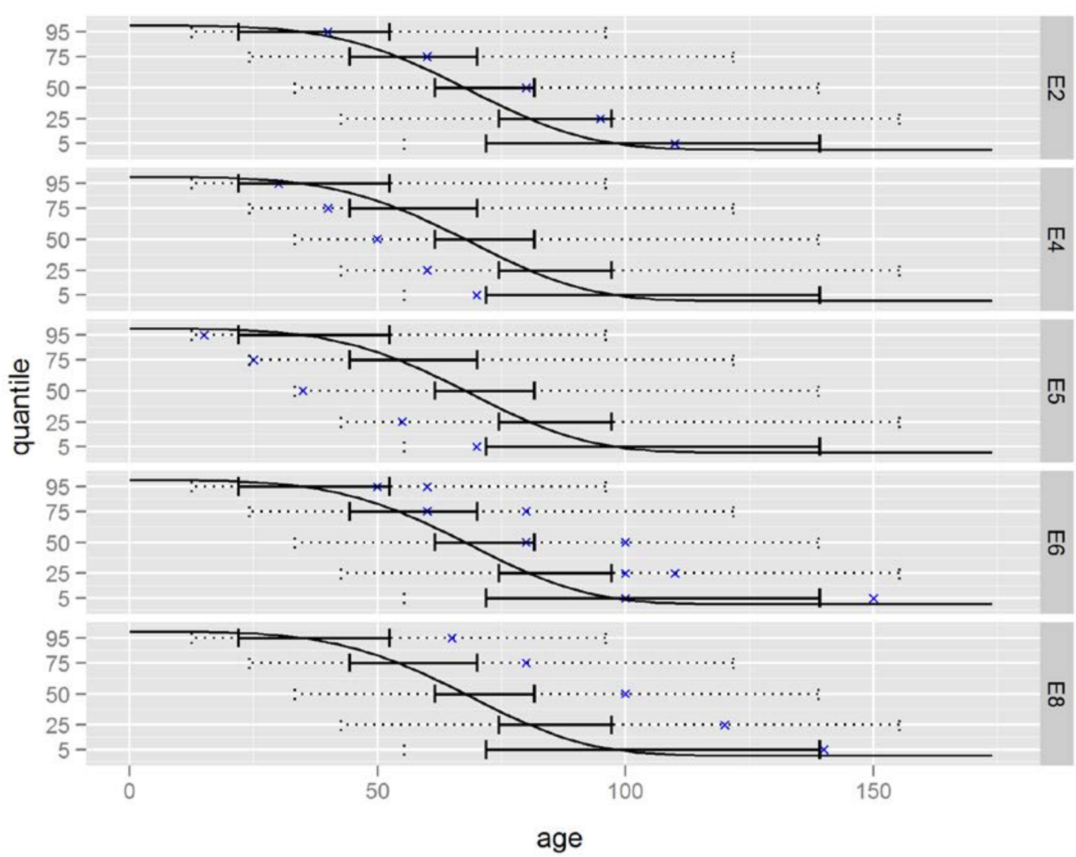

Figure 8: Comparison of ST priors and estimates from experts. Blue crosses represent quantile values as stated by the expert indicated on the right edge (E1…E8). Solid error bars give the 95 % confidence intervals for complete pooling, dotted error bars for partial pooling. The survival curve is calculated from the mean parameters ($\hat{\alpha},\hat{\beta}$) of the partial pooling prior, see Table 4.

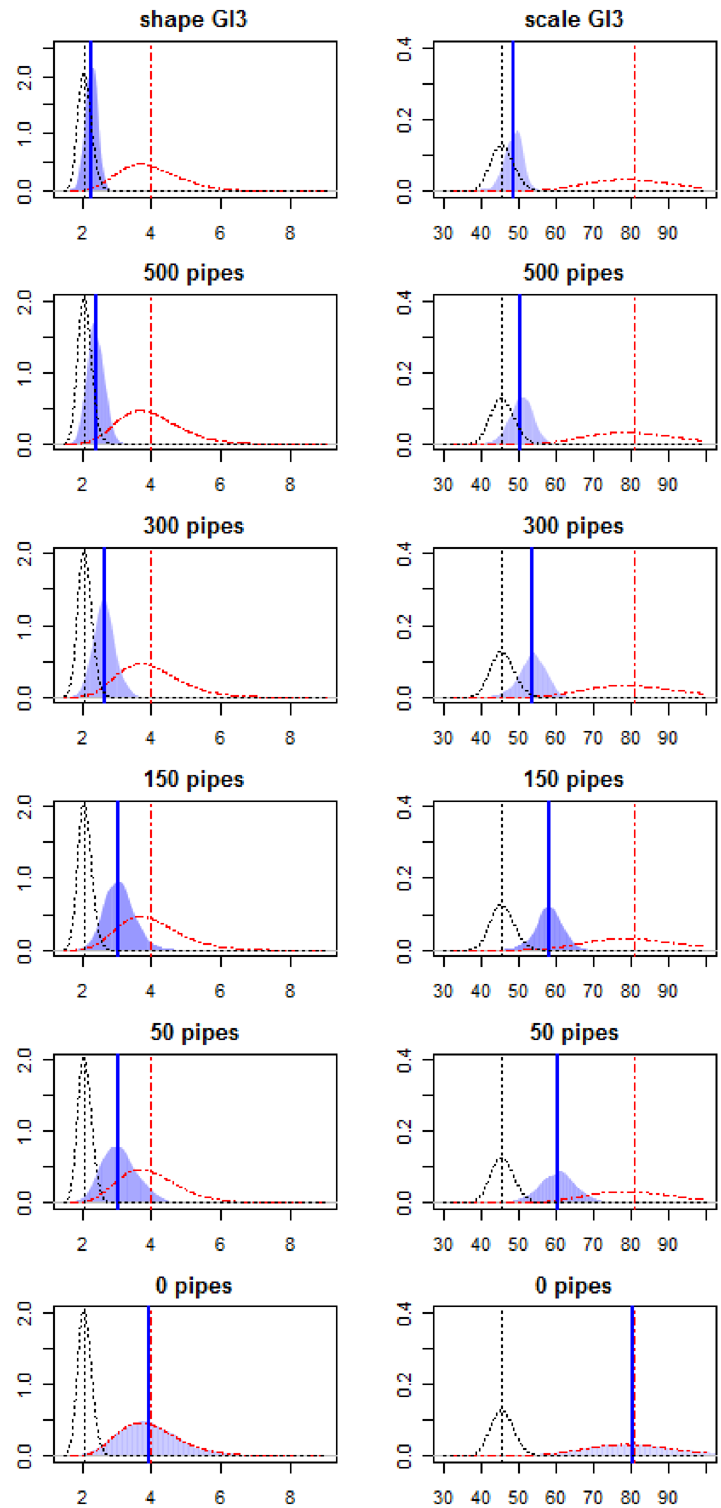

Figure 9: Bayesian inference with partial-pooling prior for grey cast iron (GI3): Posterior (blue filled), and prior (red dash-dotted) marginal distributions of the Weibull shape (left column) and scale (right column) parameters for varying amounts of data (top level = all data). As a reference, the distributions resulting from MLE with all data (black dotted) are also plotted. Vertical lines indicate the position of the corresponding means.

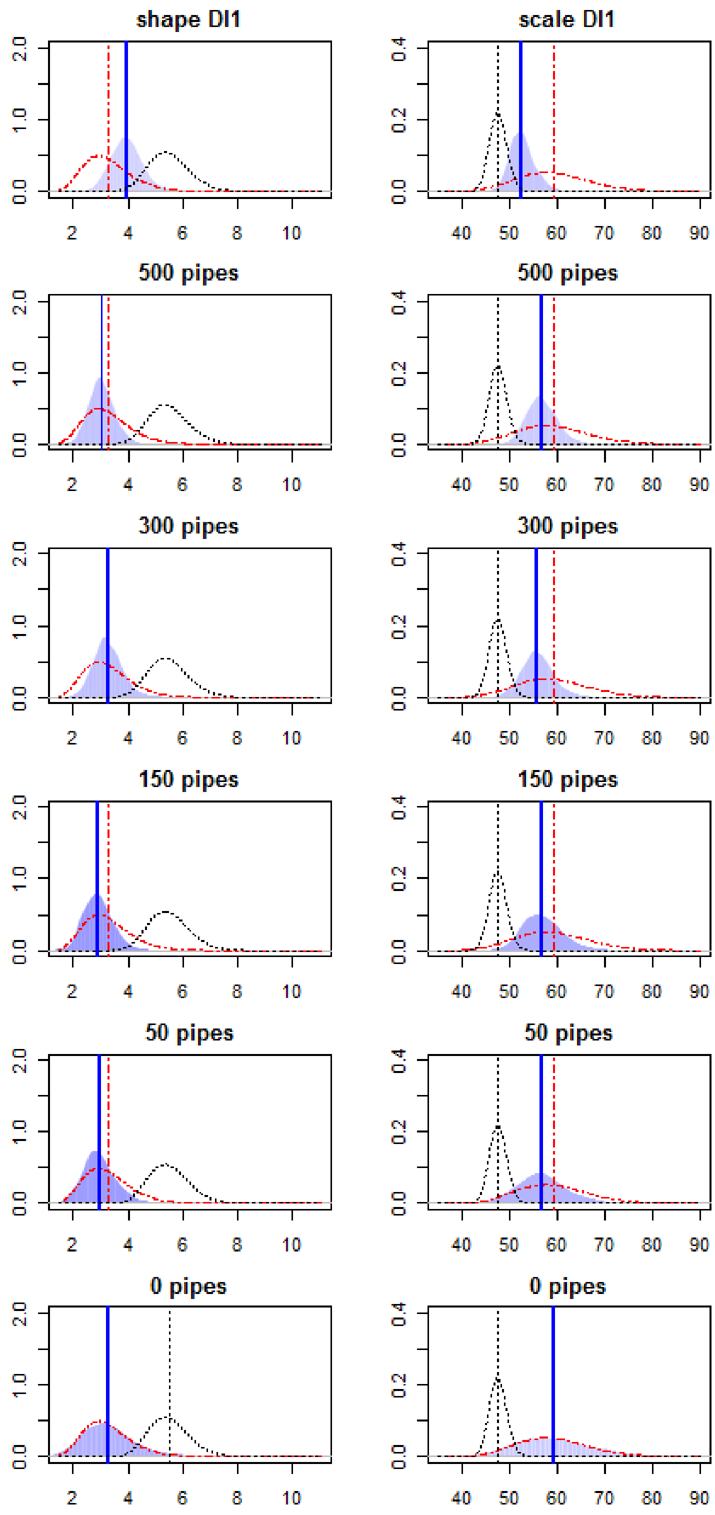

Figure 10: Bayesian inference with partial-pooling prior for ductile cast iron (DI1): Posterior (blue filled), and prior (red dash-dotted) marginal distributions of the Weibull shape (left column) and scale (right column) parameters for varying amounts of data (top level = all data). As a reference, the distributions resulting from MLE with all data (black dotted) are also plotted. Vertical lines indicate the position of the corresponding means.

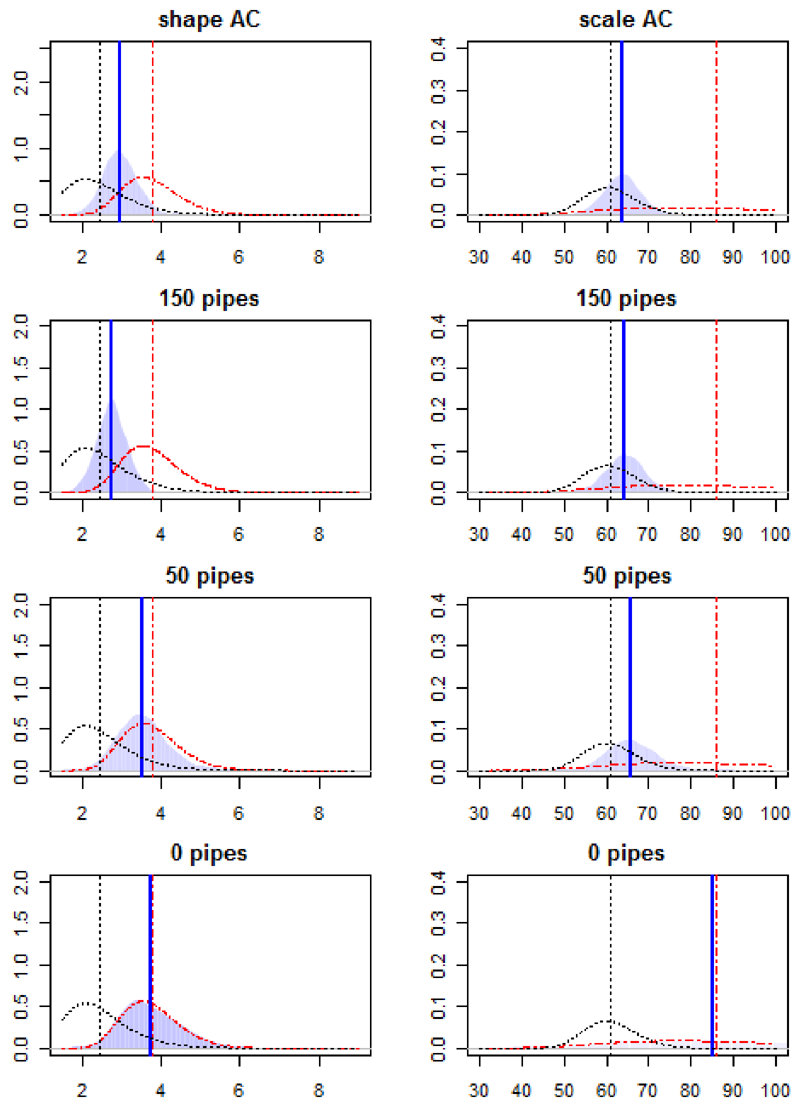

Figure 11: Bayesian inference with partial-pooling prior for asbestos cement: Posterior (blue filled), and prior (red dash-dotted) marginal distributions of the Weibull shape (left column) and scale (right column) parameters for varying amounts of data (top level = all data). As a reference, the distributions resulting from MLE with all data (black dotted) are also plotted. Vertical lines indicate the position of the corresponding means.

SUPPORTING INFORMATION TO: Scholten et al. 2013. *Combining expert knowledge and local data for improved service life modeling of water supply networks.* Environmental Modelling & Software 42: 1-16.

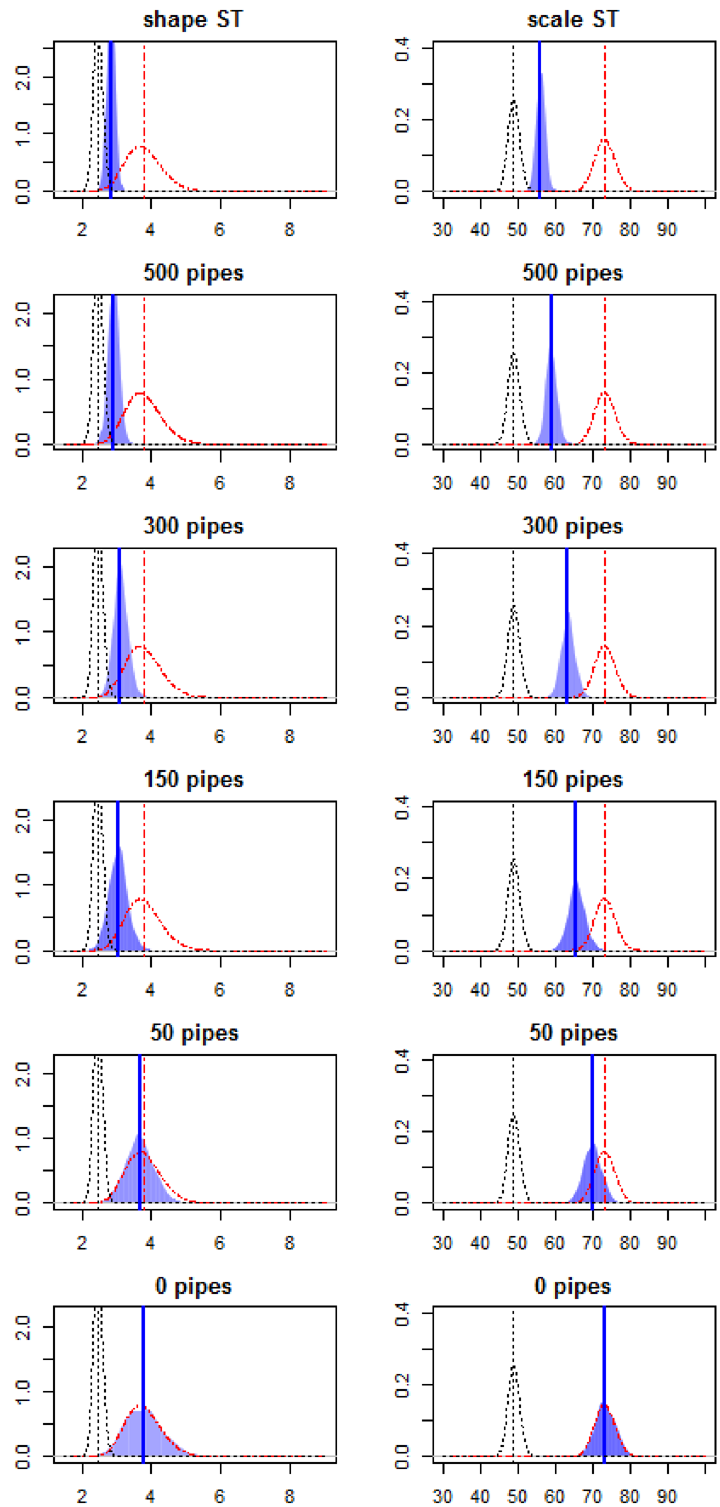

Figure 12: Bayesian inference with partial-pooling prior for steel: Posterior (blue filled), and prior (red dash-dotted) marginal distributions of the Weibull shape (left column) and scale (right column) parameters for varying amounts of data (top level = all data). As a reference, the distributions resulting from MLE with all data (black dotted) are also plotted. Vertical lines indicate the position of the corresponding means.

SUPPORTING INFORMATION TO: Scholten et al. 2013. *Combining expert knowledge and local data for improved service life modeling of water supply networks.* Environmental Modelling & Software 42: 1-16.

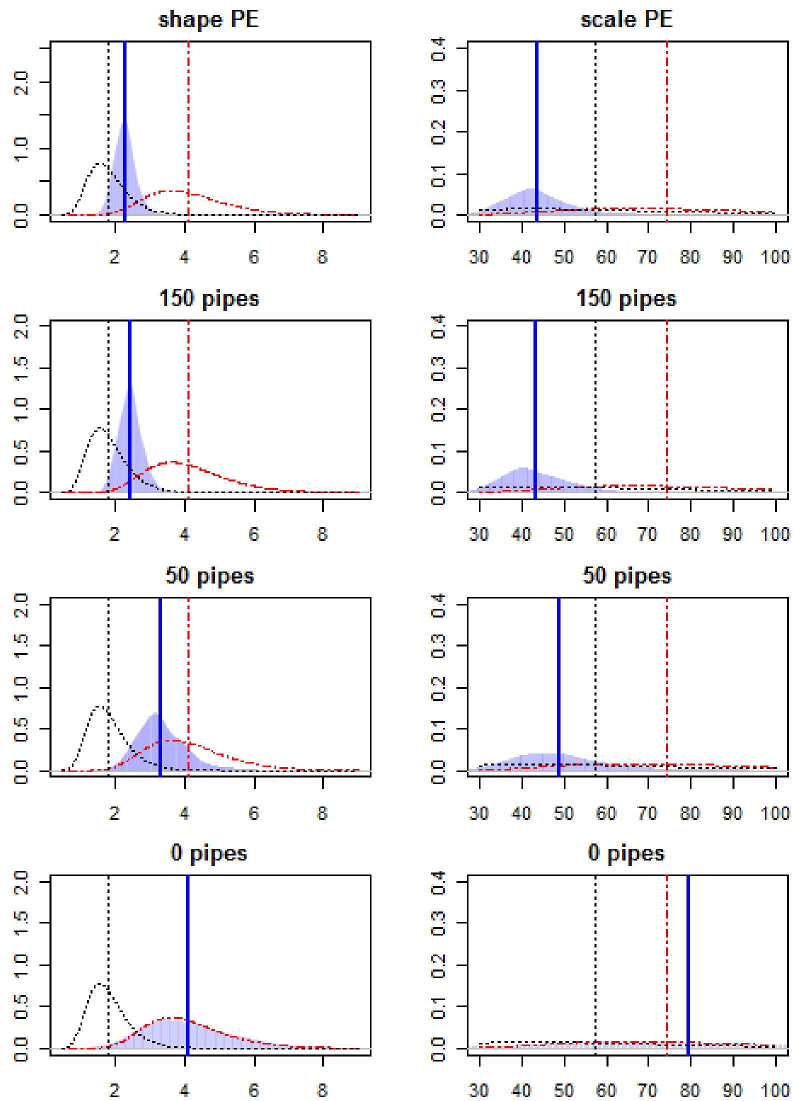

Figure 13: Bayesian inference with partial-pooling prior for polyethylene: Posterior (blue filled), and prior (red dash-dotted) marginal distributions of the Weibull shape (left column) and scale (right column) parameters for varying amounts of data (top level = all data). As a reference, the distributions resulting from MLE with all data (black dotted) are also plotted. Vertical lines indicate the position of the corresponding means.