Nat. Hazards Earth Syst. Sci., 13, 737–754, 2013
www.nat-hazards-earth-syst-sci.net/13/737/2013/
doi:10.5194/nhess-13-737-2013

Natural Hazards and Earth System Sciences
Open Access

Estimation of insurance premiums for coverage against natural disaster risk: an application of Bayesian Inference

Y. Paudel, W. J. W. Botzen, and J. C. J. H. Aerts

Institute for Environmental Studies, Amsterdam Global Change Institute (AGCI), VU University Amsterdam, Amsterdam, the Netherlands

Correspondence to: Y. Paudel (ypl400@gmail.com)

Received: 20 July 2012 – Published in Nat. Hazards Earth Syst. Sci. Discuss.: –
Revised: 6 February 2013 – Accepted: 12 February 2013 – Published: 20 March 2013

Abstract. This study applies Bayesian Inference to estimate flood risk for 53 dyke ring areas in the Netherlands, and focuses particularly on the data scarcity and extreme behaviour of catastrophe risk. The probability density curves of flood damage are estimated through Monte Carlo simulations. Based on these results, flood insurance premiums are estimated using two different practical methods that each account in different ways for an insurer's risk aversion and the dispersion rate of loss data. This study is of practical relevance because insurers have been considering the introduction of flood insurance in the Netherlands, which is currently not generally available.

1 Introduction

The potential impacts of flooding in the Netherlands are large because high values of economic assets and many people are situated in low-lying areas exposed to flooding (Klijn et al., 2007; Aerts and Botzen, 2011; de Moel and Aerts, 2011). Although flood protection standards are high, new flood risk management strategies are currently being discussed to accommodate projected trends that increase flood risk, such as climate change and socio-economic growth (Kabat et al., 2005). An example is flood insurance, and since flood risk is not generally covered by insurance in the Netherlands, questions have been raised among insurers and policy makers as to whether flood insurance is feasible (Botzen and van den Bergh, 2008).

One of the issues related to flood insurance in the Netherlands is the uncertainty related to both the probability and the consequences of extreme (low probability) flood events (Froot, 2001). Insurers are reluctant to offer insurance coverage for a risk that is ambiguous, and for which sufficiently accurate premiums cannot be priced through actuarial calculations (Jaffee and Russell, 1997; Kunreuther and Michel-Kerjan, 2007). Moreover, insurers are risk-averse to catastrophe risk, which implies that they refuse coverage or are only willing to offer catastrophe insurance if premiums are sufficiently above the expected loss (Duncan and Myers, 2000). Standard statistical methods and tools for flood risk estimation use historical data and other empirical information (Behrens et al., 2004). This is often difficult due to a lack of empirical data, and the limited loss data available are not always reliable and consistent, so they inaccurately represent actual damage (Grossi and Kunreuther, 2005).

Several studies have applied flood assessment and hydrological models to estimate (future) flood damage and its probability for the Netherlands (Vrijling, 2001; Jonkman et al., 2003, 2008; Van der Most and Wehrung, 2005; Bouwer et al., 2010). Two relevant studies are "Veiligheid Nederland in Kaart" (VNK), also known as the FLORIS study (TAW, 2000; Wouters, 2005) and "Aandacht voor Veiligheid" (AVV) (Aerts et al., 2008; Aerts and Botzen, 2011). Both studies provide estimates of current and future flood probabilities and damage under various scenarios of climate change and economic development. A shortcoming of these studies is that they only provide a single estimate of the flood probability and potential flood damage, and not the complete probability density function of damage. Moreover, because extreme events such as catastrophe damage generally follow an asymmetric distribution process with a fat right-tail, the losses located in this part of the damage distribution need to be included in risk estimates and insurance premiums. The average risk estimates from the VNK and AVV projects may therefore lead to either "upwards" or "downwards" biased risk and premium estimates. As catastrophe risks are generally assumed to follow a fat-tail distribution process, and risk estimation has to rely on limited empirical data, it is essential to take these two aspects into account in the risk assessment process. Bayesian Inference (BI) is a useful statistical tool for assessing risk, especially in applications with a lack of historical information on risk, as is the case for low-probability floods in the Netherlands (Coles and Powell, 1996; Cooley et al., 2007). BI can be used to update the probability estimate for a hypothesis (the "prior beliefs", i.e. based on AVV flood risk information) with additional empirical information (the "likelihood", i.e. based on VNK flood risk information), especially when it is unreliable to make statistical inference based on only a small amount of homogeneous data, such as either VNK or AVV flood risk information. The updated outcome (the posterior statistics) can be used to simulate hazard events and fit a risk distribution, i.e. to estimate a new probability density function for flood damage. The BI method used in this study deals with data scarcity and provides insights into uncertainties in the estimated flood damage.

Under the assumption that flood risk follows a Pareto distribution with a right fat-tail, this study aims to apply Bayesian techniques combining the VNK and AVV data to estimate the tail parameter of the flood damage distribution. In the first phase, the tail parameter for flood damage is estimated using BI, in which the prior belief is assumed to originate from a Gamma conjugate [1] family while the observations are assumed to follow a Pareto distribution. Subsequently, in the second phase, flood damage is simulated based on the estimated tail parameter, and the corresponding probability density functions are fitted for all 53 areas in the Netherlands which are exposed to flooding ("dyke ring areas"). Kunreuther et al. (2009) propose that premiums that reflect risks are an important condition for designing a natural disaster insurance system that is financially viable and provides adequate incentives for risk mitigation. Although we realize that flood insurance premiums in practice may not be fully differentiated with respect to flood risk, our analysis provides insights into the level of flood insurance premiums as if the condition of risk-based premiums were applied in the Netherlands. Using the loss information derived from the damage simulation and density functions, the risk-based insurance premiums (Actuarial-equivalence (AE amount)) will be estimated for all 53 dyke ring areas. These premiums include an extra surcharge for insurer's aversion to catastrophe risk, as proposed by Kaas et al. (2004). Finally, the AE premiums are compared with the Empirical method proposed by Kunreuther et al. (2011), which takes a modest rate for insurer's aversion to (fat-tailed) catastrophe risk into account.

The remainder of this paper is structured as follows. Section 2 describes the data and the statistical methods used to estimate the parameters of the probability density functions of flood damage, and the two methods used to calculate insurance premiums. Section 3 presents the results of the flood risk estimates and flood insurance premiums, and compares these results with existing studies. Section 4 provides discusses our findings and Sect. 5 concludes.

2 Data and methods

Figure 1 shows the overall research methodology described in detail from Sect. 2.1 to Sect. 2.4. In Sect. 2.1, the available data from AVV and VNK are described, and, if necessary, further processed and adapted for use. In Sect. 2.2, the BI is introduced and applied to estimate the tail (shape) parameter of the flood damage distribution under the assumption that flood damage follows a Pareto distribution process. Next, in Sect. 2.3, a Monte Carlo simulation of flood damage is performed using the updated tail parameters from Sect. 2.2, and subsequently probability density curves of flood damage are fitted for all 53 dyke ring areas. These sections explain in detail the probability distributions that are used in these steps. Finally, in Sect. 2.4, the resulting simulated flood damage information is used for deriving flood insurance premiums. Premiums are estimated using two different methods that respectively use the variance of risk (derived from Sect. 2.3) in a different way, and take a different account of the insurer's risk aversion to catastrophic flood risk.

2.1 Data

The low-lying areas in the Netherlands are divided into 53 dyke ring areas. Many low-lying lands (which are often called "polders") have been reclaimed from former lakes. Each dyke ring area has its own closed flood protection system of dykes, dams, and sluices that protect it from floods caused by rivers and the sea. A dyke ring is an individual administrative unit under the Water Embankment Act of 1995, which guarantees a particular level of protection against flood risk for each dyke ring area (Aerts and Botzen, 2011). For instance, a dyke ring with a safety standard of 1/1250 (a flood "return period" of once in 1250 yr)[2] has been constructed in such a way that it may withstand a flood with a probability of 1/1250. Figure 2 shows a map of the Netherlands that depicts the 53 dyke ring areas and their safety standards, which range between 1/10 000, and 1/1250. The data

[1] Conjugacy can be defined as follows: if F is a class of sampling density functions $p(x|\theta)$ and P is a class of prior distributions for θ, then the class p is conjugate for F if $p(\theta|x) \in p$ for all $P(.|\theta) \in F$ and $p(.) \in P$ (Juneja et al., 2006).

[2] The return or recurrence period is an estimate of the average time internal between two occurrences of an event, in particular, of a catastrophe, such as a flood, an earthquake or a peak river discharge. The probabilities provided by the VNK and AVV projects are an indication of the approximated return period of a dyke breach.

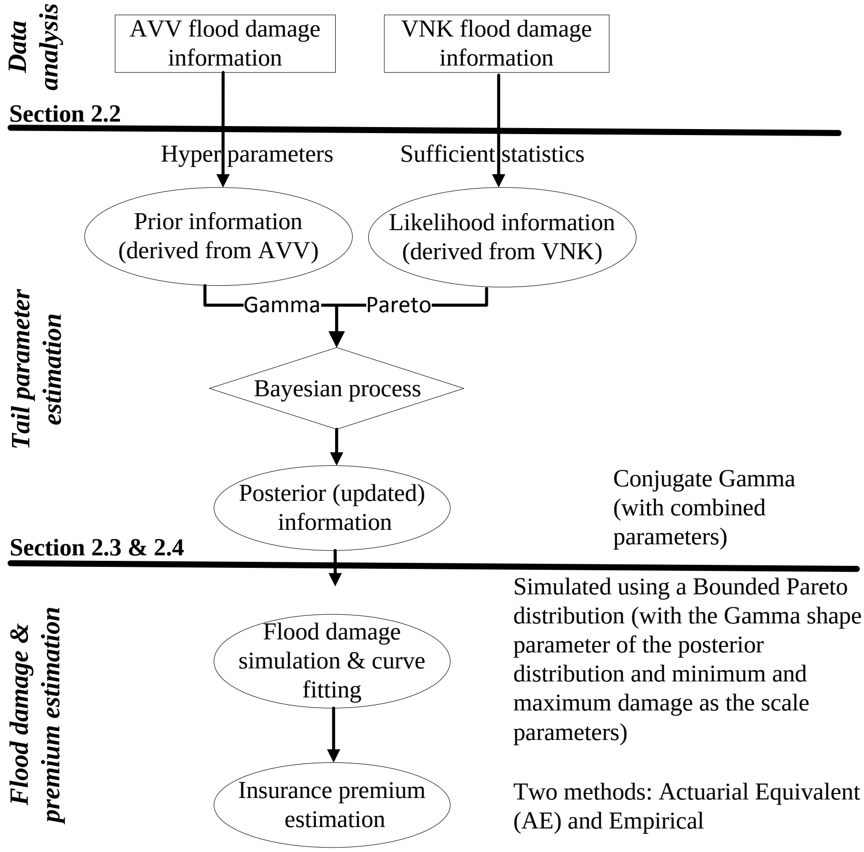

Fig. 1. Conceptual model of the overall research methodology.

used in this paper originate from two main studies, the AVV and VNK projects, on flood risk in the Netherlands.

The VNK project provides current flood risks, and the results have been reported by Wouters (2005), and by Bouwer et al. (2009) (See Table 1). VNK estimated detailed flood damage according to various flood scenarios for dyke-rings 7, 14 and 36 (Fig. 2), while a more global approach was used to estimate flood probabilities and potential flood damage for the other dyke ring areas. The study included the assessment of flood probability and damage from dyke failure mechanisms, hydraulic pressure, and multiple dyke breach scenarios (Wouters, 2005). The AVV study provides insights into the potential effects of climate and socio-economic change on flood risk over a long time horizon, as compared with the current situation. The AVV project estimates current flood risk and future flood risk was simulated using (long-term) land use and climate change scenarios of increased river discharges and sea level rise (Aerts et al., 2008; Aerts and Botzen, 2011). Appendix A provides all the data from the AVV and VNK projects, which have been used as input for the Bayesian model and flood damage simulations. The flood damage estimates from both projects are used for the BI method, while the AVV probabilities are used for the estimation of annual flood risk and premiums. Throughout this paper, the AVV flood damage data represents prior knowledge about the probability distribution of flood damage within the Bayesian framework, and the VNK flood damage data represents the flood damage likelihood (See Fig. 1). Table 1 provides the basic information provided by these two main studies on flood risk in the Netherlands for three representative dyke ring areas. Dyke ring areas 7 and 36 are representative for most of the dyke ring areas that have flood protection levels between 1/4000 and 1/1250 per year. The dyke ring Zuid Holland (along with Noord Holland) is one of the two dyke-rings with the lowest flood probability in the Netherlands (1/10 000). This dyke ring is located along the densely populated coastline, and has a high concentration of economic assets.

Data was processed (for details, see Appendix B) to derive minimum and maximum damage per dyke ring area, which are used in the flood damage simulations (Sects. 2.2 and 2.3). The assumption of a minimum and maximum amount of flood damage per dyke ring area is consistent with realty. If there is a flood event in the Netherlands, then there will always be a minimum amount of damage, while the maximum theoretical damage cannot exceed the total economic value that is exposed to flooding within a dyke ring area. The AVV and VNK data provide expected damage estimates for

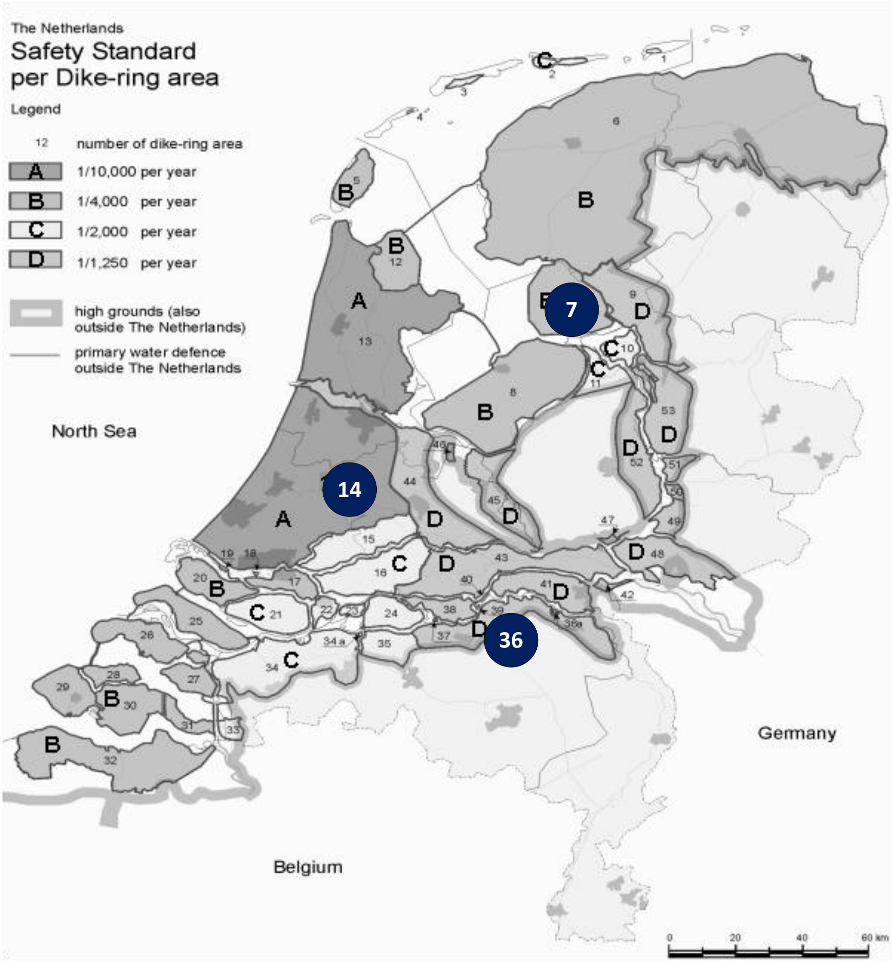

Fig. 2. Safety standards of dyke ring areas in the Netherlands (Source: TAW, 2000).

Table 1. Example of flood damage information for three dyke ring areas provided by the VNK and AVV flood risk studies in the Netherlands (damage in million euros).

Dyke ring area number	VNK			AVV	
	Minimum flood damage	Maximum flood damage	Expected flood damage	Expected flood damage	Flood probability
7	593	5301	2000	2665	0.00025
14	5254	46 950.6	18 500	23 600	0.0001
36	1073	9592.2	2800	4822	0.0008

53 dyke ring areas based on the current probability of dyke overtopping (the exceedance probability), and no minimum and maximum damage estimates, except for the three dyke ring areas, 7, 14 and 36, by VNK. Since it is assumed that flood damage is truncated on both sides, the minimum and the maximum damage for all dyke ring areas have been derived using the VNK information about dyke ring areas 7, 14 and 36. For this, the weighted upscaling and downscaling factors are estimated using the minimum and the maximum flood damage estimates for these 3 dyke-rings. These scaling factors are used to estimate the minimum and maximum flood damage for each dyke ring area under the assumption that these are proportional to the expected damage of each dyke ring (see Appendix B).

2.2 Applying Bayesian Inference for estimation of the tail parameter

Assume θ is an unknown parameter – which reflects the state of our knowledge about the flood damage data before we observed this data – defined as an element of a space Θ which denotes all possible states of an outcome, also called the parameter space. When experimenters aim to obtain more information about this unknown parameter, the flood damage observations, $x_1, \ldots\ldots x_n$, per flood return period n are assumed to be independent and identically distributed from a Pareto process where the process scale parameter $(\hat{\beta})$ is assumed to be known, and the process tail (shape) parameter $\hat{\theta}$ is unknown (see Sect. 2.2). BI is a statistical technique which attempts to estimate parameter θ by combining the prior beliefs (this means that we have some idea about it before we have seen the data) with the information observed from experiments or practice. This Bayesian technique can be represented by the Bayes theorem:

$$p(\theta\,|\,x) = \frac{p(\theta)\cdot p(x|\theta)}{p(x)} \ldots \;\; \ldots \;\; \ldots, \tag{1}$$

where $p(\theta)$ on $\theta \in \Theta$ is the prior distribution which includes our prior knowledge about θ (the AVV information); $p(x|\theta)$ is the sampling or data distribution; and $p(x)$ is the marginal distribution of x. The observed data (the VNK information) combined with the prior knowledge is given by the likelihood function $L(\theta\,|\,x_1, \ldots\ldots x_n) = p(x|\theta)$, which can be regarded as a density with respect to x, where x is fixed to a particular value (the observation). The normalized product in Eq. (1), which is the marginal distribution of x that can be represented as $p(x) = \int p(\theta) L(\theta\,|\,x)\, d\theta$, integrates over all possible values of θ for given x. Because $p(x)$ is a scalar value and not a function, we can omit the denominator in Eq. (1). The non-normalized[3] posterior distribution can be given as one of proportionality:

$$p(\theta\,|\,x) \propto p(\theta)\cdot p(x\,|\,\theta) \ldots \;\; \ldots \;\; \ldots. \tag{2}$$

The main steps in BI to derive the posterior density (the updated shape parameter for the flood damage density) for θ (Eq. 1) are:

Step 1. Formulate the likelihood or data model for the observations (the VNK damage data) and the corresponding likelihood function for the unknown parameters of the flood damage distribution;

Step 2. Specify the prior density distribution (the AVV damage data) to quantify the uncertainty about the values of the unknown parameters identified in Step 1;

Step 3. Combine the likelihood function from Step 1 with the prior distribution from Step 2 to determine the posterior distribution. The posterior distribution quantifies the joint probabilities of the values of unknown parameters of the flood damage distribution using the information from observations;

Step 4. Calculate the updated parameters of the probability density function of flood damage for each of the 53 dyke ring areas from the posterior distribution by repeating the previous steps. The mean of the parameter (the parameter of interest) can be easily derived from the posterior distribution by calculating the expectation.

We have made two main assumptions in this study to apply BI. First, it is assumed that flood loss data are independent and identically Pareto-distributed (bounded with a lower and upper threshold) continuous random numbers (See Eq. 3 in Step 1), with an unknown shape and the known scale parameter (see Appendix C). Second, for computational convenience we take the prior and posterior models from a conjugate family to estimate the shape-parameter. Conjugate in our case means that the prior and posterior distributions are from the same distribution family, which is the case in this research. To briefly elaborate on this topic: even though the Bayesian theorem is mathematically simple, due the normalizing factor (the denominator in Eq. 1) it can be a difficult task to find an analytical or numerical solution. Because the Bayesian theorem is a product of the prior and the likelihood functions, it is not always guaranteed that this product can be integrated over the relevant domain. One way to avoid this problem is by using conjugate priors. Based on this concept one can derive pairs of likelihood functions and prior distributions with appropriate mathematical properties that result in tractable closed-form solutions to the integrals (Arnold and Press, 1983). Since conjugate priors have computational advantages, we have chosen pairs of the likelihood function and the prior distribution from an informative[4] conjugate family, as will be explained below.

Step 1: Data model and the likelihood function

We first apply the Pareto method as it is useful for modelling low-probability high-impact risks, because the total aggregated damage is to a large extent determined by large losses in the right-tail of the density, which are also called "right-tail risks" (Hsieh, 2004). For this application, we assume that the VNK flood loss data (the data model) $x_1, \ldots\ldots x_n$ are an independent and identically distributed Pareto random loss variable per flood return period that is defined with two main parameters: namely, the process scale x_{m}, which is the minimum of x, and the process shape $\theta > 1$. It is assumed that

[3] The normalizing factor $p(x)$ is a constant that makes the posterior density integrate to 1. It is often omitted in practice since it is generally difficult to calculate.

[4] An informative prior expresses specific, definite information about a variable, while a non-informative prior expresses vague general information about a variable (prior) (see for more detail Juneja et al., 2006).

the first parameter is known and that the shape parameter is unknown, which we aim to update using BI in this study. The Pareto distribution can be written as (Davis, 2001)

$$p(x)=\left\{\begin{array}{ll}\theta\frac{x_m^{\theta}}{x^{\theta+1}}I(x>x_m) & \text{where } x_i\geq x_m\\ 0 & \text{otherwise}\end{array}\right\}\ldots\ \ldots\ \ldots\ \ldots, \quad (3)$$

where $\theta > 1$. [5]

The likelihood function associated with the unknown shape parameter is (Bermudez and Kotz, 2010)

$$L(\theta|x_1,\ldots\ldots x_n) \quad (4)$$
$$=\left\{\begin{array}{ll}\theta^n x_m^{\theta n}\prod_{i=1}^{n}x_i^{-(\theta+1)} & \text{where } x_i\geq x_m\\ 0 & \text{otherwise}\end{array}\right\}\ldots\ \ldots\ \ldots\ \ldots.$$

The sufficient convenient statistics[6] are based on the number of observations, n, and the data product

$\prod_{i=1}^{n} x_i$.

Step 2: Prior density

Once the data model has been specified, the prior density function for the unknown model parameter needs to be specified. The prior density describes our beliefs about the uncertainty about the model parameters, without incorporating the information from the observations. Since it is assumed that the model data follows a Pareto density with known scale and unknown shape parameters, the conjugate prior $p(\theta)$ is proportional to the Gamma density function. This Gamma function of θ is defined by the hyper-parameters a (shape) and b (scale) and is given by (Fink, 1997):

$$p(\theta)=\left\{\begin{array}{ll}\frac{(\theta)^{a-1}\frac{e^{\left(-\frac{\theta}{b}\right)}}{b^a\Gamma(a)}}{0} & \begin{array}{l}\text{for } \theta\geq 0 \text{ and } a,b>0 \text{ and}\\ \text{for } \theta<0\end{array}\end{array}\right\}\ldots\ \ldots\ \ldots\ \ldots, \quad (5)$$

where a and b are, respectively, the shape and scale estimates of the hyper-parameters, which are derived from the AVV flood damage estimates (see Appendix C).

Step 3: Posterior density

Because we have chosen a prior density that is conjugate for the likelihood, the posterior density consists of a combination of Eqs. (4) and (5), and follows the same density as the prior, namely: $p(\theta\,|\,x)=\text{gamma}(\alpha,\beta)=L(\theta|X)\cdot p(\theta)$ (Vilar-Zanon and Lozano-Colomer, 2007). Substituting Eqs. (4) and (5) in Eq. (2) yields the following specification of the posterior flood damage density (see Appendix D):

$$p(\theta\,|\,x)\propto\frac{(\theta)^{\alpha-1}e^{-\left(\frac{\theta}{\beta}\right)}}{\beta^{\alpha}\Gamma(\alpha)}\ldots\ \ldots\ \ldots\ \ldots, \quad (6)$$

where $\alpha=a+n$ and $\beta=\frac{1}{\frac{1}{b}+\ln(\prod_{i=1}^{n}x_i)-n\cdot ln(x_m)}$, and $\beta<\frac{1}{(n\cdot\ln(x_m)-\ln(\prod_{i=1}^{n}x_i))}$ (Arnold et al., 1998).

Step 4: Estimation of the parameters of interest

From the posterior distribution function specified in Step 3, the parameter of interest (the mean) of the posterior distribution of flood damage is given as follows[7]:

$$\hat{\theta}=E[p(\theta\,|\,x)]=\int_{-\infty}^{\infty}\theta\,p(\theta|x)d\theta\ldots\ \ldots\ \ldots\ \ldots. \quad (7)$$

To estimate this parameter of interest $\hat{\theta}$, we first need to simulate the posterior distribution. As discussed by Gelman (2004), this can be simulated in different ways. Because the prior distribution in this paper is assumed to be from a conjugate family, we opt for a direct simulation method, which implies the drawing of random numbers from the target (i.e. posterior) distribution (Gelman et al., 2004)[8].

2.3 Simulation of flood damage data and the fitting of probability density curves of flood damage

Once the updated tail parameters are estimated (the parameter of interest in Eq. 7), they are used, along with the updated scaling parameters for minimum and maximum damage (see Appendix B), to simulate flood damage for all 53 dyke ring areas for 250 000 flood return periods, as well as to estimate the corresponding probability density curves of flood damage. Assuming that the flood damage distribution follows the Pareto distribution as in Sect. 2.2, the probability density for each of the 53 dyke ring areas can be simulated with the following model:

$$\text{Flood damage}=f^{-1}\left(\hat{\theta},x_l,x_u\right)\ldots\ \ldots\ \ldots, \quad (8)$$

where $f(x)$ is a bounded Pareto distribution (see Appendix E for more details) with shape parameter $\hat{\theta}=E[P(\theta\,|\,x)]$, the lower bound scale parameter $x_l=\min(x)$ and the upper-bound scale parameter $x_u=\max?(x)$, which are estimated

[5] In practice, in case of sufficient data, the shape parameter of the data model can be estimated numerically from the VNK estimates with the formula $\theta=2\cdot\left(\sum_{i=1}^{n}x_{i\cdot wi}\right)^2\Big/\left\{\left(\sum_{i=1}^{n}x_{i\cdot w_i}\right)^2-\sigma^2\right\}$, with $\left(\sum_{i=1}^{n}x_{i\cdot wi}\right)^2-\sigma^2>0$. The scale parameter can be estimated with x_m (scale) $=\inf(X)$, with $X>m$. (See Appendix C for more details).

[6] A sufficient statistic has the property of sufficiency in terms of the related statistical model and its unknown parameters.

[7] Although $\hat{\theta}$ can be estimated as the product $a\cdot b$, due to the limited available data we choose to simulate $\hat{\theta}$ using the posterior distribution.

[8] For the exact procedure see Gelman (2004), pages 290–292. One can also choose to estimate θ as the product of the two hyper-parameters. However, here we decided to simulate θ by drawing random numbers from the posterior distribution. Subsequently, we derive the expectancy of θ from the simulated data.

from the VNK data for dyke ring areas 7, 14 and 36 (for more details see Appendices A and B) (Kerman and Gelman, 2007)[9].

These models were implemented in the statistical software Matlab version 2012a to perform a Monte Carlo analysis of flood damage for each dyke ring area for 250 000 flood return periods. The resulting data are used to construct probability density curves for flood damage and to estimate average flood damage over all flood return periods for each of the 53 dyke ring areas. The expected AE average amount for flood damage (the average overall flood return periods) can be represented with the standard expectancy expression $E[\mathrm{AE}]_{\text{return period}} = \sum_{n=1}^{250\,000} (x_n . p\,(x)_n)$. The annual average flood damage per dyke ring area can be derived by dividing the average $\mathrm{E[AE]}_{\text{return period}}$ per dyke ring area with the return period of a dyke breach in years for the corresponding dyke ring from AVV (as presented in column 2, Table A1 in Appendix A), $E\,[\mathrm{AE}]_{\text{annual}} = \frac{E[\mathrm{AE}]_{\text{return period}}}{\text{return periods in yr}}$.

2.4 Estimation of flood insurance premiums

Since floods are rare events, the estimates for annual flood insurance premiums can be approximated by the annual expected value of flood damage. Therefore, based on the simulated probability density curves, we can estimate average flood insurance premiums for each dyke ring area. A more refined premium differentiation on the household level is not possible, given the available information on flood risk. Hence, the premiums are partially risk-based, since they differ per dyke ring area, but not per individual insurance policy.

Since floods are rare events, the estimates for annual flood insurance premiums can be approximated by the annual expected value of flood damage, and should consider the level of insurer's risk aversion towards the extreme nature of the risk. This risk aversion is reflected as a surcharge on the premium above the expected value of the loss. This surcharge depends on the variability (variance or standard deviation) of the expected flood damage. A higher risk variance implies a higher probable maximum loss, which leads to an extra premium surcharge. In general, insurers charge a higher premium if the variability of losses is greater, because a relatively high variability indicates a high likelihood of suffering very large losses for which large cash reserves or reinsurance coverage is needed, which is costly for the insurer. This surcharge has the effect to increase the cash surplus of the insurer which protects the insurer against the possibility of insolvency (Kunreuther et al., 2011). We apply two different methods for estimating premiums that address the impact of the insurer's risk aversion attitude towards the extreme (catastrophic) nature of flood risk. The first method, AE, provides premium estimates based on the Modern Actuarial Risk Theory discussed in Kaas et al. (2004). The main emphasis of this method is on the extreme nature of damage, because it takes the full loss variance into account in the premium calculation. Moreover, a moderate degree of insurer's risk aversion rate is included in the premium, which is quite common for non-life insurance products, but may provide an underestimation in an application to heavy-tailed catastrophe losses, like those due to floods. The premium amount π to be paid by the policyholder for insurance coverage for risk X, given the exponential utility function $u\,(x) = -\alpha e^{-\alpha w}$ with the parameter of insurer risk aversion $\alpha = 0.005$, is given by (for details, see Kaas et al., 2004)

$$\pi\,(W) = E\,[\mathrm{AE}]_{\text{annual}} + \frac{1}{2} \cdot \sigma^2 \cdot r \ldots \quad \ldots \quad \ldots, \tag{9}$$

where $\pi\,(W)$ is the amount of premium necessary to insure property value W; $E\,[X]$ is the loss expectancy that can be deduced from the adjusted Pareto distribution (Eq. 8), σ^2 is the variance of the same adjusted Pareto distribution; $r\,(w) = -\frac{u''(w)}{u'(w)}$ is the insurer's risk aversion coefficient, which is in this case equal to α, and w stands for the amount to be insured; and $u(.)$ is the utility function of the insurer. Equation (8) is derived in Appendix F.

The second method is what we call the Empirical method, which has been proposed by Kunreuther et al. (2011). This method includes a surcharge of the standard deviation on the premium, which has been derived from an extensive empirical analysis of catastrophe insurance premiums in the USA for a period of 24 yr (see Chapt. 7 in Kunreuther et al., 2009). This method is a modified version of Eq. (9), and it takes the insurer's risk aversion toward the extreme nature of risk into account by making the premium dependent on the standard deviation (SD) of damage. According to the Empirical method, the premium can be calculated as follows:

$$\pi\,(\mathrm{W}) = \mathrm{E}\,[\mathrm{AE}]_{\text{annual}} + \sigma \cdot \delta \ldots \quad \ldots \quad \ldots, \tag{10}$$

where σ is the standard deviation of loss; and δ is the Empirical insurer's risk aversion rate for catastrophe risk. The coefficient δ is equal to 0.55 (Kunreuther et al., 2011).

3 Results

Descriptive statistics of the probability density functions for flood damage are provided for all 53 dyke ring areas (see Table 2). Columns 2 and 3 in Table 2 show, respectively, the estimated minimum and maximum flood damage amounts; columns 4, 5 and 6 are, respectively, the expected flood damage, the standard deviation of the simulated flood damage and the parameter of interest. Although differences between dyke ring areas in the shape parameters in column 6 appear to be small, it should be noted that the distribution function

[9]The zero failure problem is avoided here by defining the flood loss data x per flood return period. This means that in case of a flood event, which happens once over return period, there will always be some flood damage with a magnitude that lies between x_{l} and x_{u}.

used is sensitive to such small differences. Column 4 in Table 2 shows that expected flood damage amounts differ significantly between dyke ring areas. The expected flood damage provided by VNK and AVV (see Appendix A) is significantly higher compared with our estimates. The reason for this is that the damage estimates by AVV and VNK are expected flood damage amounts that are associated with what are called an "exceedance flood probability" of a very severe flood, which do not include damage amounts that fall below this extreme level. Our expected flood damage amounts are lower than the AVV and VNK expected flood damage amounts, because our estimates are based on the full probability density functions of damage between the specified minimum and maximum amounts, which include several small damage amounts.

Due to space limitations, it is unfeasible to discuss the detailed results of all 53 dyke ring areas, which is why the results for three representative dyke ring areas – the Noordoostpolder (7), Zuid Holland (14), and Land van Heusden/de Maaskant (36) – are discussed here in more detail.

Figures 3, 4 and 5 show, respectively, the resulting probability densities of flood damage for the three selected dyke ring areas. The corresponding statistics are provided in Table 3 below. The three markers in the figures indicate some representative data percentiles (the 50 per cent, the mean, and the 97.5 per cent percentiles). The loss densities are truncated on the left and right sides, in accordance with the estimates of minimum and maximum flood damage (see Appendix B). The probability of observing a damage amount is depicted on the left vertical axis, and flood frequencies for 250 000 flood return periods are shown on the right vertical axis, while the damage amounts are shown on the horizontal axis in millions of euros. The frequency densities show that the majority of loss observations in each dyke ring area are concentrated on the left-side of the curve, while every frequency curve has a long fat right-tail, which indicates a high dispersion of the loss data. As an illustration, the statistical mean of all three probability density functions of flood damage are located around the 67.9 per cent data percentile, which indicates that the loss data behave asymmetrically. This is consistent with our selection of the Pareto distribution which is fat-tailed and asymmetrical and corresponds with practical experience that flood damage is an extreme event.

Table 3 summarizes the descriptive statistics of the simulated flood damage for the three selected dyke ring areas, and the AVV and VNK flood damage estimates. The coefficient of variability[10] (also called R^2) is about 68 per cent, which suggests that the simulated flood damage has a high variance. Since the corresponding skewness[11] for all three curves is less than 2 and positive, the tail on the right side is longer compared with the left side, and the bulk of the loss values lie to the left of the mean. Furthermore, the losses on both sides are truncated with the maximum and the minimum damage. This is consistent with the fact that flood damage cannot be infinitely large and justifies the data truncation. The kurtosis[12], which indicates the peakedness of a density function, is about 8, which is higher than the kurtosis of 3 of the standard normal density. The positive kurtosis shows that the flood damage amounts are peaked and not flat compared with the standard normal density.

Flood insurance premiums for all dyke ring areas

Table 4 shows the number of houses (column 2) per dyke ring area; the Empirical, AE and AVV estimates of flood insurance premiums (respectively, in columns 3, 4 and 5) and their ratios (columns 6 and 7). The premiums that have been estimated with the Empirical and AE methods (see Sect. 2.4) are compared with the premiums that Aerts and Botzen (2011) have estimated for all dyke-rings. This comparison is of interest, since Aerts and Botzen (2011) have estimated the flood insurance premiums using only the AVV data of a single estimate of the flood probability and potential damage per dyke ring area (premium = probability*damage), while the Empirical and AE estimates are based on the mean damage that are estimated from the complete probability density of flood damage that has been derived with BI, and account for the insurer's risk aversion to the catastrophe risk (see Sect. 2.4). The data are presented in descending order with respect to the ratio of the AE and AVV premiums.

The annual Empirical premium estimates (column 3) take the standard deviation of damage and the insurer's risk aversion rate for catastrophe risk into consideration by means of a surcharge on the expected flood risk that has been derived from actual insurance markets (Kunreuther et al., 2011). The Empirical premiums are generally close to 70 per cent of the AVV premiums (column 6), which indicates that the Empirical method results in a scaling of the AVV premiums. The empirical premiums are lower, even though these include a surcharge for the rate of risk aversion which is not accounted for in the AVV premium estimate by Aerts and Botzen (2011). The mean damages per dyke ring used as input for the AE premiums are significantly lower compared with the AVV mean damages (see Table A1 in Appendix A), while the AE premiums are higher for some dyke ring areas. This can be explained by the premium surcharge of the risk aversion rate which depends on the risk variance. The AE

[10] The R^2 is a normalized measure of the dispersion of a probability distribution, which is defined as the ratio of the standard deviation to the statistical mean.

[11] The skewness measures the asymmetry of a distribution or sample data relative to the standard normal distribution, which has a skewness of zero, and any data that has an asymmetric distribution has a skewness that differs from zero.

[12] The kurtosis measures whether the data are peaked or flat with respect to the normal distribution. The kurtosis of the standard normal distribution is 3 (in the case of "excess kurtosis", it is 0); a positive kurtosis indicates a peaked distribution; and a negative kurtosis implies a flat distribution.

Table 2. Main flood risk statistics for all 53 dyke ring areas (in million euros).

Dyke ring area number	Minimum damage (x_l)	Maximum damage (x_u)	Mean damage E $[X]$	SD of simulated damage (σ)	Parameter of interest ($\hat{\theta}$)
1	25	227	58	40	1.5668
2	85	757	195	133	1.5649
3	56	505	129	88	1.5716
4	6	50	13	9	1.5626
5	650	5806	1488	1012	1.5701
6	169	1515	389	265	1.5679
7	593	5301	1360	927	1.5675
8	1977	17 670	4533	3089	1.5704
9	749	6689	1720	1174	1.5669
10	424	3786	971	662	1.5700
11	339	3029	778	531	1.56454
12	876	7825	2005	1366	1.56516
13	1017	9087	2332	1594	1.56477
14	5254	46 951	12 060	8216	1.56801
15	1441	12 874	3300	2251	1.5671
16	6017	53 766	13 802	9420	1.56986
17	2853	25 495	6537	4451	1.56912
18	141	1262	323	220	1.56481
19	395	3534	907	616	1.57236
20	2542	22 718	5827	3980	1.56801
21	904	8078	2071	1413	1.57037
22	2542	22 718	5837	3980	1.56599
23	17	151	39	26	1.56681
24	678	6058	1557	1060	1.56456
25	593	5301	1359	925	1.56129
26	706	6311	1620	1107	1.57408
27	367	3281	843	574	1.56565
28	113	1010	259	177	1.56372
29	2260	20 194	5179	3532	1.56357
30	1497	13 378	3437	2342	1.56739
31	678	6058	1552	1058	1.56715
32	254	2272	584	398	1.57184
33	4	38	10	7	1.57096
34	1582	14 136	3637	2480	1.56833
35	989	8835	2266	1547	1.56906
36	1073	9592	2460	1676	1.56347
37	1	8	2	1	1.56903
38	791	7068	1809	1229	1.56798
39	8	76	19	13	1.56445
40	11	101	26	18	1.56402
41	1469	13 126	3372	2299	1.57265
42	282	2524	649	442	1.56675
43	3898	34 834	8956	6116	1.56662
44	1554	13 883	3570	2434	1.56855
45	1525	13 631	3497	2382	1.57249
46	28	252	65	44	1.56787
47	198	1767	454	309	1.56719
48	1384	12 369	3179	2171	1.56773
49	113	1010	259	177	1.56819
50	508	4544	1167	796	1.56561
51	85	757	194	132	1.56641
52	593	5301	1362	931	1.56693
53	1525	13 631	3502	2386	1.56909

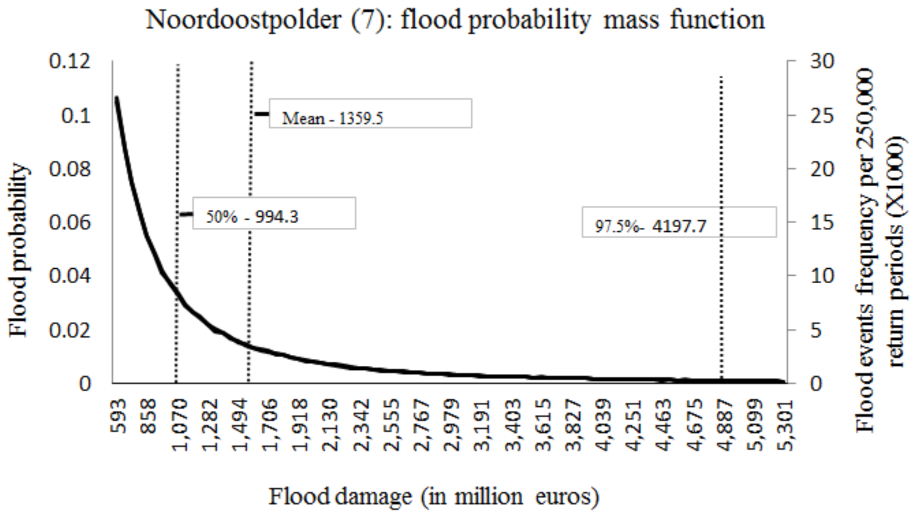

Fig. 3. Simulated flood damage density for the dyke ring Noordoostpolder (7).

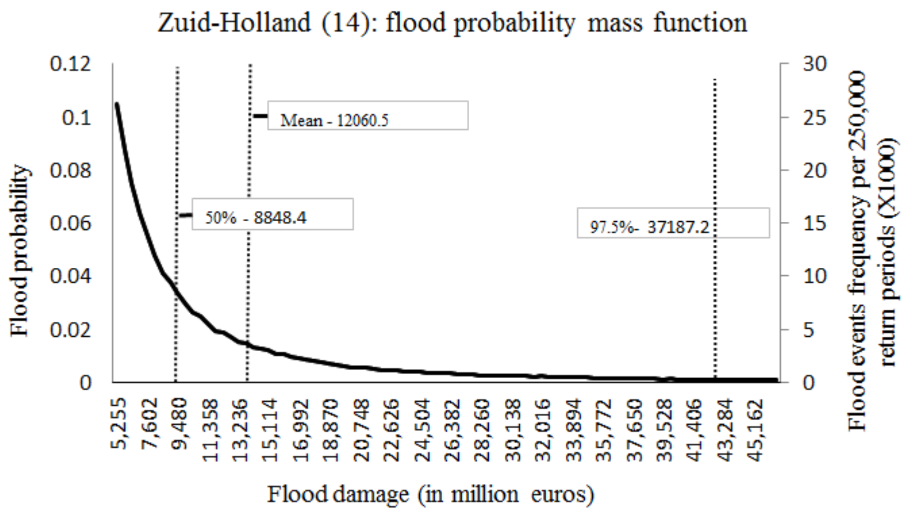

Fig. 4. Simulated flood damage density for the dyke ring Zuid Holland (14).

premiums (column 4) include a surcharge on the expected flood risk that is based on the full loss variance of the flood damage density (instead of the SD in the Empirical method), which results in substantial differences compared with the AVV estimates for some dyke ring areas. The differences between the AE and AVV premiums are largest for the dyke ring areas with a high expected damage and correspondingly high variance (e.g. dyke-rings 14, 16, and 43).

4 Discussion

The estimates of flood insurance premiums in Table 4 will be discussed with respect to three main aspects: the main differences between the premiums; how the BI method contributes to these findings; and the main implications of the results for insurers.

4.1 Main differences between the flood insurance premiums

The estimated flood damage densities per dyke ring area lie at the core of the estimations of the flood insurance premiums. Overall, the average flood damage per dyke ring, obtained through Bayesian statistical modelling, is lower if compared with the expected flood damage estimates obtained in the AVV and VNK projects (Wouters, 2005; Aerts et al., 2008). This can be attributed to using the full probability distribution from our BI approach as compared with, for example, AVV-based premiums that only use a single probability and

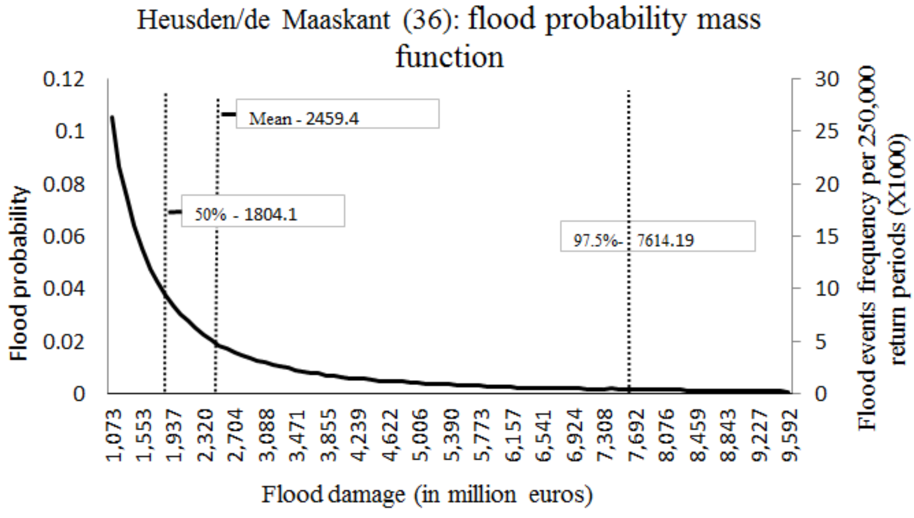

Fig. 5. Simulated flood damage density for the dyke ring Land van Heusden/de Maaskant (36).

Table 3. Descriptive statistics (damage is in million euros) of the simulated flood damage densities for dyke ring areas Noordoostpolder (7), Zuid Holland (14) and Heusden/de Maaskant (36).

Statistics	Noordoostpolder (7)	Zuid Holland (14)	Heusden/de Maaskant
Skewness	1.94	1.94	1.94
Kurtosis	6.57	6.59	6.6
Coeff. of variability	0.6819	0.6812	0.6815
Hyper-parameter shape (a)	0.521	0.505	0.517
Hyper-parameter scale (b)	1	1	1
Posterior-parameter shape (α)	1.53	1.524	1.529
Posterior-parameter scale (β)	1	1	1
Flood damage simulation shape $\hat{\theta}$	1.5675	1.56801	1.56347
Flood damage simulation scale parameter 1(x_l)	593.23	5.254.29	1.073.45
Flood damage simulation scale parameter 2 (x_u)	5301	46951	9592
50 % data percentile	994.27	8.848.48	1.804.63
97.5 % data percentile	4.197.32	37 187.01	7.614.53
VNK damage mean (model data)	2.000.0	18 500	2.800
AV-damage mean (prior information)	2.665	23 600	4,822

an extreme flood scenario with high flood damage as a basis. Furthermore, other factors, such as the choice of the density functions of the prior and likelihood information modelling, parameter uncertainty, and the Monte Carlo simulations used to fit the loss-probability curves, might also partially contribute to the difference in results of flood damage estimates.

Along with the differences in flood damage estimates, there are also significant differences in the premiums that are estimated using the AE and Empirical methods. For instance, the Empirical method emphasizes the insurer's risk aversion attitude to catastrophe risk by adding a surcharge to the expected flood risk in the premium estimate based on the standard deviation, while this is not applied in the AVV-based premiums (column 5 of Table 4). The Empirical premiums are approximately 70 per cent of the AVV premiums, which appears to be approximately constant for all dyke ring areas. This implies that the impact of using the loss standard deviation as a surcharge, on top of the risk-based premiums, is not very large if the Empirical method is used. In contrast, the AE method for calculating premiums adds a risk-averse surcharge to the expected flood risk that is based on the loss variance, which results in much higher premiums (up to 178 per cent) for some dyke ring areas compared with the Empirical and the AVV methods. Hence, these large differences, which particularly occur in those dyke ring areas with a large amount of expected damage (the first 8 dyke-rings in Table 4), can be explained by the corresponding large variances of flood damage. Such a large surcharge does not occur when the standard deviation is used for modelling insurer's risk aversion as the Empirical method does. The surcharge

Table 4. Results of annual flood insurance premiums per homeowner per dyke ring, according to the Empirical method, the Actuarial Equivalence (AE) method, and the AVV method.

Dyke ring	Number	Annual premium (in euros)			Premium ratio	
Nr.	of houses	Empirical	AE	AVV	Empirical/AVV	AE/AVV
16	82 340	87	219	123	70 %	178 %
14	1 659 248	1	2	1	75 %	174 %
43	120 526	61	122	87	71 %	140 %
17	165 235	10	17	15	68 %	116 %
20	62 823	24	39	34	71 %	115 %
22	47 243	64	104	91	70 %	114 %
29	49 060	27	42	39	70 %	108 %
8	99 069	12	17	17	70 %	103 %
44	292 938	10	14	14	72 %	99 %
41	109 400	25	34	36	71 %	95 %
45	103 282	28	38	40	70 %	95 %
53	86 300	34	46	48	70 %	95 %
30	29 532	30	41	43	70 %	95 %
48	59 881	44	58	62	71 %	94 %
15	79 164	22	29	31	70 %	93 %
36	165 555	12	15	17	72 %	90 %
35	37 524	31	38	44	71 %	87 %
21	32 152	33	40	47	71 %	85 %
12	8274	63	75	89	70 %	84 %
38	16 781	89	104	127	70 %	82 %
9	33 556	42	49	60	71 %	82 %
24	18 287	44	50	62	71 %	81 %
7	22 234	16	18	22	72 %	80 %
32	48 501	3	3	4	78 %	80 %
26	14 655	29	33	41	70 %	80 %
31	7087	57	65	81	70 %	80 %
5	5331	72	82	103	70 %	79 %
52	42 040	27	30	38	71 %	79 %
25	18 064	19	22	28	69 %	78 %
50	18 320	53	58	75	70 %	77 %
19	5696	16	18	23	72 %	76 %
10	11 128	45	49	64	70 %	76 %
27	9060	24	26	34	71 %	75 %
11	18 610	22	23	31	70 %	74 %
47	37 179	10	10	14	72 %	73 %
42	5611	96	99	136	70 %	73 %
13	412 013	1	1	1	59 %	72 %
28	3353	20	20	28	71 %	71 %
18	2054	16	16	23	71 %	71 %
33	26	100	96	137	73 %	70 %
37	12	137	133	190	72 %	70 %
51	4532	35	35	50	71 %	70 %
2	1345	75	74	106	71 %	70 %
49	7836	27	27	39	70 %	70 %
4	214	31	30	44	71 %	69 %
3	801	83	82	119	70 %	69 %
23	115	175	170	248	71 %	69 %
40	458	117	114	166	71 %	69 %
1	494	61	59	87	70 %	68 %
46	3227	17	16	24	69 %	68 %
39	169	93	90	135	69 %	67 %
34	160 741	12	16	50	23 %	32 %
6	468 014	0	0	1	21 %	22 %

on the AE premium is smaller for the last 45 dyke ring areas, shown in Table 4. This is caused by the higher weight that the AE method places on the flood damage variance, which is smaller for the last 45 dyke ring areas. In agreement with Friedman (1974), the AE premium estimates confirm that insurers are considerably risk-averse to damage with a high loss variance, and they see this type of risk either as uninsurable or as a gamble that needs a significantly high expected return.

It should be noted that Kunreuther et al. (2011) calculated the risk aversion surcharge, used in the Empirical method, based on historical surcharge information that US insurers have charged for providing coverage against hurricane damage. However, this surcharge may not completely reflect risk aversion to extreme flood events in the Netherlands, which can have a catastrophic character and result in very high losses which could ruin the insurer. A higher surcharge for risk aversion to the high amounts at stake may be applied for such events. This, for example, is done in the AE method, but we have no empirical data specific for the Netherlands on which this surcharge could be based, since empirical estimates of insurer's risk aversion to providing coverage for flood risk in the Netherlands are not available.

4.2 Bayesian Inference (BI) method

For several reasons, it can be argued that the BI method applied in this study is more suitable for estimating flood insurance premiums in the Netherlands compared with the methods that only use a single estimate of the flood probability and potential damage (the AVV and the VNK projects). First, BI provides statistical estimates of flood risk that take its stochastic and extreme nature into account by deriving the complete probability density of flood damage, and modelling the tail of this density (Bayarri and Berger, 2004). Second, BI is a suitable method for representing probabilistic relationships between different sources of information, such as the AVV and VNK data (Heckerman, 2008). However, in some BI applications there may be concerns about the reliability of the prior and likelihood data sources, and how this influences the results (Raftery et al., 1997; Malakoff, 1999; Hájek, 2007; Gelman, 2008; Chaudhuri and Ghosh, 2011). For example, the inclusion of prior information does not always lead to better results, especially when it is based on subjective beliefs. To overcome this issue, only objective data are used in this study that share similar statistical features as input. Third, the damage estimations in this study are based on extensive data simulations, which enable more detailed statistical information to be provided for the flood damage assessments. Simulation is the only way to incorporate probabilistic scenarios in the estimation of risk of low-probability floods, where there is little historical information available on such floods (Juneja et al., 2006). In contrast, VNK and AVV, as well as other flood damage studies, do not consider the probabilistic nature of risk and cannot provide any statistical inferences for damage beyond a certain level. Fourth, the BI procedure applied in this study allowed for the derivation of the full probability density of flood damage, its mean and standard deviation, which are all important inputs for the calculation of flood insurance premiums. Even though simulation provides useful information that is needed to estimate flood insurance premiums, it also has its limitations. For example, a simulation attempts to mimic the damage of a flood event based on known facts and assumptions by means of a conceptual computational environment, which results in uncertainties (Robert and Casella, 2011). Large uncertainties associated with rare events are, in our application, somewhat narrowed by truncating the loss data at the best estimates of minimum and maximum flood losses per dyke ring area.

4.3 Implications for insurers

From the findings of this study it becomes clear that insurance for flood risk is a complex product to price because of the extremely low flood frequency that entails large uncertainties. This study is the first in-depth study of the pricing of flood insurance in the Netherlands that uses the full probability density of flood damage in all 53 dyke-areas. Therefore, it provides a useful basis for insurers who are considering introducing flood insurance in the Netherlands. Our premium estimates show that flood insurance premiums can be considerably above the expected value of the flood loss in some dyke ring areas because of the risk aversion of the insurer for the catastrophic nature of flood risk. Because the risks located on the right-tail of the damage density are much more expensive to insure compared with the risks of lower damage on the left side, insurers may be reluctant to provide insurance for extreme flood losses in some high-risk areas unless they can charge sufficiently high premiums. Nevertheless, our estimated flood insurance premiums are lower than household willingness-to-pay (WTP) for flood insurance in most dyke ring areas. Botzen and van den Bergh (2012) estimate that average individual WTP for flood insurance in the current situation of flood risk is about €250 per year, which is higher than our estimated flood insurance premiums in most dyke ring areas (Botzen and van den Bergh, 2012). However, actual flood insurance premiums are in practice likely to be higher than the premiums provided in this paper due to administrative costs and a profit margin for insurers, which are not included in our estimate.

Our study follows the proposal by Kunreuther et al. (2009) to determine flood insurance premiums on the basis of estimate of actual flood risks. Nevertheless, we realize that in practice flood insurance premiums may not be fully differentiated with respect to actual flood risk, for example, because bundled coverage is provided or because it entails costs for insurers to determine and charge different premiums for every specific policy. Our analysis provides insights into the level of flood insurance premiums as if they were risk-based and assesses flood risks and flood insurance premiums for the Netherlands on a dyke ring level, which provides a relatively simple basis for premium differentiation. Overall, the estimated flood insurance premiums show large differences between dyke ring areas in the Netherlands, which is mainly due to the difference in dyke failure probabilities between these areas. This suggests that premiums should be differentiated at least on a dyke ring level if an insurance system with risk-based premiums were to be introduced. Insurance costs would differ considerably between the different low-lying areas in the Netherlands if flood insurance premiums were to reflect risk.

5 Conclusions and recommendations

This study has applied Bayesian Inference to assess the stochastic nature of flood risk and provide estimates of the probability density of flood damage for all 53 dyke ring areas in the Netherlands. Subsequently, these probability densities of flood damage have been used to estimate flood insurance premiums for these areas. While previous studies have derived a single estimate of the flood probability and expected flood damage for the low-lying areas in the Netherlands, our study has estimated the full probability density of flood damage, which allows for a more accurate estimation of flood insurance premiums. In particular, the premiums estimated in this study account for the insurer's risk aversion to the extreme nature of flood risk. This study is of practical relevance for insurers who are considering introducing flood insurance in the Netherlands.

The methodological process followed in this paper to estimate premiums for damage emerging from rare events, such as catastrophic and man-made disasters, appears to be of great relevance, as it is able to cope with the lack of empirical evidence on the corresponding expected damage. Using a practical example, this paper showed that the widespread uncertainties about flooding should be included in premium calculations by taking into account the relevant risk indicators, such as risk variance and insurer's aversion against catastrophe risk insurance. Furthermore, as we notice from the premium results, the choice of a particular method appears to make a significant difference for their levels. Therefore, it is important that the method used to estimate insurance premiums should correctly represent the real-world problem, and thus reflect the true nature of the corresponding risk. Data pre-processing with respect to consistency, reliability, and completeness is a vital part in the risk-estimation process because, regardless of the type of method used, the soundness of results can, for a large part, be assigned to input. Usually, it is assumed that the consequences of catastrophic events are unlimited. However, in practice the damage is usually limited, and lies between two extremes. In such cases, the range of possible outcomes of rare events can be somewhat narrowed if unnecessary and unrealistic information is excluded in models by truncating loss data at below a predetermined threshold, which results in more realistic premiums.

Further in-depth research is necessary to explore and analyse different aspects of Bayesian techniques tailored for rare events. This study has provided insights into uncertainty of estimated flood damage, while it should be acknowledged that another important source of uncertainty is the flood frequency. In this respect, future research could focus on obtaining better insights into uncertainties of the real probability of dyke failure. Furthermore, as it allows the integration of expert judgment and other third party information, it would be advisable to refine the prior assessment process carefully by integrating subjective information that may be of great value. Controversy arises because prior information is generally assumed to be subjective, and can have a significant impact on the final results. However, this can partly be compensated with a cross-validation of the information, as long as it is properly carried out. The risk aversion rate used for the premium calculation should reflect the actual risk in the dyke ring areas rather than those estimated for different areas. Therefore, more research will be necessary on insurer's risk aversion to catastrophe risk in the context of flood risk in the Netherlands.

Appendix A

AVV and VNK input data

Table A1 provides the flood risk estimates from the VNK and AVV projects per dyke ring area which have been used as input data in the Bayesian model. Column 2 provides the flood return period per dyke ring area and columns 3 and 4 show, respectively, the expected flood damage that corresponds to these return periods.

Appendix B

Data truncation: the derivation of minimum and maximum damage

AVV and VNK provide expected damage estimate for 53 dyke-rings areas, based on the current probability of dyke overtopping (the exceedance probability), and no minimum and maximum amount, except for the three dyke ring areas, 7, 14 and 36, by VNK. Since it is assumed that flood damage is truncated on both sides, the minimum and maximum damage amount must be derived from VNK information. Therefore, we first calculated the weighted upscaling and downscaling factors using the available information for dyke ring areas 7, 14 and 36, which were used to derive the same information for the other dyke ring areas on a proportional basis.

$$\max(\mathrm{AVV}) = \mathrm{UF} \cdot E(\text{exceedance damage}) \cdots \quad \cdots \quad \cdots \quad \cdots, \tag{B1}$$

$$\min(\mathrm{AVV}) = \mathrm{DF} \cdot E(\text{exceedance damage}) \cdots \quad \cdots \quad \cdots \quad \cdots, \tag{B2}$$

Table A1. The flood risk estimates from the VNK and AVV projects that were used as input data for the Bayesian model.

Dyke ring area number	Dyke failure Return period AVV	VNK expected damage in millions of €	AVV expected damage in millions of €
1	2000	114	114
2	2000	500	381
3	2000	254	254
4	2000	25	25
5	4000	2000	2918
6	4000	400	761
7	4000	2000	2665
8	4000	7500	8882
9	1250	5300	3362
10	2000	1200	1903
11	2000	1200	1523
12	4000	1000	3933
13	10 000	2900	4568
14	10 000	18 500	23 600
15	2000	5000	6471
16	2000	9500	27026
17	4000	8000	12 815
18	10 000	634	634
19	10 000	1776	1776
20	4000	9000	11 420
21	2000	4000	4060
22	2000	9000	11 420
23	2000	200	76
24	2000	2400	3045
25	4000	1900	2665
26	4000	2500	3172
27	4000	1300	1649
28	4000	400	508
29	4000	8000	10 151
30	4000	5300	6725
31	4000	1700	3045
32	4000	700	1142
33	4000	19	19
34	2000	3000	7105
35	2000	2000	4441
36	1250	2800	4822
37	1250	4	4
38	1250	1300	3553
39	1250	38	38
40	500	51	51
41	1250	3195	6598
42	1250	850	1269
43	1250	9000	17 510
44	1250	5440	6979
45	1250	1825	6852
46	1250	127	127
47	1250	780	888
48	1250	3360	6217
49	1250	380	508
50	1250	1820	2284
51	1250	275	381
52	1250	1595	2665
53	1250	4400	6852

Source: TAW (2000); Wouters (2005); Aerts et al. (2008).

where

$$\text{Upscaling factor (UF)} = \left[\frac{\left(\frac{\mathrm{Max}(VnK)}{E(\mathrm{VnK})}\right)_{\text{dyke ring 7}} + \left(\frac{\mathrm{Max}(VnK)}{E(\mathrm{VnK})}\right)_{\text{dyke ring 14}} + \left(\frac{\mathrm{Max}(VnK)}{E(\mathrm{VnK})}\right)_{\text{dyke ring 36}}}{3}\right]\cdots, \tag{B3}$$

$$\text{Downscaling factor (DF)} = \left[\frac{\left(\frac{\mathrm{Min}(VnK)}{E(\mathrm{VnK})}\right)_{\text{dyke ring 7}} + \left(\frac{\mathrm{Min}(\mathrm{VnK})}{E(VnK)}\right)_{\text{dyke ring 14}} + \left(\frac{\mathrm{Min}(VnK)}{E(\mathrm{VnK})}\right)_{\text{dyke ring 36}}}{3}\right]\cdots. \tag{B4}$$

$$\min(X) = m = \mathrm{DF}\cdot \mathrm{AVV}_i \cdots \cdots \cdots \cdots, \tag{B5}$$

$$\max(X) = m = \mathrm{DF}\cdot \mathrm{AVV}_i \cdots \cdots \cdots \cdots, \tag{B6}$$

Appendix C

Estimation of hyper-parameters and sufficient statistics

The estimation of hyper-parameters needs to be conducted in two steps because the estimation of hyper-parameters a and b of the prior distribution is based on information about θ, while θ itself needs to be estimated from the flood damage data x (from AVV). First, we estimate the weighted average of θ, and create several data points using flood damage information from AVV as follows: $\theta = \left\{\frac{\sum_{i=1}^{n} x_{i\cdot w_i}?}{\sigma_{x_{i\cdot w_i}}}\right\}^2$, where $\sum_{i=1}^{n} x_{i\cdot w_i}$ is the weighted average of our prior beliefs about flood damage estimated from AVV flood damage data, with $\sum_{i=1}^{n} w_i = 1,\ i = 1, 2, 3$, which are the three dyke ring areas for which detailed damage information (i.e. expected, minimum, and maximum amounts) is available; and $\sigma_{x_{i\cdot w_i}}$ is the standard deviation of damage estimated from the minimum, maximum, and expected damage (see Appendix B).

Second, based on the information about θ obtained in the first step, the necessary hyper-parameters can be estimated as $a = \left\{\frac{\sum_{i=1}^{n} \theta_{i\cdot w_i}?}{\sigma_\theta}\right\}^2$ and $b = 1$.

Appendix D

Deriving the posterior distribution from the Pareto likelihood and Gamma prior

The posterior density is derived from the prior and likelihood functions (Arnold, 1998):

$$p(\theta\,|\,x) \propto L(\theta\,|\,X)\cdot p(\theta)\ldots \cdots \cdots \cdots, \tag{D1}$$

where $p(\theta)$ is a Gamma distribution with two parameters α and β (see Sect. 2.2 about the prior distribution).

The likelihood function for damage observation is given by

$$L(\theta|x) = \theta^n x_{\mathrm{m}}^{\theta n} \prod\nolimits_{i=1}^{n} x^{-(\theta+1)}$$
$$\text{for } x \in \left[x_{m;+\infty}\right]\ldots \cdots \cdots \cdots. \tag{D2}$$

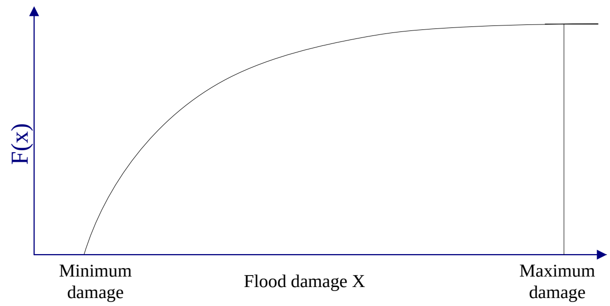

Fig. E1. A conceptual sketch of cumulative bounded distribution function.

Equation (6) can be obtained by substituting Eqs. (5) and (D2) in Eq. (D1):

$$p(\theta\,|\,x) \propto \theta^n x_{\mathrm{m}}^{\theta n} \prod\nolimits_{i=1}^{n} x_i^{-(\theta+1)}$$
$$\times(\theta)^{a-1}\,\frac{e^{\left(-\frac{\theta}{b}\right)}}{b^a\,\Gamma(a)}\cdots \cdots \cdots \cdots, \tag{D3}$$

$$\approx p(\theta\,|\,x) \propto \frac{(\theta)^{\alpha-1}\,e^{-\left(\frac{\theta}{\beta}\right)}}{\beta^\alpha\,\Gamma(\alpha)}\cdots \cdots \cdots \cdots, \tag{D4}$$

where $\alpha = a + n$, and $\beta = \frac{1}{\frac{1}{b} + \ln\left(\prod_{i=1}^{n} x_i\right) - n\cdot ln(x_{\mathrm{m}})}$, and $\beta < \frac{1}{\left(n\cdot ln(x_{\mathrm{m}}) - \ln\left(\prod_{i=1}^{n} x_i\right)\right)}$.

Appendix E

The bounded Pareto distribution

Equation (E1) gives the formula of a bounded Pareto distribution (Weisz and Brown, 2001), and Fig. E1 provides a conceptual sketch of such a distribution, which is bounded with a minimum and maximum amount of flood damage.

$$f(x) = \frac{\hat{\theta}\cdot x_1^{\hat{\theta}}\cdot x^{-(\hat{\theta}+1)}}{1 - \left(\frac{x_1}{x_{\mathrm{u}}}\right)^{\hat{\theta}}}\cdots \cdots \cdots, \tag{E1}$$

Flood damage can be simulated by drawing random numbers from Eq. (E1). Equation (8) can be written in a shorter form as follows:

$$\text{Flood damage}(x) = f^{-1}\left(x;\hat{\theta},x_{\mathrm{l}},x_{\mathrm{u}}\right)\cdots \cdots$$
$$\text{with } x_{\mathrm{l}} \le x_i \le x_{\mathrm{u}} \tag{E2}$$

where, $\hat{\theta}$, x_{l}, x_{u} are respectively, the shape parameter (obtained from Eq. 7) and two scale parameters (the lower and upper boundaries). Equation (E1), which is used to simulate flood damage for each of the 53 dyke ring areas using the bounded Pareto distribution, can be rewritten as

$$x_i = \left(-\frac{U \cdot x_{\mathrm{u}}^{\hat{\theta}} - U \cdot x_{\mathrm{l}}^{\hat{\theta}} - x_{\mathrm{u}}^{\hat{\theta}}}{x_{\mathrm{u}}^{\hat{\theta}} \cdot x_{\mathrm{l}}^{\hat{\theta}}}\right)^{-\frac{1}{\hat{\theta}}} \ldots\ldots\ldots, \qquad \text{(E3)}$$

where U is a uniformly distributed random number between 0 and 1.

Appendix F

Derivation of the AE premium

Given a utility function $u(x)$, an insured value w and damage X, the maximum premium charged by an insurer (π^+) can be approximated as has been described in Kaas et al. (2004). The maximum insurance premium π^+ can be derived by equating the expected utility in the absence of insurance with the expected utility with insurance:

$$E\left[u\left(w - X\right)\right] = \left[u\left(w - \pi^+\right)\right]\ldots, \qquad \text{(F1)}$$

σ^2 and $[X]$ are, respectively, the risk variance and mean of damage X. Using the first term of Taylor series expansion of $u(x)$ in $w - E[X]$, we obtain

a. Utility in terms of the premium:

$$u\left(w - \pi^+\right) \approx u\left(w - E\left[X\right]\right) + \left(E\left[X\right] - \pi^+\right) u' \left(w - E\left[X\right]\right)\ldots\ldots\ldots, \qquad \text{(F2)}$$

b. Utility in terms of the damage:

$$u\left(w - X\right) \approx u\left(w - E\left[X\right]\right) + \left(E\left[X\right] - X\right) u' \left(w - E\left[X\right]\right) + \frac{1}{2}\left(E\left[X\right] - X\right)^2 u'' \left(w - E\left[X\right]\right) \qquad \text{(F3)}$$

Taking expectation of both sides of Eqs. (F2) and (F3) and substituting this in (F1) gives

$$\pi^+ \approx E\left[X\right] - \frac{1}{2}\sigma^2 \frac{u''\left(w - E\left[X\right]\right)}{u'\left(w - E\left[X\right]\right)}\ldots\ldots\ldots \qquad \text{(F4)}$$

By defining the risk-version coefficient $r(w)$ as $r = -\frac{u''(w)}{u'(w)}$, (F4) can be rewritten as

$$\pi^+ \approx E\left[X\right] + \frac{1}{2}\sigma^2 r\left(w - E\left[X\right]\right)\ldots\ldots\ldots \qquad \text{(F5)}$$

The expression (F5) is the same as Eq. (9) if $\mathrm{E}[X]$ is substituted with $\mathrm{E}[\mathrm{AE}]_{\mathrm{annual}}$.

Acknowledgements. The authors wish to thank Theo Dijkstra, Erwann Michel-Kerjan and Howard Kunreuther for their helpful comments on earlier drafts of this article. Part of this research has been co-funded by The Netherlands Organisation for Scientific Research (NWO).

Edited by: H. Kreibich
Reviewed by: four anonymous referees

References

Aerts, J., Sprong, T., and Bannink, B. A.: Aandacht voor Veiligheid, Leven met Water, Klimaat voor Ruimte, DG Water, 3–196, 2008.

Aerts, J. C. J. H. and Botzen, W. J. W.: Climate change impacts on pricing long-term flood insurance: A comprehensive study for the Netherlands, Glob. Environ. Change, 21, 1045–1060, 2011.

Arnold, B. C. and Press, S. J.: Bayesian inference for pareto populations, J. Econometrics, 21, 287–306, 1983.

Arnold, B. C., Castillo, E., and Sarabia, J. M.: Bayesian analysis for classical distributions using conditionally specified priors, The Indian J. Statis., 60, 228–245, 1998.

Bayarri, M. J. and Berger, J. O.: The Interplay of Bayesian and Frequentist Analysis, Statist. Sci., 19, 58–80, citeulike-article-id:5183336, 2004.

Behrens, C. N., Lopes, H. F., and Gamerman, D.: Bayesian analysis of extreme events with threshold estimation, Statis. Model., 4, 227–244, doi:10.1191/1471082X04st075oa, 2004.

Bermudez, P. D. and Kotz, S.: Parameter estimation of the generalized Pareto distribution-Part II, J. Statist. Plann. Inference, 140, 1374–1397, 2010.

Botzen, W. J. W. and van den Bergh, J. C. J. M.: Insurance against climate change and flooding in the Netherlands: Present, future, and comparison with other countries, Risk Anal., 28, 413–426, 2008.

Botzen, W. W. J. and van den Bergh, J. C. J. M.: Monetary valuation of insurance against flood risk under climate change, Int. Economic Rev., 1–38, 2012.

Bouwer, L. M., Bubeck, P., Wagtendonk, A. J., and Aerts, J. C. J. H.: Inundation scenarios for flood damage evaluation in polder areas, Nat. Hazards Earth Syst. Sci., 9, 1995–2007, doi:10.5194/nhess-9-1995-2009, 2009.

Bouwer, L. M., Bubeck, P., and Aerts, J. C. J. H.: Changes in future flood risk due to climate and development in a Dutch polder area, Glob. Environ. Change, 20, 463–471, 2010.

Chaudhuri, S. and Ghosh, M.: Empirical likelihood for small area estimation, Biometrika, 98, 473–480, doi:10.1093/biomet/asr004, 2011.

Coles, S. G. and Powell, E. A.: Bayesian methods in extreme value modelling: A review and new developments, Int. Statist. Rev., 64, 119–136, 1996.

Cooley, D., Nychka, D., and Naveau, P.: Bayesian spatial modeling of extreme precipitation return levels, J. Am. Statis. Assoc., 102, 824–840, 2007.

Davis, R. E.: Decision Policy Optimization Via Certain Equivalent Functions for Exponential Utility, in: Advances in Mathematical Programming and Financial Planning, edited by: Lawrence, K. D., Reeves, G. R., and John, B. G., Elsevier Science, 89–111, 2001.

de Moel, H. and Aerts, J.: Effect of uncertainty in land use, damage models and inundation depth on flood damage estimates, Nat. Hazards, 58, 407–425, doi:410.1007/s11069-11010-19675-11066, 2011.

Duncan, J. and Myers, R. J.: Crop insurance under catastrophic risk, Am. J. Agric. Econom., 82, 842–855, doi:10.1111/0002-9092.00085, 2000.

Fink, D.: A compendium of conjugate priors, Citeseer, 46, available at: http://www.people.cornell.edu/pages/df36/CONJINTRnew%20TEX.pdf, 1997.

Friedman, B.: Risk Aversion and the Consumer Choice of Health Insurance Option, in: The Review of Economics and Statistics, The MIT Press, 56, 209–214, 1974.

Froot, K. A.: The market for catastrophe risk: a clinical examination, J. Financ. Econ., 60, 529–571, 2001.

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B.: Bayesian Data Analysis, Chapman & Hall/CRC, 2004.

Gelman, A.: Objections to Bayesian statistics, Bayesian Analysis, 3, 445–450, 2008.

Grossi, P. and Kunreuther, H.: Catastrophe Modeling: A New Approach to Managing Risk, in: An Introduction to Catastrophe Models and Insurance, edited by: Grossi, P., and Kunreuther, H., 1 edn., Springer, New York, 23–42, 2005.

Hájek, A.: The reference class problem is your problem too, Synthese, 156, 563–585, doi:10.1007/s11229-006-9138-5, 2007.

Heckerman, D.: A Tutorial on Learning with Bayesian Networks in: Innovations in Bayesian Networks, edited by: Holmes, D. and Jain, L., Studies in Computational Intelligence, Springer Berlin/Heidelberg, 33–82, 2008.

Hsieh, P. H.: A Data-Analytic Method for Forecasting Next Record Catastrophe Loss, J. Risk Insur., 71, 309–322, 2004.

Jaffee, D. M. and Russell, T.: Catastrophe insurance, capital markets, and uninsurable risks, J. Risk Insur., 64, 205–230, 1997.

Jonkman, S. N., Gelder, P. H. A. J. M. V., and Vrijling, J. K.: An overview of quantitative risk measures for loss of life and economic damage, J. Hazard. Mater., 99, 1–30, 2003.

Jonkman, S. N., Kok, M., and Vrijling, J. K.: Flood risk assessment in the Netherlands: A case study for dike ring South Holland, Risk Anal., 28, 1357–1373, 2008.

Juneja, S., Shahabuddin, P., Shane, G. H., and Barry, L. N.: Rare-Event Simulation Techniques: An Introduction and Recent Advances, in: Handbooks in Operations Research and Management Science, edited by: Lenstra, J. K. and Nemhauser, G. L., Elsevier, 291–350, 2006.

Kaas, R., Goovaerts, M., Dhaene, J., and Denuit, M.: Modern actuarial risk theory, Kluwer Academic, New York, 2–10, 2004.

Kabat, P., van Vierssen, W., Veraart, J., Vellinga, P., and Aerts, J.: Climate proofing the Netherlands, Nature, 438, 283–284, 2005.

Kerman, J. and Gelman, A.: Manipulating and summarizing posterior simulations using random variable objects, 17, 235–244, 2007.

Klijn, F., Baan, P. J. A., De Bruijn, K. M., and Kwadijk, J.: Overstromingsrisico's in Nederland in een veranderend klimaat, WL—delft hydraulics, Delft, Netherlands, 1–166, 2007.

Kunreuther, H. C. and Michel-Kerjan, E. O.: Climate change, insurability of largescale disasters, and the emerging liability challenge, University of Pennsylvania Law Review, 155, 1795–1842, 2007.

Kunreuther, H. C., Michel-Kerjan, E. O., Doherty, N. A., Grace, M. F., Klein, R. W., and Pauly, M. V.: At War with the Weather: Managing Large-Scale Risks in a New Era of Catastrophes, First edn., The MIT Press, Cambridge, MA, 2009.

Kunreuther, H., Michel-kerjan, E., and Ranger, N.: Insuring Climate Catastrophes in Florida: An Analysis of Insurance Pricing and Capacity under Various Scenarios of Climate Change and Adaptation Measures, Center for Risk Management and Decision Processes, The Wharton School, University of Pennsylvania, USA, London, 2011-07, 1–24, 2011.

Malakoff, D.: Bayes Offers a 'New' Way to Make Sense of Numbers, Science, 286, 1460–1464, doi:10.1126/science.286.5444.1460, 1999.

Raftery, A. E., Madigan, D., and Hoeting, J. A.: Bayesian model averaging for linear regression models, J. Am. Statis. Assoc., 92, 179–191, 1997.

Robert, C. P. and Casella, G.: Monte Carlo statistical methods, Springer, New York, London, 83–102, 2011.

TAW: Van overschrijdingskans naar overstromingskans, Technische Adviescommissie voor de Waterkering (RWS-DWW), DelftP00.04a, 3–25, 2000.

Van der Most, H. and Wehrung, M.: Dealing with uncertainty in flood risk assessment of dike rings in the Netherlands, Nat. Hazards, 36, 191–206, 2005.

Vilar-Zanon, J. L. and Lozano-Colomer, C.: On pareto conjugate priors and their application to large claims reinsurance premium calculation, Astin Bulletin, 37, 405–428, 2007.

Vrijling, J. K.: Probabilistic design of water defense systems in The Netherlands, Reliability Eng. Syst. Safety, 74, 337–344, 2001.

Weisz, J. and Brown, M.: Supporting Low-Latency Broadcast in End System Multicast, Carnegie mellon University, 2001.

Wouters, K.: Veiligheid Nederland in kaart globale schadeberekening, Dienst Weg- en Waterbouwkunde, Rijkswaterstaat, Delft, available at: http://www.deltacommissie.com/doc/Aandacht%20voor%20veiligheid%20.pdf, 2005.