

Bayesian updating of a prediction model for sewer degradation

H. Korving¹ and J.M. van Noortwijk²

¹ Witteveen+Bos Consulting Engineers, P.O. Box 233, 7400 AE, Deventer, The Netherlands, and Delft University of Technology, Faculty of Civil Engineering and Geosciences, P.O. Box 5048, 2600 GA, Delft, The Netherlands (E-mail: j.korving@witteveenbos.nl)

² HKV Consultants, P.O. Box 2120, 8203 AC Lelystad, The Netherlands, and Delft University of Technology, Faculty of Electrical Engineering, Mathematics and Computer Science, P.O. Box 5031, 2600 GA, Delft, The Netherlands (E-mail: j.m.van.noortwijk@hkv.nl)

Abstract

Sewer degradation is mainly a stochastic process. The future condition of sewers can be predicted with models based on condition states. In The Netherlands, the ‘SPIRIT’ model is being developed which combines expert opinion and visual inspections to predict sewer degradation. The statistical method implemented in this model is based on Bayesian statistics. The likelihood function of condition states is updated with new sewer inspections. A Dirichlet distribution is used to describe ‘subjective’ prior knowledge. Initially, prior knowledge only consists of expert opinions. The application of the statistical method is illustrated with an example. The results show that the model can be solved analytically, which reduces calculation time. In addition, expert opinions and inspections can be combined very easily and at any moment. Finally, the weight of experts and inspections is determined on the basis of the prior information and data instead of estimated by means of subjective expert knowledge.

Keywords

Asset management, Bayesian statistics, Dirichlet distribution, expert knowledge, informative prior, prediction model, sewer condition.

INTRODUCTION

Decisions on sewer rehabilitation rely on a prioritisation of sewers based on their condition states. Current conditions are measured by means of visual inspections and future conditions can be predicted with a sewer degradation model. Sewer degradation, however, is a stochastic process that varies widely and is generally not captured by available data. Therefore, changes in sewer condition are modelled not on the level of individual deterioration processes, but on the level of sewer condition states in which all explanatory processes are aggregated (e.g., Baur and Herz, 2002; Micevski et al., 2002). A comparable model is being developed in The Netherlands (Beenen et al., 2005). This model, called SPIRIT, is a prediction tool for sewer degradation. It combines expert opinions and sewer inspections in order to predict sewer condition in 5 and 10 years time.

This paper discusses the statistical model that is implemented in SPIRIT for predicting the condition of a sewer, given its current condition, based on both expert knowledge and inspections. Bayesian analysis has been applied for updating the prediction model. As a result, statistical uncertainty due to lack of data can be taken into account. The paper describes the applied methodology including the statistical model and updating procedure. The methodology is illustrated with an example.

STATISTICAL MODEL

Bayes’ theorem

Bayesian statistics models both natural variability and statistical uncertainty. Using Bayes’ theorem, a prior distribution can be updated as soon as new observations are available. Natural variability

represents the unpredictability of physical processes, such as rainfall. Statistical uncertainty results from a lack of data. The more new observations are available, the smaller the statistical uncertainty.

Bayes' theorem (i.e., the conditional probability theorem) can be written as

$$\pi(\theta|\mathbf{y}) = \frac{\ell(\mathbf{y}|\theta)\pi(\theta)}{\int_{\theta} \ell(\mathbf{y}|\theta)\pi(\theta)d\theta} = \frac{\ell(\mathbf{y}|\theta)\pi(\theta)}{\pi(\mathbf{y})}, \quad (1)$$

where $\pi(\theta|\mathbf{y})$ is the posterior density of $\theta = (\theta_1, \dots, \theta_d)'$ after observing data $\mathbf{y} = (y_1, \dots, y_n)'$, $\ell(\mathbf{y}|\theta)$ is the likelihood function of observations $\mathbf{y} = (y_1, \dots, y_n)'$ when the parametric vector $\theta = (\theta_1, \dots, \theta_d)'$ is known, $\pi(\theta)$ is the prior density of $\theta = (\theta_1, \dots, \theta_d)'$ before observing data $\mathbf{y} = (y_1, \dots, y_n)'$, and $\pi(\mathbf{y})$ is the marginal density of the observations $\mathbf{y} = (y_1, \dots, y_n)'$.

The likelihood function $\lambda(\mathbf{y}|\theta)$ represents the natural variability of a random quantity Y when θ is known. Conversely, the prior density $\pi(\theta)$ and the posterior density $\pi(\theta|\mathbf{y})$ represent statistical uncertainty of θ . This statistical uncertainty concerns parameter uncertainty. The likelihood of n independent observations y_1, y_2, \dots, y_n is defined as

$$\ell(\mathbf{y}|\theta) = \ell(y_1, y_2, \dots, y_n|\theta) = \prod_{i=1}^n \ell(y_i|\theta). \quad (2)$$

Likelihood function of condition-state inspection data

The probability of condition states are described with a multinomial distribution with 5 states $i = 1, 2, \dots, 5$ according to EN13508-2. The sum of probabilities of the condition states equals 1. When $k+1$ is the total number of condition states and p_i is the probability of condition state i , $i = 1, \dots, k$, the probability of condition state $k+1$ equals

$$p_{k+1} = 1 - \sum_{i=1}^k p_i \quad (3)$$

The random quantity Y_{ij} equals 1 when the j th inspection results in condition state i and 0 otherwise. The likelihood function of the j th inspection $\mathbf{y}_j = (y_{1j}, \dots, y_{kj})'$ is,

$$\Pr\{\mathbf{Y}_j = \mathbf{y}_j\} = \Pr\{Y_{1j} = y_{1j}, \dots, Y_{k+1,j} = y_{k+1,j}\} = \left(\prod_{i=1}^k p_i^{y_{ij}} \right) \cdot \left(1 - \sum_{i=1}^k p_i \right)^{y_{k+1,j}}. \quad (4)$$

Note that a random quantity is referred to as an upper case Y and an observation is referred to as a lower case y . Assuming that the n inspections are independent, the likelihood function of the inspection data y_{ij} can be rewritten as

$$\prod_{j=1}^n \Pr\{\mathbf{Y}_j = \mathbf{y}_j\} = \prod_{j=1}^n \Pr\{Y_{1j} = y_{1j}, \dots, Y_{k+1,j} = y_{k+1,j}\} = \prod_{j=1}^n \left(\prod_{i=1}^k p_i^{y_{ij}} \right) \cdot \left(1 - \sum_{i=1}^k p_i \right)^{y_{k+1,j}} \quad (5)$$

with $0 < p_i < 1$, $i = 1, \dots, k$, and $\sum_{i=1}^k p_i \leq 1$. This likelihood function is known as the multinomial distribution with parametric vector $\mathbf{p} = (p_1, \dots, p_k)'$. In simplified notation,

$$X_i = \sum_{j=1}^n Y_{ij}, \quad i = 1, \dots, k+1, \quad \mathbf{X} = (X_1, \dots, X_{k+1})', \quad \mathbf{x} = (x_1, \dots, x_{k+1})', \quad (6)$$

likelihood function (5) can be rewritten as,

$$\Pr\{\mathbf{X} = \mathbf{x}\} = \prod_{j=1}^n \Pr\{\mathbf{Y}_j = \mathbf{y}_j\} = \left(\prod_{i=1}^k p_i^{\sum_{j=1}^n y_{ij}} \right) \cdot \left(1 - \sum_{i=1}^k p_i \right)^{\sum_{j=1}^n y_{k+1,j}} = \left(\prod_{i=1}^k p_i^{x_i} \right) \cdot \left(1 - \sum_{i=1}^k p_i \right)^{x_{k+1}}. \quad (7)$$

The random quantity X_i represents the number of times in which during n inspections condition state i is observed. Note that

$$\sum_{i=1}^{k+1} X_i = \sum_{i=1}^{k+1} \sum_{j=1}^n Y_{ij} = n \quad (8)$$

Uncertainty of condition-state probabilities

The prior distribution of the parametric vector $\mathbf{p} = (p_1, \dots, p_k)'$ can be updated with the observations. This requires the formulation of a prior distribution expressing prior knowledge of these parameters. Using Bayes' theorem, the prior distribution of \mathbf{p} is updated to the posterior distribution given the observations $\mathbf{x} = (x_1, \dots, x_{k+1})'$.

Given the multinomial likelihood function,

$$\ell(\mathbf{x}|\mathbf{p}) = \Pr\{\mathbf{X} = \mathbf{x}|\mathbf{P} = \mathbf{p}\} = \prod_{j=1}^n \Pr\{\mathbf{Y}_j = \mathbf{y}_j\} = \left(\prod_{i=1}^k p_i^{x_i} \right) \cdot \left(1 - \sum_{i=1}^k p_i \right)^{x_{k+1}} \quad (9)$$

the question arises which type of prior distribution should be applied. In this case a 'conjugate prior' is chosen (DeGroot, 1970, Chapter 9; Bernardo and Smith, 1994, Chapter 3). The main advantage of conjugate priors is that the posterior can be calculated in explicit form. A limitation is that the prior distribution type affects the estimated posterior probability.

The conjugate distribution for the multinomial likelihood function is the Dirichlet distribution. A random vector $\mathbf{P} = (P_1, \dots, P_k)'$ has a Dirichlet distribution with parametric vector $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_{k+1})' > 0$ if the probability density function can be written as (Bernardo and Smith, 1994),

$$\pi(\mathbf{p}) = \pi(p_1, \dots, p_k) = \text{Di}_k(\mathbf{p}|\boldsymbol{\alpha}) = \frac{\Gamma\left(\sum_{i=1}^{k+1} \alpha_i\right)}{\prod_{i=1}^{k+1} \Gamma(\alpha_i)} \cdot \left(\prod_{i=1}^k p_i^{\alpha_i-1} \right) \cdot \left(1 - \sum_{i=1}^k p_i \right)^{\alpha_{k+1}-1} \quad (10)$$

with parameters $\alpha_i > 0$, $i = 1, \dots, k+1$, $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_{k+1})'$ and $\Gamma(\alpha)$ the gamma function which is defined as

$$\Gamma(\alpha) = \int_{t=0}^{\infty} t^{\alpha-1} \exp\{-t\} dt, \quad \alpha > 0. \quad (11)$$

Further, the following simplified notation is used,

$$\alpha_0 = \sum_{i=1}^{k+1} \alpha_i. \quad (12)$$

Examples of the application of the Dirichlet distribution can be found in Van Noortwijk et al. (1992), Van Noortwijk (2000) and Korving (2004).

Posterior distribution of condition-state probabilities

Assuming that the prior distribution of P is given by the Dirichlet distribution and the observations \mathbf{x} by the multinomial likelihood function, the posterior distribution of P can be calculated. Substituting likelihood function (5) in Bayes' theorem gives the posterior of P , which is also a Dirichlet distribution with parameters $\alpha_1 + x_1, \dots, \alpha_{k+1} + x_{k+1}$:

$$\pi(\mathbf{p}|\mathbf{x}) = \text{Di}_{k+1}(\mathbf{p}|\boldsymbol{\alpha} + \mathbf{x}) = \frac{\Gamma\left(\sum_{i=1}^{k+1} \alpha_i + x_i\right)}{\prod_{i=1}^{k+1} \Gamma(\alpha_i + x_i)} \cdot \left(\prod_{i=1}^k p_i^{\alpha_i + x_i - 1}\right) \cdot \left(1 - \sum_{i=1}^k p_i\right)^{\alpha_{k+1} + x_{k+1} - 1}. \quad (13)$$

The derivation can be found in e.g. DeGroot (1970, Chapter 9) or Bernardo and Smith (1994, Chapter 3).

The expected value of the posterior marginal distribution of the probability P_i is equal to (Bernardo and Smith, 1994),

$$E(P_i) = \frac{\alpha_i}{\sum_{i=1}^{k+1} \alpha_i} = \frac{\alpha_i}{\alpha_0}, \quad i = 1, \dots, k+1, \quad (14)$$

and the variance equals

$$\text{Var}(P_i) = \frac{\alpha_i(\alpha_0 - \alpha_i)}{\alpha_0^2(\alpha_0 + 1)}, \quad i = 1, \dots, k+1 \quad (15)$$

The statistical uncertainty in the probability of an individual condition state i can be estimated from the posterior density of P_i . The posterior density of the random quantity P_i can be described with a beta distribution with parameters α_i and $\alpha_0 - \alpha_i$ because for $k = 1$ the Dirichlet distribution reduces to a beta distribution (Bernardo and Smith, 1994),

$$\pi(p_i) = \text{Di}_1(p_i|\alpha_i, \alpha_0 - \alpha_i) = \text{Be}(p_i|\alpha_i, \alpha_0 - \alpha_i) = \frac{\Gamma(\alpha_0)}{\Gamma(\alpha_i)\Gamma(\alpha_0 - \alpha_i)} p_i^{\alpha_i - 1} (1 - p_i)^{\alpha_0 - \alpha_i - 1}. \quad (16)$$

The estimated statistical uncertainty is based on the 0.05 and 0.95 quantile (5th and 95th percentile) of this Beta distribution. The p quantile of the uncertain probability P_i written as q is calculated with the inverse of the beta distribution,

$$q = \text{Be}^{-1}(p|\alpha_i, \alpha_0 - \alpha_i). \quad (17)$$

The posterior mean of P_i can be written as the weighted average of the prior mean in (14) and the sample mean of the observations,

$$\hat{p}_i = \frac{x_i}{\sum_{i=1}^{k+1} x_i} = \frac{x_i}{n}, \quad i = 1, \dots, k+1, \quad (18)$$

with n , according to (8), equal to the number of inspections or observations. The weighted average can be simplified as,

$$E(P_i|\mathbf{x}) = \frac{\alpha_i + x_i}{\sum_{i=1}^{k+1} (\alpha_i + x_i)} = \frac{\alpha_0}{\alpha_0 + n} \cdot E(P_i) + \frac{n}{\alpha_0 + n} \cdot \hat{p}_i, \quad i = 1, \dots, k+1. \quad (19)$$

This formula shows the weight of the prior information,

$$w_n = \frac{\alpha_0}{\alpha_0 + n}, \quad (20)$$

and the weight of the observations,

$$1 - w_n = \frac{n}{\alpha_0 + n}. \quad (21)$$

The weight of the observations increases as the number of inspections n increases. For $n \rightarrow \infty$, the weight of the prior information approaches 0 and the weight of the observations approaches 1.

The elegance of the presented Bayesian approach is that α_0 represents the weight assigned to the prior information. Consequently, the prior information counts as α_0 observations. In Van Noortwijk et al. (1992), α_0 is called the number of virtual observations as opposed to the number of actual observations n .

Assessing the prior distribution of condition-state probabilities

In order to estimate the prior distribution of the uncertain probabilities of the condition states $\mathbf{P} = (P_1, \dots, P_k)'$, we use expert judgement. This expert opinion is available in terms of subjective estimates of the condition-state probabilities. We assume that the expected prior probability of condition state i equals r_i for $i = 1, \dots, k+1$,

$$E(P_i) = \frac{\alpha_i}{\sum_{i=1}^{k+1} \alpha_i} = \frac{\alpha_i}{\alpha_0} = r_i, \quad i = 1, \dots, k+1 \quad (22)$$

This means that all parameters of the Dirichlet distribution (10) can be expressed in terms of the number of virtual observations α_0 as follows

$$\alpha_i = r_i \alpha_0, \quad i = 1, \dots, k+1 \quad (23)$$

This leads to a variance and coefficient of variation of the condition-state probability which only depends on α_0 :

$$\text{Var}(P_i) = \frac{r_i(1-r_i)}{\alpha_0 + 1}, \quad \text{CV}(P_i) = \sqrt{\frac{1-r_i}{r_i} \cdot \frac{1}{\alpha_0 + 1}}, \quad i = 1, \dots, k+1 \quad (24)$$

Hence, the number of virtual observations α_0 controls $\text{Var}(P_i)$ and $\text{CV}(P_i)$, and remains to be determined. The larger α_0 , the smaller the prior uncertainty in the condition-state probabilities, and the more the prior information counts in determining the predictive probability of a condition state (see Equation (19)).

Van Noortwijk et al. (1992) propose the following three methods to subjectively assess the number of virtual observations α_0 :

1. Directly ask α_0 : i.e., ask as how many observations the expert opinion counts.

2. Test the sensitivity to new information using the posterior mean of a state probability P_i for some i . According to Equation (19) the posterior mean is a weighted average of the prior mean and the sample mean. Ask how the estimate of P_i would change if state i would be observed x_i times and solve Equation (19) for α_0 . To check the consistency, this can be done for several values of i .
3. Specify the length of the 90% uncertainty (confidence) interval for the beta-distributed P_i and solve for α_0 (numerically or by approximation; this can be done for several values of i). For an expert, estimating the 90% uncertainty interval of a condition-state probability is less abstract than estimating the corresponding variance or coefficient of variation using Equation (24).

For the moment, we propose to directly estimate the number of virtual observations α_0 and to let it be equal to the number of experts. The underlying assumptions are that each expert is assigned equal weight and the experts are statistically independent. However, more research is needed to justify these assumptions.

EXAMPLE

The presented Bayesian approach is applied to an example regarding sewer degradation. It is implemented in the prediction model SPIRIT (Beenen et al., 2005). The example concerns observation type ‘cracks and fractures’ (BAB) according to EN13508-2. Prior information (i.e. expert knowledge) expressed as a Dirichlet distribution of condition-state probabilities is updated using inspection data. The posterior distribution which is also a Dirichlet distribution is calculated with (13).

The available data consists of 2 experts and 9 ‘double inspections’, i.e. one inspection at $t = t_0$ and one at $t = t_0+5$ years. (Table 1 and Table 2, respectively).

Table 1. Expected sewer degradation in 5 years based on expert knowledge. Given the current condition of a sewer, the minimum, median and maximum degradation after 5 years is presented for each expert.

expert	condition state ($t = t_0$)	condition state ($t = t_0+5$ years)		
		min	median	max
1	3	3	3	3
2	3	3	3	4

Table 2. Observed condition states (cracks and fractures) at $t = t_0$ and $t = t_0+5$ years.

	state				
	1	2	3	4	5
$t = t_0$	-	-	9	-	-
$t = t_0+5$ years	0	0	4	4	1

In order to illustrate that the expected condition is affected by the available prior information, the number of ‘virtual observations’ is varied:

1. The prior distribution of condition-state probabilities is an informative Dirichlet distribution according to Equation (10) with $\alpha_0 = 2$ and minimum and maximum degradation (Table 1) interpreted as the 25th and 75th percentile.
2. The prior distribution of condition-state probabilities is an informative Dirichlet distribution according to Equation (10) with $\alpha_0 = 8$ and minimum and maximum degradation (Table 1) interpreted as the 25th and 75th percentile.

Expert weighted as single inspection

According to the assumptions in the SPIRIT model, the minimum and maximum condition state as determined by the experts are interpreted as the 25th and 75th percentile of the distribution function of condition states. From this, the expected condition-state probabilities after 5 years are calculated. Each expert is counted as 1 inspection. As a result, the total number of virtual observations equals 2.

The resulting posterior estimates of the condition-state probabilities are given in Figure 1. The figure shows the prior probabilities of condition states including the 90% uncertainty interval (5th and 95th percentile). The weight of prior information (i.e. expert knowledge) and inspections equals 0.18 and 0.82, respectively.

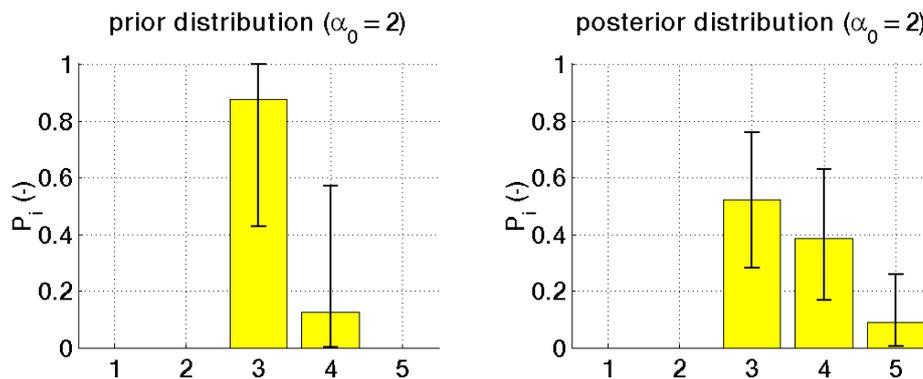


Figure 1. Prior and posterior probabilities of condition states for cracks and fractures with informative prior distribution (Dirichlet) and number of virtual observations $\alpha_0 = 2$.

Expert weighted as four inspections

When each expert is weighted as 4 inspections, the number of virtual observations equals $\alpha_0 = 8$. The resulting posterior estimates of the condition-state probabilities are given in Figure 2. Consequently, the weight of prior information (i.e. expert knowledge) and inspections changes and equals 0.47 and 0.53, respectively. By assigning a larger weight to the experts, the uncertainty interval narrows and prior information becomes more important.

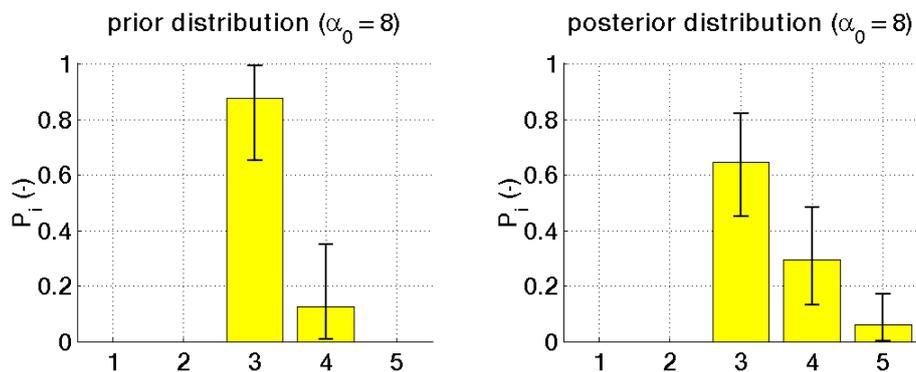


Figure 2. Prior and posterior probabilities of condition states for cracks and fractures with informative prior distribution (Dirichlet) and number of virtual observations $\alpha_0 = 8$.

The example shows that the condition-state probabilities in the SPIRIT model can be updated using Bayes' theorem. The expert judgements are considered as virtual observations. Important assumptions are that experts are independent and the weight assigned to each expert is equal. In

addition, the more virtual and actual observations are available, the smaller the uncertainty in the condition-state probabilities.

CONCLUSIONS

The objective of this paper is to describe a statistical model for predicting sewer condition based on a combination of expert knowledge and sewer inspections. The prediction model is updated with inspections using Bayesian statistics. A Dirichlet distribution is used to model prior knowledge on condition-state probabilities.

In order to estimate the prior Dirichlet distribution of condition-state probabilities, the following subjective information is needed:

- expected probability of condition states;
- weight of prior information expressed as ‘virtual observations’.

The uncertainty of the posterior distribution of condition-state probabilities is determined by the sum of virtual and actual observations.

The advantages of modelling statistical uncertainty of condition-state probabilities with a Dirichlet distribution are:

- The proposed Bayesian model can be solved analytically, which reduces calculation time.
- Expert opinions and inspections can be combined very easily and at any moment.
- The weight of experts and inspections is determined on the basis of the prior information and data instead of estimated by means of expert knowledge.

Currently, it is assumed that the number ‘virtual’ observations equal the number of experts. This requires that the opinion of each expert is given equal weight and the experts are independent. Expert opinions, however, can be dependent. Therefore, the assumption of independent and equally-weighted experts needs to be investigated and more research is needed to assess the number of virtual observations.

REFERENCES

- Baur R. and Herz R. (2002). Selective inspection planning with ageing forecast for sewer pipes. *Water Science and Technology*, 46(6-7): 389-396.
- Beenen, A.S., Vis, H., van Gestel, P.J., and Rabenort, A. (2005). SPIRIT: Towards an interactive tool for predictive asset management of sewer systems. In: *Proceedings of the 10th International Conference on Urban Drainage, Copenhagen/Denmark, 21-26 August 2005*.
- Bernardo, J.M., and Smith, A.F.M. (1994). *Bayesian Theory*. Chichester: John Wiley & Sons.
- DeGroot, M.H. (1970). *Optimal Statistical Decisions*. New York: McGraw-Hill.
- Korving, H. (2004). Probabilistic Assessment of the Performance of Combined Sewer Systems. PhD Thesis, Delft University of Technology, Delft.
- Micevski, T., Kuczera, G, and Coombes, P. (2002). Markov model for storm water pipe deterioration. *Journal of Infrastructure Systems*, 8(2): 49-56.
- Van Noortwijk, J.M. (2000). Optimal maintenance decisions on the basis of uncertain failure probabilities. *Journal of Quality in Maintenance Engineering*, 6(2):113-122.
- Van Noortwijk, J.M. Dekker, R., Cooke, R.M., and Mazzuchi, T.A. (1992). Expert judgment in maintenance optimization. *IEEE Transactions on Reliability*, 41(3): 427-432.