AGU100 ADVANCING EARTH AND SPACE SCIENCE

Water Resources Research

RESEARCH ARTICLE
10.1029/2019WR025068

Key Points:
- Hierarchical Bayesian model (HBM) captures spatiotemporal variability in flood damage processes better than established depth-damage functions
- Region- and event-specific characteristics that explain the variability in damage processes improve spatiotemporal transfer of the HBM
- The HBM is tested for transferability using empirical damage data from six events in the Elbe, Danube, Rhine and Oder catchments in Germany

Supporting Information:
- Supporting Information S1

Correspondence to:
N. Sairam,
nivedita@gfz-potsdam.de

Citation:
Sairam, N., Schröter, K., Rözer, V., Merz, B., & Kreibich, H. (2019). Hierarchical Bayesian approach for modeling spatiotemporal variability in flood damage processes. *Water Resources Research*, *55*, 8223–8237. https://doi.org/10.1029/2019WR025068

Received 4 MAR 2019
Accepted 9 SEP 2019
Published online 29 OCT 2019

Hierarchical Bayesian Approach for Modeling Spatiotemporal Variability in Flood Damage Processes

Nivedita Sairam[1,2], Kai Schröter[1], Viktor Rözer[3], Bruno Merz[1,4], and Heidi Kreibich[1]

[1]GFZ German Research Centre for Geosciences, Section Hydrology, Potsdam, Germany, [2]Geography Department, Humboldt University, Berlin, Germany, [3]Grantham Research Institute on Climate Change and the Environment, London School of Economics, London, UK, [4]Institute for Environmental Sciences and Geography, University of Potsdam, Potsdam, Germany

Abstract Flood damage processes are complex and vary between events and regions. State-of-the-art flood loss models are often developed on the basis of empirical damage data from specific case studies and do not perform well when spatially and temporally transferred. This is due to the fact that such localized models often cover only a small set of possible damage processes from one event and a region. On the other hand, a single generalized model covering multiple events and different regions ignores the variability in damage processes across regions and events due to variables that are not explicitly accounted for individual households. We implement a hierarchical Bayesian approach to parameterize widely used depth-damage functions resulting in a hierarchical (multilevel) Bayesian model (HBM) for flood loss estimation that accounts for spatiotemporal heterogeneity in damage processes. We test and prove the hypothesis that, in transfer scenarios, HBMs are superior compared to generalized and localized regression models. In order to improve loss predictions for regions and events for which no empirical damage data are available, we use variables pertaining to specific region- and event-characteristics representing commonly available expert knowledge as group-level predictors within the HBM.

1. Introduction

Implementation of efficient flood risk management requires accurate and reliable quantification of flood risk. Flood loss estimation models are crucial in determining monetary losses incurred due to floods (Bubeck & Kreibich, 2011; Merz et al., 2010). These models need to capture the damage processes due to flooding using the relationships between incurred loss and its impacting and resisting factors (Merz et al., 2013; Thieken et al., 2005). Most common flood loss models are depth-damage functions, which estimate the loss from the type or use of the element at risk (e.g., residential building) and the inundation depth (Figueiredo et al., 2018; Gerl et al., 2016). Gerl et al. (2016) categorized flood loss models based on the model development approach into synthetic/engineering models (e.g., Dottori et al., 2016; Klaus et al., 1994; Parker et al., 1987; Penning-Rowsell 1977; Smith, 1994) and empirical models (e.g., Carisi et al., 2018; Elmer et al., 2010; Kreibich et al., 2010; Nicholas et al., 2001; Thieken et al., 2008; Zhai et al., 2005).

Commonly, empirical flood loss models are developed using damage data from single events covering a small spatial extent (catchment/region; Chinh et al., 2017, Carisi et al., 2018). These models have the advantage that they are able to incorporate local- and event-specific differences either explicitly through additional predictors or implicitly through a specific stage damage function. However, research has shown that models trained from specific events do not perform well when transferred in space and/or time (Cammerer et al., 2013). The low skill of such localized models in transfer settings is a consequence of the spatiotemporal heterogeneity in the factors influencing building loss during different flood events and process types (Vogel et al., 2018). Local exposure and vulnerability are commonly affected by predominant building style, household income, regulations, and flood insurance practice (Jongman et al., 2012). Significant variability in hazard intensity, such as flood duration, flow velocity, contamination, and sediment load, is generally observed for different events. Between consecutive flood events, the level of adaptation and exposure can vary, resulting in temporal variability in damage processes (Kreibich, Botto, et al., 2017).

Flood intensity is influenced by duration of inundation, along with inundation depth (Rözer et al., 2019). Households experiencing longer inundation duration experience higher building damage (Thieken et al., 2005). Return period is an indicator of the extremity of the flood event in a given region. Return period is positively correlated to flooding intensity and negatively correlated to flood experience (Elmer et al., 2010). Households in regions experiencing frequent flooding have high flood experience resulting in increased awareness, preparedness, and widespread implementation of private precautionary measures, such as flood proofing buildings and sealing oil tanks (Bubeck et al., 2013). These characteristics strongly influence the damage processes in private households; however, it is quite challenging to collect data concerning these attributes at the object level (household). Hence, the development of generalized flood loss models suitable for various regions and events is not trivial. In order to overcome these challenges in the representation of damage processes, we propose a Hierarchical Bayesian model (HBM) for flood loss estimation using water depth at the household level as a predictor. This is a probabilistic model that provides uncertainty quantification and also explicitly accounts for spatiotemporal variability in the damage processes.

HBMs can be theoretically conceptualized and implemented to account for causal effects in processes (Feller & Gelman, 2015; Gelman, 2006; Kruschke & Vanpaemel, 2015; Levy 2012). Hence, these models have been widely used in various fields involving experimental observations or survey data. Sun et al. (2015) implemented hierarchical Bayesian clustering to identify spatiotemporal trends in precipitation extremes; Ahn et al. (2017) developed a HBM to forecast seasonal stream flows. Das et al. (2018) showed the potential of using a hierarchical modeling approach for modeling irrigation withdrawals over the United States, especially for data-sparse years. However, as per our knowledge, there are no studies that implemented a HBM for flood loss estimation.

The localized model considers that each region and event has distinct damage processes that are independent of the other regions and events. The generalized model assumes that all regions and events have the same damage processes (given the explanatory variables, i.e., flood loss predictors). The hierarchical (multilevel) approach aims to achieve a middle ground between completely generalized and localized regression models. It provides flexibility in defining a meaningful structure to flood loss models. In order to facilitate spatiotemporal transferability of flood loss models, the damage processes pertaining to different events and regions are modeled separately while also accounting for similar processes across regions and events. Bayesian probabilistic modeling is used for flood loss estimation because of its inherent ability to quantify uncertainty in the observations and include it in the posterior distributions of the predictions. A Bayesian approach combined with a hierarchical model structure provides estimates of uncertainty at the level of individual objects (household) and groups, that is, events and regions. An additional advantage of the hierarchical approach is the possibility to include information pertaining to different levels in the hierarchy as explanatory variables in the model structure. This allows us to use region- and event-related aggregated data or expert knowledge from secondary data sources such as government reports or media and news, pertaining to flood damage processes to parameterize the model with the intention to improve loss predictions during spatiotemporal transferability.

In this study, we use empirical flood loss data from six flood events in the Elbe, Danube, Rhine, and Oder catchments in Germany in order to test the following hypotheses:

1. Implementing a HBM for flood loss estimation captures spatiotemporal variability (regions and events) in the damage processes better and improves loss prediction, compared to the generalized and localized regression models.
2. Including group-level predictors with information representing specific region- and event-characteristics using expert knowledge improves flood loss prediction of the HBM.

The paper is organized as follows: The empirical data used in this study is described in section 2.1. The functional form and different model structures of HBM and localized and generalized models are discussed in section 2.2. Methods and metrics to assess model performance are discussed in section 2.3. The best performing HBM structure is chosen in section 3.1. The HBM, localized, and generalized model parameters are explained in section 3.2. The development of a HBM with group-level predictors is described in section 3.3. The predictive performance of the models and inferences are explained in section 3.4.

2. Data and Methods

2.1. Data

Object (household)-level empirical flood loss data are available via computer-aided cross-sectional telephone surveys of private households that have suffered from losses due to floods in 2002, 2005, 2006, 2010, 2011, and

Table 1
Sample Size, the Summary (Median) of Water Depth (wd) in Meters, Exposed Building Value (bv) in EUR, Absolute and Relative Losses to Residential Buildings (bloss in EUR, rloss), Inundation Duration (d) in Hours, Footprint Areas of the Buildings (ba) in Square Meter and Return Period of the Event (rp) in Years, Prevalence of Private Precaution (pre)—Percentage of Households That Implemented One or More Private Precautionary Measures, Including Waterproof Sealing, Flood Adapted Use and Flood Adapted Interior Fitting, Prevalence of Flood Experience (fe)—Percentage of Households that Have Experienced at Least One Flood Event in the Past, Prevalence of Building Types (bt)—Percentage of Buildings That Are Single-Family Houses (bt1), Multifamily Houses (bt2), and Semidetached Houses (bt3) for Each of the Spatiotemporal Groups

										bt				
Catchment	Event	Sample size	wd	bv	*bloss*	*rloss*	pre	fe	d	bt1	bt2	bt3	ba	rp
Danube	2002	225	1.7	354,785	6,258	0.015	36	40	15	28	27	45	170	53
	2005	104	2.0	406,012	7,874	0.015	65	52	24	25	37	38	200	39
	2013	79	3.0	571,536	45,000	0.060	68	32	96	24	13	63	206	>1,000
Elbe	2002	518	3.5	302,005	43,805	0.096	21	17	120	30	22	48	144	190
	2006	42	2.9	307,800	6,962	0.018	86	78	156	10	29	61	142	28
	2011	58	2.7	475,456	9,140	0.015	78	67	24	29	14	57	160	30
	2013	492	2.7	427,680	23,250	0.051	80	58	168	13	15	72	150	112
Oder	2010	75	3.3	376,200	32,258	0.060	73	29	30	16	21	63	150	366
Rhine	2011	70	2.2	531,300	2,092	0.004	99	81	48	26	20	54	205	19
Total		1,663												

Note. Values adjusted for inflation to values as of 2013 using the building price index (DESTATIS, 2013).

2013 in the Elbe, Danube, Rhine, and Oder catchments in Germany using a standardized questionnaire. Using the flood masks derived from satellite data (DLR, Center for Satellite Based Crisis information, https://www.zki.dlr.de/), a list of affected streets was derived. The telephone numbers of households in these streets were obtained from public telephone directory. The survey campaigns always focused on a single event and used a questionnaire with about 180 questions regarding aspects of hazard, exposure, vulnerability, and residential building and content losses. Water depth above ground level is determined using the reported water level in the highest affected story by applying corrections based on the presence of a basement and height of the ground floor. Relative loss to buildings, *rloss*, is the ratio of absolute building loss (Euro) to its total replacement value (Euro) at the time of the event (Elmer et al., 2010). Hence, *rloss* has a range of 0 to 1, where 0 indicates no building damage and 1 indicates total loss of the building. More information about the individual flood events, the surveys, and their results were published in Thieken et al. (2007), Kreibich et al. (2011), Kreibich, Botto, et al., (2017), Kienzler et al. (2015), and Vogel et al. (2018). For our study, we selected from these surveys all data sets that refer to residential buildings with basements (for unbiased measurements of water depth) and for which information on water depth and relative building loss is available. In the context of spatiotemporal transferability of flood loss models, the event during which the household experienced flooding is used to group the households temporally and the catchment in which the household is located is used for spatial grouping. Nine region and event groups with considerable number of completed data sets (>25) are considered in this study, resulting in total 1,663 data sets. Information regarding each of the spatiotemporal groups is reported in Table 1.

From Table 1, the events in the Elbe catchment in 2002 and 2013 were extreme floods, which affected a large number of households. Though the events were both extreme, owing to an increase in prevalence of flood experience and private precaution, the losses caused due to the 2013 floods in the Elbe is significantly lower than the losses caused due to 2002 floods. In both Elbe and Danube catchments, there is an increase in prevalence of flood experience and private precaution after the 2002 event. The June 2013 event in the Danube catchment resulted in large spatial extent of flood peaks with high magnitudes. This flood was in hydrological terms the most severe flood in Germany at least for the last six decades (Schröter et al., 2015). Also, the average duration of inundation in most areas during the 2013 event was close to 4 days. Therefore, despite high flood experience and improvements in private precaution, this event resulted in high losses. In the Rhine catchment, though the median water depth experienced by households during the 20-year return period event in 2011 was 2.2 m, these households suffered the least amount of losses. A possible explanation for this is that more than 80% of these households had high flood experience and 99% of the households had implemented one or more private precautionary measures.

2.2. Modeling Flood Damage Processes

2.2.1. Functional Form and Bayesian Parameter Estimation

A flood loss model based on a depth-damage function is set up to estimate relative loss (*rloss*) suffered by individual residential buildings. A square root function of water depth (wd) in meters is used ($\sqrt{\text{wd}}$), since this functional form has been proven to be suitable (Merz et al., 2013; Rözer et al., 2019; Schröter et al., 2014; Wagenaar et al., 2017). Values of *rloss* lie between 0 and 1. In contrast to deterministic models that assume certainty in the process and determine the outcome as a point estimate, probabilistic models result in a probability distribution representing the uncertainty in the model structure, parameters, and noise in the data. Random variables and probability distributions are incorporated in probabilistic models. A probabilistic flood loss model is set up to estimate relative losses. Since, *rloss* values are bounded between 0 and 1, prediction from regression models using unbounded distributions may result in implausible values. Therefore, *rloss* is modeled as a beta distribution bounded between 0 and 1 (Rözer et al., 2019). The shape parameters of the beta distribution, α and β, can be algebraically determined using mean μ and precision φ (equation (1)). μ is the mean *rloss* which is a function of $\sqrt{\text{wd}}$ and φ represents the precision (inverse of variance) of the distribution of estimated *rloss* values for each household.

The function parameters of μ (slope and intercept) and φ are estimated using Markov Chain Monte Carlo (MCMC) sampling. Since μ is the expected value of *rloss* that needs to be positive, we use a log-link function. To estimate the parameter values, we start with our general belief about the distribution of the parameters (priors) and then use evidence the data (represented as likelihood). Monte Carlo simulations create a large number of replications of these parameters that represent the damage processes, which results in approximate posterior distributions for relative loss estimates (*rloss*). The MCMC sampling assumes memoryless property or Markov property by which, during an iteration, if the current state of the estimated parameters represents the data generation process better than the immediate previous one, it is added to the chain of parameter values. Hence, when a large number of iterations are run, the parameter values are not influenced by where the sampling began initially. Though the posterior distribution of the parameters is estimated from the priors and the likelihood, the evidence from data dominates the prior beliefs. However, giving appropriate priors helps us to improve efficiency of the parameter search and also rejects implausible parameter values. For the flood loss model represented by equation (1), weakly informative generic priors are provided. For example, the water depth is constrained to be positively correlated with *rloss*.

$$\begin{aligned} rloss &\sim \text{beta}(\alpha, \beta) \\ \alpha &= \mu \times \varphi \\ \beta &= (1-\mu) \times \varphi \\ \mu &= \text{E}(rloss) \\ \log(\mu) &= f\left(\sqrt{\text{wd}}\right). \end{aligned} \tag{1}$$

2.2.2. HBM

A HBM is a multilevel probabilistic regression model that estimates a set of coefficients for each group while the predictors are used to model the outcomes. There is a second probability distribution over these group-level parameters that govern the variability between the groups. Model parameters that remain constant across all the groups are termed as shared parameters or fixed effects. Parameters that vary across different groups are termed as varying effects.

Given the functional form from equation (1), the damage processes can be modeled to vary either randomly between region and event groups (varying intercept model) or conditioned on $\sqrt{\text{wd}}$ (varying slope model) or a combination of both as shown in Table 2. A varying intercept suggests that damage processes may vary randomly between the groups, whereas a varying slope suggests that the damage processes vary conditioned on the square root of water depth. A model structure with varying intercept is commonly applicable when the median building losses conditioned on water depth at each region and event group are different. A model structure with varying slope is recommended when the spread/variance of building losses conditioned on water depth at each region and event group is different. In a varying intercept model, the variability in damage processes between groups of households remains the same irrespective of the water depth experienced by the households. For example, consider a small flood event in a well-prepared neighborhood, if

Table 2
Specification of the Eight HBM Structures (M1–8) Tested in This Study

HBM structure	Description	Model structure specification
M1	Varying intercept between spatial groups (regions)	$\log(\mu_i) = \theta \times \sqrt{\text{wd}_i} + \varepsilon_r$ $\varepsilon_r \sim \text{normal}(\mu'_r, \sigma'_r)$
M2	Varying intercept between spatiotemporal groups (regions-events)	$\log(\mu_i) = \theta \times \sqrt{\text{wd}_i} + \varepsilon_{re}$ $\varepsilon_{re} \sim \text{normal}(\mu'_{re}, \sigma'_{re})$ $\mu'_{re} \sim \text{normal}(\mu'_r, \sigma'_r)$
M3	Varying slope between spatial groups (regions)	$\log(\mu_i) = \theta_r \times \sqrt{\text{wd}_i} + \varepsilon$ $\theta_r \sim \text{normal}(\mu_r, \sigma_r)$
M4	Varying slope between spatiotemporal groups (regions-events)	$\log(\mu_i) = \theta_{re} \times \sqrt{wd_i} + \varepsilon$ $\theta_{re} \sim \text{normal}(\mu_{re}, \sigma_{re})$ $\mu_{re} \sim \text{normal}(\mu_r, \sigma_r)$
M5	Varying slope and intercept between spatial groups (regions)	$\log(\mu_i) = \theta_r \times \sqrt{\text{wd}_i} + \varepsilon_r$ $\theta_r \sim \text{normal}(\mu_r, \sigma_r)$ $\varepsilon_r \sim \text{normal}(\mu'_r, \sigma'_r)$
M6	Varying slope between spatiotemporal groups (regions-events) and varying intercept between spatial groups (regions)	$\log(\mu_i) = \theta_{re} \times \sqrt{\text{wd}_i} + \varepsilon_r$ $\theta_{re} \sim \text{normal}(\mu_{re}, \sigma_{re})$ $\mu_{re} \sim \text{normal}(\mu_r, \sigma_r)$ $\varepsilon_r \sim \text{normal}(\mu'_r, \sigma'_r)$
M7	Varying slope between spatial groups (regions) and varying intercept between spatiotemporal groups (regions-events)	$\log(\mu_i) = \theta_r \times \sqrt{\text{wd}_i} + \varepsilon_{re}$ $\theta_r \sim \text{normal}(\mu_r, \sigma_r)$ $\varepsilon_{re} \sim \text{normal}(\mu'_{re}, \sigma'_{re})$ $\mu'_{re} \sim \text{normal}(\mu'_r, \sigma'_r)$
M8	Varying slope and intercept between spatiotemporal groups (regions-events)	$\log(\mu_i) = \theta_{re} \times \sqrt{\text{wd}_i} + \varepsilon_{re}$ $\theta_{re} \sim \text{normal}(\mu_{re}, \sigma_{re})$ $\mu_{re} \sim \text{normal}(\mu_r, \sigma_r)$ $\varepsilon_{re} \sim \text{normal}(\mu'_{re}, \sigma'_{re})$ $\mu'_{re} \sim \text{normal}(\mu'_r, \sigma'_r)$

Note. In the model structure specification, $\sqrt{\text{wd}_i}$ is the square root of water depth at ith household, θ and ε are the coefficients of water depth and intercept, respectively. The shared parameters that are common to all region and event groups are represented without subscripts, that is, θ and ε. Subscript *i* refers to *i*th household; subscript *re* refers to the group of households belonging to a particular region and event group; subscript *r* refers to the group of households belonging to a particular region group. The priors of the parameters are represented as ~. HBM = Hierarchical Bayesian Model.

majority of the households do not have expensive fittings or valuables in the lower floors, then the overall exposure value is reduced. Hence, irrespective of the experienced water depth, all the households in the region will incur less damage on average compared to a group of households with low preparedness. A model structure with varying slope suggests that the variability in damage processes is dependent of the water depth and more reflected in the estimated building loss for households experiencing higher water depths. An example of damage processes with varying slopes is the effect of contamination. Contaminated water causes more damage to building structure even at smaller water depth, and the magnitude of damage due to contamination also increases with increasing water depth. Similarly, a reduction in incurred damage is seen due to measures such as water barriers. However, the effectiveness of these measures is dependent on the water depth. Beyond a certain level of water depth, the measures can only reduce loss and not prevent it completely.

When the model structure includes varying effects between different groups, there is always an overarching probability distribution in the hierarchy governing these variations. For example, in a HBM structure, where the slope and intercept are made to vary between regions, a second distribution governs the variability of the slope and intercept across the regions (see Table 2: Model Structure M5). Similarly, when the slope and intercept are made to vary between region and event groups, there are overarching distributions at two levels, governing their variability across the region and event groups and also across regions (see Table 2: Model Structure M8). A number of HBM structures can be formulated using a depth-damage function. For choosing the best model structure, we select eight meaningful model structures based on the premise that the variability in damage processes of households across multiple events is always conditioned on the region in which the households are located (see Table 2: Model Structures M2, M4, M6, and M8). Among the tested model

10.1029/2019WR025068

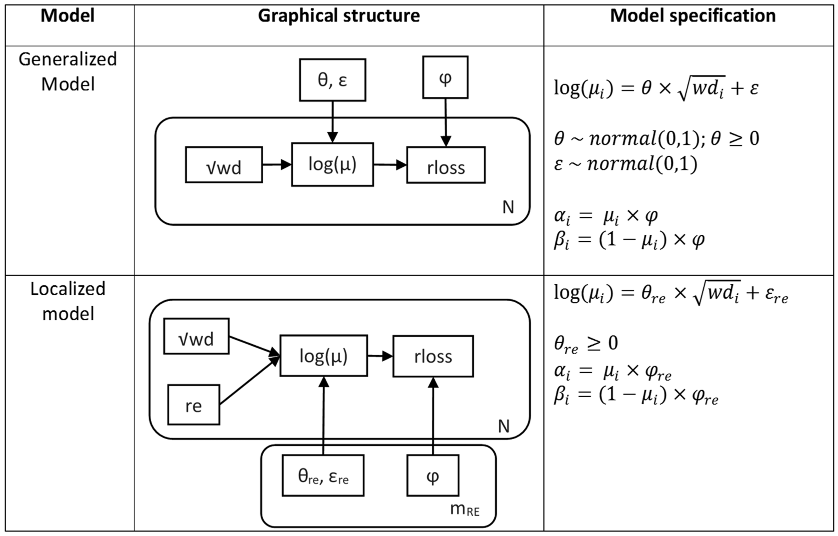


Figure 1. Generalized and localized models—graphical structure and specification. The graphical illustration is adopted from Levy et al. (2012). A box with rounded corners represents a particular level in the hierarchy, and the indicators at its bottom right corner refer to the number of entities in the particular level. *N* refers to the total number of households in the model (1,663); m_{re} refers to the number of region and event groups ((9) see Table 1). Subscript *i* refers to *i*th household; subscript *re* refers to the group of households belonging to a particular region and event group. In the localized model structure, variable *re* refers to the region and event group of each household.

structures, the one with the best prediction capability is chosen and compared against the generalized and localized models that are introduced below.

Since we do not intend to implement strict constraints over the parameters, weakly informative generic priors are provided for the shared parameters (Gelman et al., 2017). θ~*normal*(0,1), ε~normal(0,1) and the coefficient of $\sqrt{\text{wd}}$; θ is constrained to be positive. Weakly informative generic priors are also provided for region-level hyper-priors in the varying slope and intercept models, μ_r,σ_r and μ'_r,σ'_r, respectively. μ_r~normal(0,1), σ_r~cauchy(0,10), μ'_r~normal(0,1), and σ'_r~cauchy(0,10). The parameters for standard deviation, σ_r and σ'_r, are constrained to be nonnegative.

2.2.3. Generalized Model

In a generalized model, a single set of parameters is estimated, irrespective of any grouping. Hence, there is only one level in the model structure. Most flood loss models developed using empirical data from multiple regions and events are generalized models (Kreibich, Di Baldassarre, et al., 2017; Merz et al., 2013). The damage processes across all events and regions are generalized given the flood loss predictors. Adopting this approach to parameterize the depth-damage function (equation (1)) generalizes the damage processes conditioned on $\sqrt{\text{wd}}$ and results in a single slope estimate (θ), intercept (ε), and precision (φ), as shown in Figure 1. Weakly informative priors are provided for θ and ε; θ and φ are constrained to be positive.

2.2.4. Localized Model

A localized model uses an independent set of parameters for each group. Flood loss models developed using empirical data from specific events and regions can be considered as localized models. The localized model approach to parameterize the depth-damage function from equation (1) results in slope (θ_{re}), intercept (ε_{re}), and the precision parameter (φ_{re}), as shown in Figure 1. θ_{re} and φ_{re} are constrained to be positive. These parameters are estimated independently for every region and event group (*re*). In the absence of sufficient data for each region and event group, the localized modeling approach may result in unreliable, noisy estimates.

2.3. Analyzing the Predictive Performance of Models

The predictive performance of the models is determined by comparing the predicted relative loss estimates to the observed relative losses. Two validation tests, that is, out-of-sample and out-of-group validations, are performed using three performance metrics—expected log-pointwise predictive density (*elpd*), Mean Absolute Error (MAE), and Mean Bias Estimate (MBE). *elpd* (equation (2)) is a measure of the predictive accuracy of the model for data points considered (Vehtari et al., 2017).

$$elpd = \sum_{i=1}^{n} \int p_t(\widetilde{y}_i) \log p(\widetilde{y}_i|y) d\widetilde{y}_i, \tag{2}$$

where $p_t(\widetilde{y}_i)$ is the true density of observed *rloss* for *i*th household and $p(\widetilde{y}_i|y)$ is the posterior predictive distribution for *rloss* for *i*th household using the model. The sum of predictive densities over *n* households involved in the validation is used to reflect the accuracy of the model. The advantage of using expected pointwise predictive density is that *elpd* is a fully Bayesian method that estimate out-of-sample predictive performance of the model using the entire posterior distribution, whereas commonly used information criteria only consider goodness of fit using maximum likelihood of the predictions, which is a point estimate (Gelman et al., 2014).

$$MAE = \frac{1}{n} \sum_{i=1}^{n} \left| \widetilde{rloss}_i - rloss_i \right| \times building\ value_i, \tag{3}$$

$$MBE = \frac{1}{n} \sum_{i=1}^{n} \widetilde{rloss}_i - rloss_i. \tag{4}$$

Prediction errors in the point estimates (median of posterior distributions) of *rloss* from the probabilistic depth-damage functions are reported using *MAE* and *MBE*. In equations (3) and (4), *n* refers to the total number of households in the validation data set; $rloss_i$ and $\widetilde{rloss}_i$ are the observed and predicted relative loss point estimates for *i*th household. Models resulting in lower values of *MAE* and lower absolute values of *MBE* have better prediction capabilities.

2.3.1. Out-of-Sample Validation

Out-of-sample validation measures the model performance in predicting losses incurred by households that have not been used in model development but belong to the same regions and events used in model development. *elpd* for out-of-sample validation is estimated using leave-one-out cross-validation (LOO-CV), by determining the model prediction accuracy while excluding households, one at a time. This process is approximated using Pareto smoothed importance sampling (PSIS), implemented by Vehtari et al. (2017). The shape parameter of the Pareto smoothed distribution $\widehat{k}$ is required to be less than 0.7 for the *elpd* estimate to be reliable (Vehtari et al., 2017). While applying PSIS approximation, as a conservative estimate, the difference in *elpd* between the models is considered significant when it is greater than 4 times the standard error (SE) whose corresponding *p* value is <0.0001. *MAE* for out-of-sample validation are determined using a tenfold cross-validation performed by iteratively removing 10 equal-sized random samples of households without replacement, one at a time, refitting the model and predicting the losses suffered by the held-out households.

2.3.2. Out-of-Group Validation

Out-of-group validation is used to measure the model performance in predicting losses incurred by households that experienced a new event. The new event may either occur in a region that has already been included in the model development (temporal transferability) or a new region (spatial transferability). Out-of-group validation is performed using leave-one-group-out cross-validation. To estimate a model's capability in predicting losses for a new event, households are held out while fitting the model, one event at a time, and losses incurred by the held-out households are predicted. Similarly, a model's prediction capability for new regions is estimated by removing the households belonging to individual regions, one region at a time, refitting the model and predicting the losses for households in the held-out region. Since the localized models are completely localized, they cannot be tested for transferability in the same way. The localized models developed for each region and event group are applied to the other region and event groups. During the transfer, the average of out-of-group prediction errors from the individual models is used to determine the performance of the localized model. In order to nullify the bias due to varying numbers of households in different region and event groups, stratified bootstrap sampling with equal number of households (400 from each region and 200 from each region- and event-group) with replacement is performed while estimating *elpd* for out-of-group validation.

Table 3
Out-of-Sample Predictive Performances of Potential Hierarchical Bayesian Model Structures

	Out-of-sample LOO-CV with PSIS approximation	
Model comparison	*elpd* difference (SE)	Model comparison read as > superior = equal <inferior
M1 vs. M2	−60 (12)	M1 < M2
M2 vs. M3	53 (12)	M2 > M3
M2 vs. M4	−21 (4)	M2 < M4
M4 vs. M5	74 (15)	M4 > M5
M4 vs. M6	0 (1)	M4 = M6
M4 vs. M7	17 (4)	M4 > M7
M4 vs. M8	0 (0.6)	M4 = M8

Note. LOO-CV with PSIS approximation is used to estimate and compare the out-of-sample expected log-pointwise. The standard errors (SEs) of the comparisons are shown in brackets. *elpd*, expected log-pointwise predictive density; LOO-CV, leave-one-out cross-validation; PSIS, smoothed importance sampling.

3. Results and Discussion

3.1. HBM Structure

The out-of-sample predictive performances of the eight potential HBM structures are provided in Table 3. When the *elpd* difference is significant (i.e., >4 SE) and positive, then the first model performs better than the second and vice versa. For all the model comparisons, the PSIS $\hat{k}$ values were less than the recommended estimate of 0.7, indicating that the *elpd* estimate is reliable (Vehtari et al., 2017). The *elpd* difference between M4 and the two model structures M6 and M8 are insignificant, implying that these model structures show similar out-of-sample predictive performance. M2 performs better than M1 and M3. M4, M6, and M8 perform better than M5 and M7. Among these three models showing similar performance, M4 has the least complexity (least number of parameters). M4 also performs better than M2. The Kruskal-Wallis (Hollander & Wolfe, 1973) test also confirms that in the varying intercept models (M6 and M8), there is no significant difference in the intercepts for various spatial and spatiotemporal groups. Therefore, we choose M4 (varying slope between spatiotemporal groups) as the appropriate model structure. Its graphical structure and model specifications are shown in Figure 2. Thus, this HBM structure is proposed for flood loss estimation and, in the following, tested against the generalized and localized models.

According to the chosen structure, HBM—M4 (Figure 2), *rloss* is modeled using

1. θ_{re}—effect of *wd* on *rloss* that is specific to each region and event levels,
2. ε—shared intercept for all region and event levels,
3. μ_{re}, σ_{re}—distribution parameters at region- and event-level governing θ_{re},
4. μ_r, σ_r—distribution parameters at region-level governing μ_{re},
5. ϕ—common precision parameter for distribution of *rloss*, and
6. weakly informative priors for ε, μ_r, σ_r, and σ_{re}.

The best performing model structure, HBM—M4, includes a single shared intercept, ε, and varying slope, θ_{re}. In addition to the distribution governing the varying slope (θ_{re}) at the region and event level, HBM—M4 comprises a second governing distribution at the region level. The varying slope across different groups of households accounts for variability in damage processes conditioned on $\sqrt{\text{wd}}$. A single shared intercept across different groups implies no random variability in damage processes independent of water depth across the groups. The distribution parameters at the region level (μ_r, σ_r) capture the variability of damage processes across regions, which is consistent across multiple events in the same region.

Based on expert knowledge regarding the drivers of damage processes in different region and event groups, the best performing model structure, HBM—M4, is justifiable. The implementation of private precautionary measures was increased by more than 40% after the 2002 floods in Germany. However, the implemented measures did not completely prevent losses during extreme fluvial floods (Table 1: Median water depth for all the events was more than 1.5 m). Since most of the property-level flood barriers were overtopped during these events (Hudson et al., 2014; Sairam et al., 2019), the implemented measures could mostly reduce the

10.1029/2019WR025068

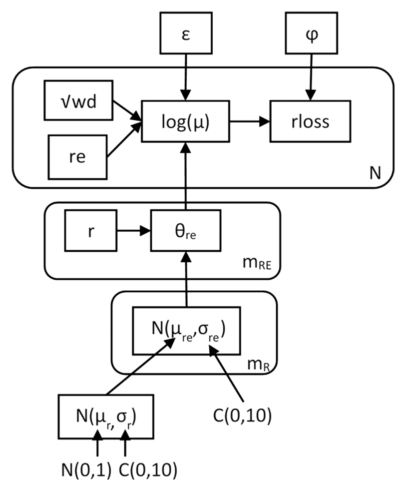
Model	Graphical structure	Model specification
HBM - M4		$\log(\mu_i) = \theta_{re} \times \sqrt{wd_i} + \varepsilon$ $\theta_{re} \sim normal(\mu_{re}, \sigma_{re})$ $\mu_{re} \sim normal(\mu_r, \sigma_r)$ $\sigma_{re} \sim cauchy(0,10)$ $\mu_r \sim normal(0,1)$ $\sigma_r \sim cauchy(0,10)$ $\alpha_i = \mu_i \times \varphi$ $\beta_i = (1 - \mu_i) \times \varphi$ $rloss_i \sim beta(\alpha_i, \beta_i)$

Figure 2. Hierarchical Bayesian model (HBM) with Structure M4 (HBM—M4) graphical structure and specification. The graphical illustration is adopted from Levy et al. (2012). A box with rounded corners represents a particular level in the hierarchy, and the indicators at its bottom right corner refer to the number of entities in the particular level. For example, *N* refers to the total number of households in the model (1,663); m_{RE} refers to the number of region and event groups (9); m_R refers to the number of region-groups ((4); see Table 1). Subscript *i* refers to *i*th household; subscripts *r* and *re* refer to the group of households belonging to a particular region group and region and event group, respectively. Variables *r* and *re* refer to the region group and region and event group of each household.

impact of flooding but not prevent it completely. In this respect, the variability in damage processes across region and event groups is always influenced by the experienced water depth. Hence, a single shared intercept (ε) between region and event groups in HBM—M4 is reasonable.

The varying slope (θ_{re}) in HBM—M4 is reasonable since the variability in damage processes between region and event groups is more pronounced in households experiencing higher water depths, especially during extreme events. Extreme events generally affect larger areas, and the households that are not affected during more frequent events might experience flooding. These households generally have low preparedness. Hence, encountering an extreme event with high water depths may result in higher amount of incurred losses (Elmer et al., 2010). θ_{re} in HBM—M4 potentially captures this characteristic of damage processes due to differences in exposure to flooding and preparedness. Some exposure and vulnerability characteristics pertaining to a particular region such as predominant building construction types and socioeconomic characteristics of households do not vary across frequent events. Hence, in addition to the distribution governing the varying slope (θ_{re}) at the region and event level, HBM—M4 comprises a second governing distribution at the region level. Thus, along with event-specific variability, the model is also capable of capturing such region-specific variability such as land use and predominant building construction types, which are consistent across multiple events, occurring in a short time span.

3.2. Model Parameters

The HBM—M4 has a single shared intercept ($\varepsilon = -3.75$) but separate slopes for each region and event group with overarching distributions as shown in Figure 3a. The parameters of the overarching distributions (μ_{re}, σ_{re}, and μ_r, σ_r) provide finite variance for the slopes across region and event groups. The distribution parameters of the slope, intercept, and the overarching distributions are provided in sections S1–S4 in the supporting information. In the HBM, slopes with large deviations from the governing distribution means are penalized. This effect is termed as "shrinkage" (Levy et al., 2012). In the absence of shrinkage, the variance of slopes across the groups can range from zero to infinity. Alternatively, complete shrinkage generalizes the damage processes as the variance of slopes between the region and event groups reduces to zero. Thus, the aspect of shrinkage, which is ubiquitous to hierarchical models, helps to achieve a balance between bias and variance. The slopes from the HBM—M4 pertaining to each region and event group significantly vary

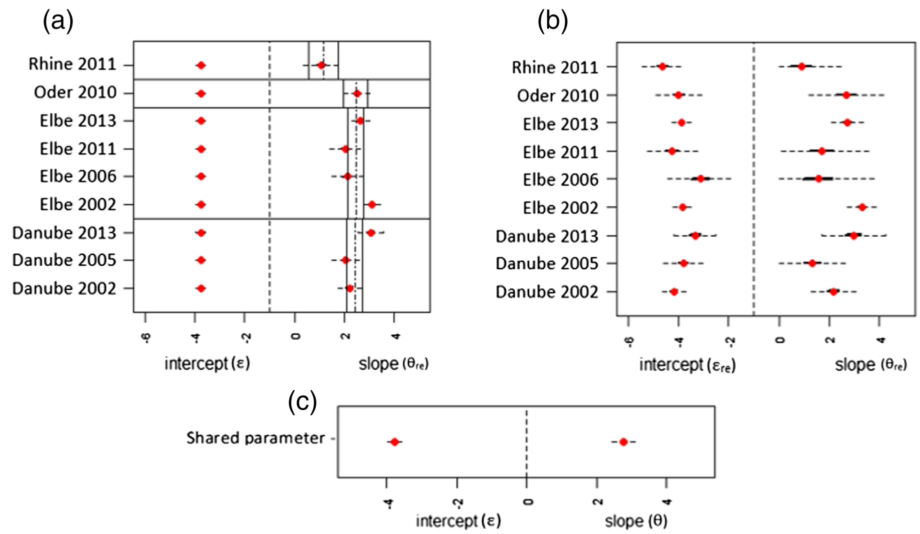


Figure 3. (a) Intercept (ε) and slope (θ_{re}) parameters estimated from the hierarchical Bayesian model—M4 with μ_{re} for each region is represented as dot-dash vertical lines with solid vertical lines showing the 95% confidence interval, (b) intercept (ε_{re}) and slope (θ_{re}) parameters estimated from the localized model structure, and (c) intercept (ε) and slope (θ) parameters estimated from the generalized model structure.

from each other (Figure 3a). This proves the presence of large variability in damage processes between event and region groups. For example, the depth-damage relationship for the extreme flood event in 2002 in the Elbe region has the highest slope ($\theta_{\text{Elbe2002}} = 3.29$) reflecting the strong influence of water depth on building loss. However, the succeeding event in 2013 in the Elbe has a much smaller slope ($\theta_{\text{Elbe2013}} = 2.79$) indicating the improved resistance of households to flood damage compared to the 2002 event. The means of the distributions governing the variabiality of the slopes (θ_{re}) within each region, μ_{re} (Figure 3a), do not show much variation between the Elbe ($\mu_{\text{Elbe}(02,06,11,13)} = 2.23$) and Danube ($\mu_{\text{Danube}\ (02,05,13)} = 2.26$) regions. This suggests that the variability in damage processes across different events within the same region is much higher than the variability across regions.

The localized model results in independent sets of slope and intercept estimates for each region and event groups as shown in Figure 3b. From the distributions of slope and intercept for every region and event group provided in sections S1 and S4, we find that the parameters estimated using localized models have large uncertainty except for the extreme events of 2002 and 2013 with large sample of empirical data set. Consistent inferences regarding damage processes cannot be made from these noisy parameter estimates. The generalized model results in a single slope parameter (coefficient of $\sqrt{\text{wd}}$, $\theta = 2.78$) and intercept ($\varepsilon = -3.77$) for all region and event groups (Figure 3c). The damage processes represented by the parameters of the generalized model are more inclined toward extreme events (such as Elbe 2002 and 2013 and Danube 2013) and may not accurately capture the damage processes of small events (such as Elbe 2006 and Danube 2005). The distributions of the slope and intercept from the generalized model are provided in sections S1 and S4.

3.3. HBM With Group-Level Predictors

For many regions, detailed empirical data concerning flood depths and incurred losses may be unavailable at the household level. Undertaking household-level surveys are quite tedious and also implausible if there is no available record of flooding in the region or if the last flood event occurred a long time ago. Within the hierarchical framework, there is a possibility to include group-level predictors that may potentially improve model predictions. Hence, in order to improve risk assessment for region and event groups for which empirical loss data are unavailable, we include group-level predictors that are explanatory variables obtained on basis of aggregated data or expert knowledge, pertaining to a region and event group. For example, there may be cases where residents leave a region after an extreme event. In these cases, the temporal variability in building occupancy and overall exposure can be included as group-level predictors in order to explain the variability in damage processes.

Table 4
Results of Stepwise Regression Predicting Varying Slopes of Hierarchical Bayesian Model—M4

Step	Model	R^2	Adjusted R^2	AIC	BIC
1	—			−3.84	24.10
2	fe_{re}	0.71	0.66	−12.89	15.25
3	$fe_{re} + d_{re}$	0.92	0.89	−22.18	6.15
4	$fe_{re} + d_{re} + rp_{re}$	0.96	0.93	−26.49	2.04

Note. AIC, Akaike information criterion; BIC, Bayesian information criterion. R^2 is the coefficient of determination. It is a measure of how well the slopes of HBM—M4 are replicated by the regression model. Adjusted R^2 is a variant of R^2 that is penalized for increasing number of explanatory variables.

Identifying group-level predictors that improves flood loss predictions during spatiotemporal transfer requires a good understanding of the variability in damage processes across region and event groups. In our study, we statistically derive the group-level predictors by attributing the varying slopes (θ_{re}) from HBM—M4 to loss influencing/resisting aspects pertaining to respective region and event groups. A stepwise linear regression (Venables & Ripley, 2002) with 1,000 iterations is performed to predict the varying slopes of depth-damage functions from the HBM—M4 using the attributes from Table 1. The model is updated in steps with the best predictors using generalized Akaike information criterion and Bayesian information criterion. Both Akaike information criterion and Bayesian information criterion score the model based on goodness of fit and also penalize the model for overfitting based on the number of parameters. These criteria are used for determining the best predictors in a regression model. We determine that among the group-level attributes influencing flood losses, from Table 1, prevalence of flood experience (fe_{re}), duration of inundation (d_{re}), and return period of the event (rp_{re}) are crucial in explaining the spatiotemporal variability in damage processes (Table 4).

We hypothesize that region- and event-group-level predictors such as the percentage of households that have prior flood experience, the median duration of inundation in the region, and return period of the event improve the performance of the HBM—M4 during transferability scenario. The HBM—M4 with group-level predictors includes interaction terms, fe_{re}, d_{re}, and rp_{re}, representing the prevalence of flood experience, median duration of inundation, and return period in every region-event groups as shown in Figure 4. The varying slope θ_{re} is defined as a linear function of fe_{re}, d_{re}, and rp_{re} with their Coefficients A, B, and C, respectively, and Intercept D. The distributions of the Parameters A, B, and C obtained via MCMC sampling are included in section S5.

3.4. Predictive Performance of Models

The out-of-sample and out-of-group prediction errors (*MAE* and *MBE*) are summarized according to region and event groups in Tables 5a and 5b, respectively. The *elpd* comparison for the models are provided aggregated for all the region and event groups in Table 5c.

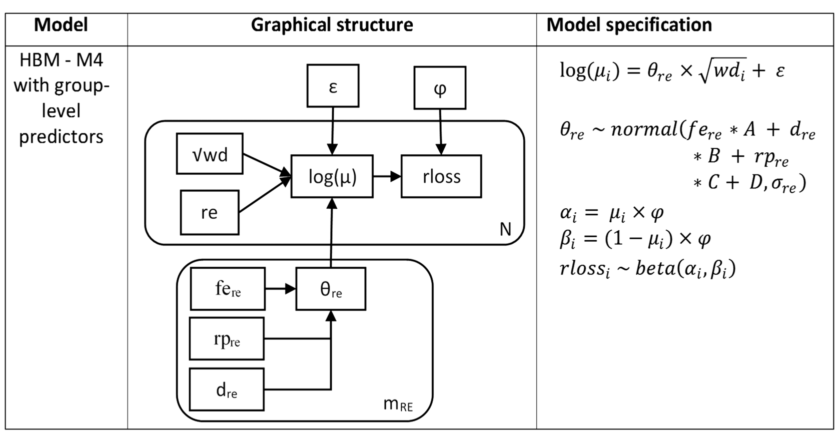


Figure 4. Hierarchical Bayesian model (HBM)—M4 with group-level predictors. The graphical illustration is adopted from Levy et al. (2012). In the structure, the box with rounded corners represents a particular level in the hierarchy, and the indicators at its bottom right corner refer to the number of entities in the particular level. For example, N refers to the total number of households in the model (1,663); m_{RE} refers to the number of region and event groups ((9); see Table 1). In the model specification, subscript i refers to ith household; subscript re refers to the group of households belonging to a particular region and event group. The variable re refers to the region and event group of each household.

10.1029/2019WR025068

Table 5
Accuracy Assessment of Generalized, Localized, and Hierarchical Models

(a) MAE of medians of posterior relative loss distributions. Error from the best performing model is shown in bold.

Accuracy Assessment	Model	Danube 2002	Danube 2005	Danube 2013	Elbe 2002	Elbe 2006	Elbe 2011	Elbe 2013	Oder 2010	Rhine 2011
k-fold (out-of-sample)	Localized	0.016	0.021	0.066	0.074	0.024	0.025	0.038	0.045	**0.005**
	Generalized	0.023	0.036	0.048	**0.070**	0.042	0.035	0.038	0.046	0.037
	HBM—M4	**0.014**	**0.020**	**0.043**	0.074	**0.020**	**0.015**	**0.031**	**0.043**	0.008
Out-of-group (leave-one-event-out)	Localized	0.228	0.033	0.049	0.091	0.041	0.037	0.037	NA	
	Generalized	0.025	0.035	0.044	0.091	0.042	0.038	0.039		
	HBM—M4	0.019	0.029	0.035	0.090	0.029	0.025	0.033		
	HBM—M4 with group-level predictors	**0.013**	**0.020**	**0.018**	**0.075**	**0.026**	**0.016**	**0.032**		
Out-of-group (leave-one-region-out)	Localized	0.019	0.028	0.049	0.098	0.035	0.029	0.038	0.050	0.032
	Generalized	0.024	0.034	0.051	0.103	0.028	0.025	0.031	0.048	0.040
	HBM—M4	**0.013**	**0.019**	**0.048**	**0.087**	**0.018**	**0.014**	**0.030**	**0.042**	**0.021**

(b) MBE of medians of posterior relative loss distributions. Error from the best performing model is shown in bold.

Accuracy assessment	Model	Danube 2002	Danube 2005	Danube 2013	Elbe 2002	Elbe 2006	Elbe 2011	Elbe 2013	Oder 2010	Rhine 2011
k-fold (out-of-sample)	Localized	−0.002	**−0.003**	−0.019	0.008	−0.006	0.003	−0.006	0.005	−0.004
	Generalized	−0.005	−0.019	−0.024	0.039	0.033	−0.026	−0.004	−0.003	−0.029
	HBM—M4	**−0.001**	**−0.003**	**−0.001**	**0.004**	**−0.003**	**−0.000**	**−0.003**	**0.001**	**−0.002**
Out-of-group (leave-one-event-out)	Localized	−0.005	−0.005	0.052	0.078	−0.011	−0.011	0.026	NA	
	Generalized	−0.006	−0.019	0.018	0.104	−0.035	−0.026	0.013		
	HBM—M4	0.004	−0.002	0.014	0.065	−0.004	−0.006	0.005		
	HBM—M4 with group-level predictors	**0.003**	**0.000**	**−0.001**	**0.008**	**−0.002**	**−0.001**	**0.002**		
Out-of-group (leave-one-region-out)	Localized	**0.001**	0.023	0.011	−0.035	−0.007	−0.008	0.014	0.016	−0.020
	Generalized	−0.006	−0.019	−0.002	0.074	−0.015	−0.013	0.010	−0.011	−0.031
	HBM—M4	**0.001**	**0.007**	**0.000**	**0.015**	**0.003**	**−0.002**	**0.008**	**0.009**	**−0.004**

(c) Differences in expected log-pointwise predictive density

Accuracy assessment		Model comparison	*Elpd* difference (SE)	Model comparison read as > superior,=equal, <inferior
Out-of-sample	Leave-one-out (LOO-CV) with PSIS approximation	Localized vs. Generalized	196 (30)	Localized > generalized
		Localized vs. HBM—M4	76 (30)[a]	Localized = HBM—M4
Out-of-group	Leave-one-event-out (temporal transfer)	Localized vs. Generalized	−568 (120)	Localized < Generalized
		Generalized vs. HBM—M4	−91 (31)	Generalized < HBM—M4
		HBM—M4 vs. HBM—M4 with group-level predictors	−131 (44)	HBM—M4 < HBM—M4 with group-level predictors
	Leave-one-region-out (spatial transfer)	Generalized vs. HBM—M4	−57 (24)	Generalized < HBM—M4

[a]Insignificant difference (*elpd* difference from LOO-CV with PSIS approximation <4 SE)

In terms of out-of-sample prediction accuracy (*k*-fold cross-validation), the HBM—M4 has smaller point-estimate error (*MAE*) compared to the generalized model except for the 2002 event in Elbe. The generalized model resulted in the least out-of-sample *MAE* for the 2002 event in Elbe. From the posterior distribution plots, we find that the slopes and intercepts (see Figures S1 and S4) of Elbe 2002 are very close to the slope and intercept of the generalized model. Since a large sample of households in the data set suffered the 2002 event in Elbe, the generalized model parameters are strongly influenced by this region- and event-group characteristics leading to a better fit for Elbe 2002 compared to the HBM—M4.

For the 2011 event in Rhine, the localized model results in least *MAE* for out-of-sample predictions compared to the HBM—M4. One plausible reason is that the damage processes that occurred in Rhine 2011 are very different from that of the other events. Though households that suffered the 2011 event in Rhine experienced water depths comparable with the other events and had similar values of exposed buildings, the incurred damage was much lesser (Table 1). While investigating further, we also see that from the posterior distributions of parameters of Rhine 2011 from HBM—M4, generalized and localized models, the slopes and intercepts from the localized model (Figures S1 and S4) in Rhine 2011 are very different from the rest of the events. Additionally, the unavailability of empirical loss data pertaining to other events from the region hinders the modeling of the regional variability in damage processes. Though the HBM—M4 results in an overall best fit (refer to section 3.1), generalizing the damage processes (varying slope and constant intercept—M4) between Rhine 2011 and other events leads to overestimation of losses pertaining to the 2011 event in Rhine. The HBM—M4 performs better than localized and generalized models in terms of least absolute value of MBE for out-of-sample predictions.

The Bayesian model comparison through *elpd* difference aggregated for all region and event groups (Table 5 b) shows significant improvement in the prediction accuracy of the localized model over the generalized model. However, the localized model and HBM—M4 show no significant difference (*elpd* difference < 4 SE) in their out-of-sample prediction capabilities (LOO-CV). For all LOO-CV model comparisons (Table 5 b), the PSIS $\widehat{k}$ values were less than 0.7, indicating a reliable estimation of *elpd* (Vehtari et al., 2017).

The ability of the models to perform in spatiotemporal transfer scenarios is tested using out-of-group prediction accuracy. The out-of-group prediction errors from localized models pertaining to each region and event group are averaged and compared with the prediction errors of the individual hierarchical and generalized models. The out-of-group validation for held-out households from each event and region groups is performed for seven region and event groups out of nine. Since the 2010 event in Oder and 2011 event in Rhine are the only events from the regions in our data set, they are not used in temporal transfer. All the nine region and event groups are used in predicting held-out households from the regions (Out-of-group CV). The HBM—M4 performs best during spatiotemporal transfer compared to the generalized and localized models in terms of point estimate errors *MAE* and *MBE* (Tables 5a and 5b). This result agrees with the conclusions from previous studies (Cammerer et al., 2013; Jongman et al., 2012; Schröter et al., 2014; Vogel et al., 2018; Wagenaar et al., 2016) that models built using data from the respective regions representing the local characteristics result in better damage estimates compared to more generalized or transferred localized models. Similar results are seen when the *elpd* differences are estimated between the models (Table 5c). Hence, the predictive performance of the HBM—M4 is significantly higher than that of the generalized and localized models during spatiotemporal transfer.

The performance of the HBM—M4 with group-level predictors is assessed for held-out households using Out-of-group CV. The HBM—M4 with group-level predictors performs better than HBM—M4 in terms of point estimate errors and *elpd* estimates (Tables 5a–5c). Thus, introducing aggregated variables or information through expert knowledge pertaining to every region and event group as group-level predictors within the hierarchical framework helps to improve predictive capability of the HBM during spatiotemporal transfer.

4. Conclusions

A HBM is developed for capturing spatiotemporal variability in flood damage processes. Parameterization of the widely used depth-damage functions, that is, square root functions of water depth, with shared intercept and varying slope across region and event groups results in a HBM for flood loss estimation. Aggregated variables attributing to region- and event-characteristics, namely, flood experience of the households, duration of inundation, and return period of the event, are used as group-level predictors to estimate the varying slopes in the HBM and improve loss predictions for regions and events where no empirical loss data are available. Such region- and event-specific information could also be provided via expert knowledge. We tested and proved the hypothesis that, in transfer scenarios, HBMs are superior compared to localized and generalized regression models.

Additional advantages of implementing this model for flood loss estimation are the following:

1. The HBM is developed based on depth-damage functions, which can be further improved with expert region- and event-specific information that is mapped on model parameters (slope and intercept). Hence, the model development requires only object-level empirical data consisting of water depth and incurred flood losses.
2. Since the HBM is a probabilistic model, it inherently provides quantification of uncertainty in the predicted loss estimates. This is valuable for improved decision making.
3. Owing to the availability of input data (water depth), the HBM is widely applicable and will as such significantly improve flood loss modeling, particularly in spatiotemporal model transferability settings.

In this study, the development and validation of the HBM and localized and generalized regression models are performed based on empirical flood loss data from six flood events in the Elbe, Danube, Rhine, and Oder catchments in Germany. However, these models are easily scalable and might be even more valuable in international flood loss model transferability applications.

Acknowledgments
This research has received funding from the European Union's Horizon 2020 research and innovation program under Grant Agreement 676027 MSCA ETN System-Risk. Flood damage data of the 2005, 2006, 2010, 2011, and 2013 events along with instructions on how to access the data are available via the German flood damage database, HOWAS21 (http://howas21.gfz-potsdam.de/howas21/). Flood damage data of the 2002 event was partly funded by the reinsurance company Deutsche Rückversicherung (www.deutscherueck.de) and may be obtained upon request. The surveys were supported by the German Research Network Natural Disasters (German Ministry of Education and Research (BMBF), 01SFR9969/5), the MEDIS project (BMBF; 0330688) the project "Hochwasser 2013" (BMBF; 13N13017), and by a joint venture between the German Research Centre for Geosciences GFZ, the University of Potsdam, and the Deutsche Ruckversicherung AG, Dusseldorf.

References

Ahn, K.-H., Palmer, R., & Steinschneider, a S. (2017). A hierarchical Bayesian model for regionalized seasonal forecasts: Application to lowflows in the northeastern United States. *Water Resources Research*, *53*, 503–521. https://doi.org/10.1002/2016WR019605

Bubeck, P., Botzen, W. J., Kreibich, H., & Aerts, J. C. (2013). Detailed insights into the influence of flood-coping appraisals on mitigation behaviour. *Global Environmental Change*, *23*(5), 1327–1338.

Bubeck, P., & Kreibich, H. (2011). Direct costs and losses due to the disruption of production processes. ConHaz Report of WP1.

Cammerer, H., Thieken, A. H., & Lammel, J. (2013). Adaptability and transferability of flood loss functions in residential areas. *Natural Hazards and Earth System Sciences*, *13*(11), 3063–3081.

Carisi, F., Schröter, K., Domeneghetti, A., Kreibich, H., & Castellarin, A. (2018). Development and assessment of uni- and multivariable flood loss models for Emilia-Romagna (Italy). *Natural Hazards and Earth System Sciences*, *18*(7), 2057–2079. https://doi.org/10.5194/nhess-18-2057-2018

Chinh, D., Dung, N., Gain, A., & Kreibich, H. (2017). Flood loss models and risk analysis for private households in Can Tho City, Vietnam. *Water*, *9*(5), 313. https://doi.org/10.3390/w9050313

Das, P., Patskoski, J., & Sankarasubramanian, A. (2018). Modeling the irrigation withdrawals over the coterminous US using a hierarchical modeling approach. *Water Resources Research*, *54*, 3769–3787. https://doi.org/10.1029/2017WR021723

DESTATIS (Federal Statistical Office): Statistisches Jahrbuch – Deutschland und Internationales (2013). available at: https://www.destatis.de/DE/Publikationen/StatistischesJahrbuch/StatistischesJahrbuch2013.pdf (last access: 20 May 2018), 2013 (in German).

Dottori, F., Figueiredo, R., Martina, M. L. V., Molinari, D., & Scorzini, A. R. (2016). INSYDE: A synthetic, probabilistic flood damage model based on explicit cost analysis. *Natural Hazards and Earth System Sciences*, *16*(12), 2577–2591. https://doi.org/10.5194/nhess-16-2577-2016

Elmer, F., Thieken, A. H., Pech, I., & Kreibich, H. (2010). Influence of flood frequency on residential building losses. *Natural Hazards and Earth System Sciences*, *10*(10), 2145–2159.

Feller, A., & Gelman, A. (2015). Hierarchical models for causal effects. In *Emerging trends in the social and behavioral sciences: An interdisciplinary, searchable, and linkable resource*, (pp. 1–16). Hoboken, N.J.: John Wiley & Sons.

Figueiredo, R., Schröter, K., Weiss-Motz, A., Martina, M. L., & Kreibich, H. (2018). Multi-model ensembles for assessment of flood losses and associated uncertainty. *Natural Hazards and Earth System Sciences*, *18*(5), 1297–1314.

Gelman, A. (2006). Multilevel (hierarchical) modeling: What it can and cannot do. *Technometrics*, *48*(3), 432–435.

Gelman, A., Hwang, J., & Vehtari, A. (2014). Understanding predictive information criteria for Bayesian models. *Statistics and Computing*, *24*(6), 997–1016.

Gelman, A., Simpson, D., & Betancourt, M. (2017). The prior can often only be understood in the context of the likelihood. *Entropy*, *19*(10), 555.

Gerl, T., Kreibich, H., Franco, G., Marechal, D., & Schröter, K. (2016). A review of flood loss models as basis for harmonization and benchmarking. *PLoS ONE*, *11*(7), e0159791. https://doi.org/10.1371/journal.pone.0159791

Hollander, M., & Wolfe, D. A. (1973). *Nonparametric statistical analysis*, (pp. 115–120). New York: J Wiley and Sons.

Hudson, P. G. M. B., Botzen, W. J. W., Kreibich, H., Bubeck, P., & Aerts, J. C. J. H. (2014). Evaluating the effectiveness of flood damage mitigation measures by the application of propensity score matching. *Natural Hazards and Earth System Sciences*, *14*(7), 1731–1747.

Jongman, B., Kreibich, H., Apel, H., Barredo, J. I., Bates, P. D., Feyen, L., et al. (2012). Comparative flood damage model assessment: Towards a European approach. *Natural Hazards and Earth System Sciences*, *12*, 3733–3752.

Kienzler, S., Pech, I., Kreibich, H., Müller, M., & Thieken, A. H. (2015). After the extreme flood in 2002: Changes in preparedness, response and recovery of flood-affected residents in Germany between 2005 and 2011. *Natural Hazards and Earth System Sciences*, *15*(3), 505–526.

Klaus, J., Pflügner, W., Schmidtke, R. F., Wind, H., & Green, C. (1994). Models for flood hazard assessment and management. In E. C. Penning-Rowsell, & M. Fordham (Eds.), *Floods across Europe. Hazard assessment, modelling and management*, (pp. 67–106). London: Middlesex University Press.

Kreibich, H., Botto, A., Merz, B., & Schröter, K. (2017). Probabilistic, multivariable flood loss modeling on the mesoscale with BT-FLEMO. *Risk Analysis*, *37*(4), 774–787.
Kreibich, H., Di Baldassarre, G., Vorogushyn, S., Aerts, J. C., Apel, H., Aronica, G. T., et al. (2017). Adaptation to flood risk: Results of international paired flood event studies. *Earth's Future*, *5*, 953–965. https://doi.org/10.1002/2017EF000606
Kreibich, H., Seifert, I., Merz, B., & Thieken, A. H. (2010). Development of FLEMOcs—A new model for the estimation of flood losses in the commercial sector. *Hydrological Sciences Journal–Journal des Sciences Hydrologiques*, *55*(8), 1302–1314.
Kreibich, H., Seifert, I., Thieken, A. H., Lindquist, E., Wagner, K., & Merz, B. (2011). Recent changes in flood preparedness of private households and businesses in Germany. *Regional Environmental Change*, *11*(1), 59–71.
Kruschke, J. K., & Vanpaemel, W. (2015). Bayesian estimation in hierarchical models. In J. R. Busemeyer et al. (Eds.). *The Oxford Handbook of computational and mathematical psychology* (pp. 279–299). Oxford University Press.
Levy, R. (2012). Probabilistic models in the study of language. *Online Draft, Nov*.
Merz, B., Kreibich, H., & Lall, U. (2013). Multi-variate flood damage assessment: A tree-based data-mining approach. *Natural Hazards and Earth System Sciences*, *13*(1), 53–64.
Merz, B., Kreibich, H., Schwarze, R., & Thieken, A. (2010). Review article. "Assessment of economic flood damage". *Natural Hazards and Earth System Sciences*, *10*(8), 1697–1724.
Nicholas, J., Holt, G. D., & Proverbs, D. G. (2001). Towards standardising the assessment of flood damaged properties in the UK. *Structural Survey*, *19*(4), 163–172.
Parker, D. J., Green, C. H., & Thompson, P. M. (1987). *Urban flood protection benetits: A project appraisal guide*. Aldershot: Gower technical press.
Penning-Rowsell, E. (1977). *Benefits of flood alleviation*. Saxon House.
Rözer, V., Kreibich, H., Schröter, K., Müller, M., Sairam, N., Doss-Gollin, J., et al. (2019). Probabilistic models significantly reduce uncertainty in Hurricane Harvey pluvial flood loss estimates. *Earth's Future*, *7*, 384–394. https://doi.org/10.1029/2018EF001074
Sairam, N., Schröter, K., Lüdtke, S., Merz, B., & Kreibich, H. (2019). Quantifying flood vulnerability reduction via private precaution. *Earth's Future.*, *7*, 235–249. https://doi.org/10.1029/2018EF000994
Schröter, K., Kreibich, H., Vogel, K., Riggelsen, C., Scherbaum, F., & Merz, B. (2014). How useful are complex flood damage models? *Water Resources Research*, *50*, 3378–3395. https://doi.org/10.1002/2013WR014396
Schröter, K., Kunz, M., Elmer, F., Mühr, B., & Merz, B. (2015). What made the June 2013 flood in Germany an exceptional event? A hydro-meteorological evaluation. *Hydrology and Earth System Sciences*, *19*(1), 309–327. http://doi.org/10.5194/hess-19-309-2015
Smith, D. I. (1994). Flood damage estimation—A review of urban stage-damage curves and loss functions. *Water S. A.*, *20*(3), 231–238.
Sun, X., Lall, U., Merz, B., & Dung, N. V. (2015). Hierarchical Bayesian clustering for nonstationary flood frequencyanalysis: Application to trends ofannual maximum flow in Germany. *Water Resources Research*, *51*, 6586–6601. https://doi.org/10.1002/2015WR017117
Thieken, A. H., Kreibich, H., Muller, M., & Merz, B. (2007). Coping with floods: Preparedness, response and recovery of flood-affected residents in Germany in 2002. *Hydrological Sciences Journal*, *52*(5), 1016–1037.
Thieken, A. H., Müller, M., Kreibich, H., & Merz, B. (2005). Flood damage and influencing factors: New insights from the August 2002 flood in Germany. *Water Resources Research*, *41*, W12430. https://doi.org/10.1029/2005WR004177
Thieken, A. H., Olschewski, A., Kreibich, H., Kobsch, S., & Merz, B. (2008). Development and evaluation of FLEMOps—A new flood loss estimation model for the private sector. *WIT Transactions on Ecology and the Environment*, *118*, 315–324.
Vehtari, A., Gelman, A., & Gabry, J. (2017). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing*, *27*(5), 1413–1432.
Venables, W. N., & Ripley, B. D. (2002). *Modern applied statistics with S*, (4th ed.). New York: Springer.
Vogel, K., Weise, L., Schröter, K., & Thieken, A. H. (2018). Identifying driving factors in flood-damaging processes using graphical models. *Water Resources Research*, *54*, 8864–8889. https://doi.org/10.1029/2018WR022858
Wagenaar, D., de Jong, J., & Bouwer, L. M. (2017). Multi-variable flood damage modelling with limited data using supervised learning approaches. *Natural Hazards and Earth System Sciences*, *17*(9), 1683.
Wagenaar, D. J., De Bruijn, K. M., Bouwer, L. M., & Moel, H. D. (2016). Uncertainty in flood damage estimates and its potential effect on investment decisions. *Natural Hazards and Earth System Sciences*, *16*(1), 1–14.
Zhai, G., Fukuzono, T., & Ikeda, S. (2005). Modeling flood damage: Case of Tokai Flood 2000 1. *JAWRA Journal of the American Water Resources Association*, *41*(1), 77–92.