

Please cite as: Nivedita Sairam, Kai Schröter, Francesca Carisi, Dennis Wagenaar, Alessio Domeneghetti, Daniela Molinari, Fabio Brill, Sally Priest, Christophe Viavattene, Bruno Merz, Heidi Kreibich (2020) Bayesian Data-Driven approach enhances synthetic flood loss models, *Environmental Modelling & Software*, Volume 132, 2020.

The published version is available from <https://doi.org/10.1016/j.envsoft.2020.104798>

The following is a post-print version.

# Bayesian Data-Driven Approach Enhances Synthetic Flood Loss Models.

Nivedita Sairam<sup>1,2</sup>, Kai Schröter<sup>1</sup>, Francesca Carisi<sup>3</sup>, Dennis Wagenaar<sup>4</sup>, Alessio Domeneghetti<sup>3</sup>, Daniela Molinari<sup>5</sup>, Fabio Brill<sup>1,7</sup>, Sally Priest<sup>6</sup>, Christophe Viavattene<sup>6</sup>, Bruno Merz<sup>1,7</sup>, Heidi Kreibich<sup>1</sup>

<sup>1</sup>GFZ German Research Centre for Geosciences, Section 4.4. Hydrology, Potsdam, Germany.

<sup>2</sup>Geography Department, Humboldt University, Berlin, Germany.

<sup>3</sup>DICAM-Department of Civil, Chemical, Environmental and Materials Engineering, Alma Mater Studiorum - University of Bologna (Bologna, Italy)

<sup>4</sup>Deltares, Delft, The Netherlands

<sup>5</sup>Department of Civil and Environmental Engineering, Politecnico di Milano, Piazza Leonardo da Vinci 32, 20133, Milano (Italy)

<sup>6</sup>Flood Hazard Research Centre, Middlesex University, The Burroughs, Hendon, London, UK.

<sup>7</sup>Institute for Environmental Sciences and Geography, University of Potsdam, Germany

## Key Points

1. Bayesian Data-Driven approach integrates knowledge from the vast compendium of established synthetic models with empirical loss data.
2. This approach improves accuracy and quantifies reliability of synthetic flood loss models using local empirical data.
3. Continuous integration of empirical data from multiple flood events, using Bayesian Data-Driven approach improves loss predictions for a potential future event.

## Abstract

Flood loss estimation models are developed using synthetic or empirical approaches. The synthetic approach consists of what-if scenarios developed by experts. The empirical models are based on statistical analysis of empirical loss data. In this study, we propose a novel Bayesian Data-Driven approach to enhance established synthetic models using available empirical data from recorded events. For five case studies in Western Europe, the resulting Bayesian Data-Driven Synthetic (BDDS) model enhances synthetic model predictions by reducing the prediction errors and quantifying the uncertainty and reliability of loss predictions for post-event scenarios and future events. The performance of the BDDS model for a potential future event is improved by integration of empirical data once a new flood event affects the region. The BDDS model, therefore, has high potential for combining established synthetic models with local empirical loss data to provide accurate and reliable flood loss predictions for quantifying future risk.

## 1. Introduction

Due to changing climate and increased settlements and assets in the flood plains, risk to life and property due to flooding is rising (Barredo 2009, Merz et al. 2012, Domeneghetti et al. 2015). Decisions concerning Flood Risk Management (FRM) focusing on new flood defense schemes and resilience initiatives are generally based on risk assessment encompassing of future hazard scenarios and the resulting damages. Models focusing on the hazard components (hydrology and hydraulics) are constantly being developed and improved by the research community, and are outside the scope of this paper; especially, the integration of physics-based models with Machine Learning algorithms have led to the development of high-resolution hazard maps (Teng et al. 2017, da Costa et al. 2019). In

47 addition to flood hazard modelling, accounting for flood damage processes is crucial to predict losses.  
48 Flood damage processes are modelled using loss models, also called as vulnerability functions (Ward  
49 et al. 2019). Flood loss models are an essential component of the risk chain as they quantify flood risk  
50 in terms of economic losses (Merz et al., 2010). Flood loss models are generally developed using two  
51 approaches: 1. Synthetic or Engineering functions, 2. Empirical modelling. Synthetic models use expert  
52 opinions or engineering solutions that result in a set of What-If scenarios to estimate flood losses. They  
53 are not based on statistical analysis of observed data (Penning-Rowsell and Chatterton, 1977). One of  
54 the major advantages of synthetic loss models is their non-dependency on empirical data. However,  
55 the development of detailed damage scenarios covering all damage possibilities and building  
56 characteristics requires high effort (Smith, 1994). Since these models are synthesized based on a  
57 variety of data sources, such as expert knowledge and technical papers, the advantage is that these  
58 models are more generalized and lead to higher levels of standardization compared to empirical  
59 models and therefore are more suited to being used for actions that require accountability, such as  
60 investment decision-making (Smith, 1994; Merz et al. 2010; Amadio et al., 2019). For practical  
61 applications, the outputs from the synthetic models are required to capture the observed damage  
62 processes. However, except in very few models such as the INSYDE (Dottori et al. 2019), the empirical  
63 loss values do not constitute the model development.  
64

65 Empirical models are developed based on real damage information observed from past events and  
66 hence, require large amounts of high-quality detailed data on flood damages and the damage-  
67 influencing factors, such as water depth (Merz et al. 2010, Smith, 1994). These models aim to represent  
68 the relationship between flood damage and its influencing factors using patterns that occurred in the  
69 past events. The empirical models may be based on data from a single event (localized model) or  
70 cumulative data from multiple events (generalized model). Flood loss models purely based on localized  
71 empirical datasets are unable to reliably predict building damages for other events (Wagenaar et al.  
72 2018). In contrast, generalized models (e.g. Bayesian Network, multi-level parameterization) based on  
73 data from multiple events cover a wider range of damage processes and perform better for new events  
74 (Wagenaar et al. 2018, Sairam et al. 2019). As empirical models are based on real damage data, it is  
75 expected that they capture the observed damage processes and are less prone to surprises (Merz et  
76 al. 2015). However, an important disadvantage is their requirement for detailed damage surveys.  
77 These are often expensive and time consuming. Survey campaigns that are conducted after extreme  
78 events may result in a large sample of respondents that reported damage. However, in the case of  
79 surveys conducted after small localized events, the resulting datasets are often insufficient to model  
80 different damage processes.

81 Owing to lack of detailed object-level damage data, only a few studies have validated the flood loss  
82 models against observed loss estimates (Gerl et al. 2016; Amadio et al., 2019). An advantage of the  
83 empirical approach is the possibility to use a part of the empirical data for validation during model  
84 development. However, since synthetic models are generally developed when empirical data is  
85 unavailable, both calibration and validation of synthetic models remain a challenge. Both synthetic and  
86 empirical flood loss models may be deterministic or probabilistic. More than 95% of the state-of-the-  
87 art flood loss models are deterministic (Gerl et al. 2016).

88 Deterministic models result in one damage estimate based on the influencing factors. On the other  
89 hand, probabilistic models provide a distribution of losses. In reality, there exists variability in damage  
90 predictions given by the loss model based on the influencing factors. This may be due to the inherent  
91 stochastic nature of damage processes and other reasons such as uncertainty in empirical data, model  
92 structure and missing influencing factors in the model (Schröter et al. 2014, Winter et al. 2018).  
93 Decision makers and administrators are required to consider thoroughly the reliability of the flood loss  
94 models, in order to base FRM decisions and investments on the loss predictions. Hence, flood loss

95 models should provide loss predictions along with an estimate of their uncertainty and reliability. A  
96 probabilistic flood loss model estimates the probability of occurrence of all possible loss scenarios for  
97 each object and results in a distribution of predicted losses. Probabilistic models potentially account  
98 for all sources of uncertainty in model parameters, structure and variability in the modelled processes  
99 based on observed data and assumptions concerning damage processes. Hence, there is an increasing  
100 interest in developing probabilistic approaches for flood loss modelling (Schröter et al. 2014, Wagenaar  
101 et al. 2018, Rözer et al. 2019, Lüdtke et al. 2019). In the presence of large detailed empirical datasets,  
102 advanced approaches for the development of probabilistic loss models are given by Wagenaar et al.  
103 (2018) and Rözer et al. (2019). Thus, another advantage of the empirical approach is the possibility to  
104 develop probabilistic models whose reliability can be determined. Since the synthetic models are not  
105 fitted to observed losses during development, they are commonly not calibrated. Hence, it is  
106 impossible to estimate the reliability of the synthetic model without validating the model against  
107 empirical loss data (Zischg et al. 2018).

108 We propose to combine the empirical and synthetic approaches to harness advantages of both  
109 concepts. To this end, we use relevant empirical loss data for enhancing the synthetic model  
110 predictions. The objective of this study is to propose and validate a Bayesian Data-Driven approach  
111 that calibrates the predictions of existing synthetic flood loss models using relevant empirical loss data  
112 at the object-level (residential buildings), within a probabilistic framework. The resulting flood loss  
113 estimation model is a Bayesian Data-Driven Synthetic (BDDS) Model. The BDDS model associates  
114 probability distributions with synthetic model outputs and can explain variability across households  
115 due to characteristics, which are not taken into account by the synthetic loss model. The BDDS model  
116 requires a synthetic model and local empirical data to calibrate the model for that region. The synthetic  
117 model can refer to any spatial scale (regional, national, continental). The BDDS model is aimed at  
118 enhancing the synthetic loss model by providing truly probabilistic loss predictions that are sharp  
119 (narrow width of distribution of predictions), calibrated and reliable for both central values and  
120 dispersion.

121 The BDDS model is tested for improvement in predictive capability compared to the standard national  
122 synthetic model, based on case studies from four countries in Western Europe – UK, Netherlands, Italy  
123 and Germany. We develop the BDDS model for residential buildings using the loss predictions from the  
124 synthetic flood loss models and empirical loss data from one or several (if available) flood events from  
125 the specific case study regions. Moreover, the BDDS model allows integrating synthetic model  
126 predictions with a continuous collection of empirical data after each flood event, in order to enhance  
127 prediction of flood losses due to potential flood events that may occur in the future.

128 The paper is organized as follows: Section 2 explains the Methods and Data including setting up the  
129 framework for BDDS model (2.1), BDDS model construction (2.2) and metrics for assessing model  
130 performances (2.3); explanation of case studies, object-level empirical data and the synthetic models  
131 used in the study (2.4). Results including damage prediction for post-event scenarios and future events  
132 are reported and discussed in Section 3. Section 4 includes concluding points focusing on  
133 implementation of the model, scope for future work and software availability.

134

## 135 **2. Methods and Data**

### 136 **2.1. Setting up the framework for BDDS model:**

137 The BDDS model describes the relationship between empirical losses and their corresponding  
138 deterministic loss predictions from synthetic models by means of a full Bayesian approach. The  
139 parameters of the BDDS model are indicators pertaining to the deviation between the synthetic model  
140 predictions and empirical observations. Also, the full joint posterior probability distribution of the  
141 BDDS model parameters can be obtained along with the predictive distribution of flood losses given

142 the synthetic model and empirical losses from events that occurred in the region. From the credibility  
 143 intervals of the predictive distributions, it is possible to estimate the uncertainty in the flood loss  
 144 predictions.

145  
 146 The BDDS model is based on the premise that the empirical losses and synthetic loss predictions may  
 147 be seen as components of a statistical model, in which the synthetic loss predictions are considered as  
 148 exogenous variables (one that is determined outside the model, and imposed on the model) that are  
 149 used to determine the observed losses. The BDDS model estimates losses using a linear function with  
 150 empirical loss as the dependent variable regressed against the synthetic loss prediction. We assume  
 151 that the BDDS model is identifiable for households within a region: i.e., the damage processes that  
 152 occur in households belonging to one region are the same. Hence, the BDDS model assumes a single  
 153 set of parameters for each region.

154  
 155 In order to make the loss predictions comparable across the different case studies, we use relative loss  
 156 to buildings,  $rloss$ , which is the ratio of absolute building loss to its total reconstruction value in the  
 157 respective currencies, at the time of the event (Elmer et al., 2010). The rloss values are between 0 and  
 158 1, where 0 indicates no damage and 1 indicates complete damage, requiring reconstruction of the  
 159 building. The BDDS model is given in 1.

160  
 161 
$$\widetilde{rloss} | rloss_{syn} \sim Beta(\alpha, \beta) \quad \text{Equation - 1}$$
  
 162 
$$\alpha = \mu \times \varphi$$
  
 163 
$$\beta = (1 - \mu) \times \varphi$$
  
 164 
$$\mu = inv\ logit(\lambda \times rloss_{syn} + \varepsilon)$$

165 In this model definition, the observed rloss is represented as  $\widetilde{rloss}$  and the rloss predictions from  
 166 synthetic model is represented as  $rloss_{syn}$ . Since the observed losses are not included in the  
 167 synthetic model development, the BDDS model definition uses a set of parameters to alter the  
 168 synthetic model predictions to agree with the observations.  $\widetilde{rloss}$  is modelled as a beta distribution  
 169 with logit transformation, since, unbounded distributions might result in implausible values for  $\widetilde{rloss}$   
 170 (Rözer et al. 2019). The beta distribution holds two parameters  $\alpha$  and  $\beta$  which are algebraically  
 171 determined using location parameter  $\mu$  and variance parameter  $\varphi$ .  $\mu$  is a function of the synthetic rloss  
 172 predictions ( $rloss_{syn}$ ) with parameters slope ( $\lambda$ ), intercept ( $\varepsilon$ ). These parameters are estimated by  
 173 modelling the deviations of the empirical loss data from the synthetic model predictions using Markov  
 174 Chain Monte Carlo (MCMC) sampling implemented using STAN (Carpenter et al. 2017). We initially  
 175 provide priors that describe our general belief about the distribution of the parameters. For example,  
 176  $\varphi$  is required to be positive and hence given a un-informative generic prior,  $gamma(0.01, 0.01)$ . We  
 177 provide un-informative generic priors to  $\lambda$  and  $\varepsilon$  to determine the parameterization of BDDS model  
 178 based on the availability of evidence from empirical loss data. The MCMC sampling creates a large  
 179 number of replications of these parameters explaining the data generation process of flood losses. This  
 180 results in approximate posterior distributions of  $\widetilde{rloss}$ .

1812.2. BDDS model construction

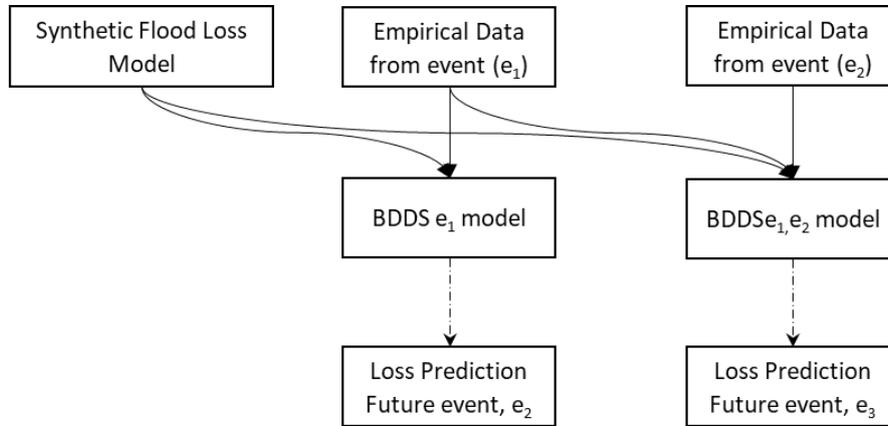
182 In reality, we are particularly interested in the capability of the BDDS model to estimate expected flood  
 183 losses to buildings after an event (post-event scenarios) or predict expected losses for a potential  
 184 future event. Therefore, we focus only on the temporal update of BDDS considering two scenarios:

- 185 1. Post-event: Comparison of a BDDS model developed using empirical data from one event against  
 186 synthetic loss predictions, for the same event using 10-fold Cross Validation (local 10-fold CV). The  
 187 empirical dataset from the event is split into 10 parts, a BDDS model is trained with 9 parts of the  
 188 dataset and validated on the left-out data (10<sup>th</sup> part). This is repeated 10 times, i.e., until all of the  
 189 dataset is validated. The model definition of the post-event scenario is given by Equation 2.

190 Future event: Comparison of a BDDS model developed using empirical data from one or more  
 191 events against synthetic loss predictions, for a future event that occurs in the same region  
 192 (Temporal one-step ahead Cross Validation; see Figure 1). Since flood damage processes are  
 193 influenced by human-flood interactions such as preparedness and land use changes (Barendrecht  
 194 et al. 2019), events occurring in the same region may show significant changes in terms of damage  
 195 processes over time. Based on empirical evidence, it is expected that exposure and vulnerability  
 196 show rather similar characteristics within one region than between regions (Schröter et al 2014,  
 197 Sairam et al 2019).

198  
 199 A BDDS model (BDDS  $e_1$ ) is developed using synthetic model and empirical flood loss data from the  
 200 first event ( $e_1$ ). This model provides calibrated probabilistic loss predictions for the future event,  
 201  $e_2$ . After the occurrence of the event  $e_2$ , a BDDS model (BDDS  $e_1, e_2$ ) is developed using the same  
 202 synthetic model and empirical loss data from both events  $e_1$  and  $e_2$ . This model results in calibrated  
 203 probabilistic loss predictions for the event  $e_3$ , which may potentially happen in the future. The  
 204 BDDS model definition of the future event scenario is given by Equation 3.

205  
 206 Synthetic models are also sometimes updated to consider significant changes in damage processes  
 207 over time. For example, in the UK, the MCM damage datasets have been incrementally updated and  
 208 improved for over 40 years. Since the MCM online publication (<https://www.mcm-online.co.uk/>) in  
 209 2013, the MCM functions are updated considering available evidences on changes in building contents  
 210 and structure as well as repair, drying and reconstruction costs and other socio-economic  
 211 determinants. For predicting damages from potential future events, the recent models are preferable.  
 212 Considering the available multi-event case studies, none of the corresponding synthetic models were  
 213 updated between the events.



214  
 215 Figure 1: Framework for Temporal one-step ahead CV using a synthetic flood loss model and  
 216 continuous collection of empirical flood loss data. The components involved in the development of  
 217 BDDS model are shown with solid lines and the predictions are shown as dot-dash lines.  
 218

219 
$$p(\overline{rloss}_{b'e'} | \overline{rloss}_{be}) = \int_{\theta} p(\overline{rloss}_{b'e'} | \theta) p(\theta | \overline{rloss}_{be}) d\theta$$
 Equation - 2

220 
$$p(\overline{rloss}_{b'e'} | \overline{rloss}_{be}) = \int_{\theta} p(\overline{rloss}_{b'e'} | \theta) p(\theta | \overline{rloss}_{be}) d\theta$$
 Equation - 3

221  
 222 The BDDS model definition for the two scenarios of CV are given in equations 2 and 3, respectively. We  
 223 are particularly interested in the posterior predictive distribution of the target variable  $\overline{rloss}$   
 224 of residential buildings  $b'$  that are not included in training the BDDS model conditioned on the observed  
 225 losses from the empirical dataset,  $\overline{rloss}_{be}$  from buildings  $b$  and events  $e$ . For the post-event damage

226 prediction, the posterior prediction consists of residential buildings that are from the same event  $e$  as  
 227 the empirical data used in the BDDS model training/calibration (Equation 2). For the future event  
 228 damage prediction, the posterior prediction of  $\widehat{rloss}$  are estimated for residential buildings from a  
 229 future event  $e'$  that was not used in the BDDS model training/calibration.  $\theta$  contains the beta model  
 230 parameters ( $\varphi$ ,  $\lambda$  and  $\varepsilon$ ) as shown in Equation 1. Hence, after specifying a prior for  $\theta$ , one finds the  
 231 posterior distribution  $p(\theta|\widehat{rloss}_{be})$ .

232

### 232.3. Metrics for assessing model performances

234 The influence of the BDDS model in enhancing synthetic flood loss models is quantified by comparing  
 235 the predictive performance of the BDDS model against the synthetic model. The predictive  
 236 performance is evaluated in terms of accuracy of the point estimate based on the median of the  
 237 predictive distribution (50<sup>th</sup> percentile of the distribution), using the Mean Absolute Error (MAE) and  
 238 Mean Bias Error (MBE); the reliability and uncertainty of the predictions are evaluated by means of the  
 239 Hit rate (HR) and Interval Score (IS) metrics (Gneiting et al. 2007). The HR represents the percentage  
 240 of predictions where the observed data falls into the 90% High Density Interval (HDI) of the prediction  
 241 ( $HDI_{90}$ ; values between the 5<sup>th</sup> and 95<sup>th</sup> percentiles of the distribution); the interval score (IS) penalizes  
 242 the mean width of the 90% HDI, if the prediction lies outside the 90% HDI.

$$243 \quad MAE = \frac{1}{n} \sum_{i=1}^n |\widehat{rloss}_i - rloss_i| \quad \text{Equation - 4}$$

$$244 \quad MBE = \frac{1}{n} \sum_{i=1}^n \widehat{rloss}_i - rloss_i \quad \text{Equation - 5}$$

$$245 \quad HR = \frac{1}{n} \sum_{i=1}^n h_i; \quad h_i = 1 \text{ if } rloss_i \in HDI_{90i}; \quad 0, \text{ otherwise} \quad \text{Equation - 6}$$

$$246 \quad IS = HDI_{90i} + \frac{1}{n} \sum_{i=1}^n \frac{2}{\beta} (\min(HDI_{90i}) - \widehat{rloss}_i) \{ \widehat{rloss}_i < \min(HDI_{90i}) \} + \frac{2}{\beta} (\widehat{rloss}_i -$$

$$247 \quad \max(HDI_{90i}) \{ \widehat{rloss}_i > \max(HDI_{90i}) \} \quad \text{Equation - 7}$$

248

249 Where  $\widehat{rloss}$  is the observed rloss from empirical dataset,  $rloss$  is the 50<sup>th</sup> percentile of the predictive  
 250 distribution and  $\beta$  scales the score based on the considered HDI;  $\beta = 1 - (0.95 - 0.05)$ , for 90% HDI.  
 251 Least MAE and least absolute value of MBE indicate the better performing model. High HR is  
 252 characteristic of reliable estimates. A smaller IS indicates narrow 90% HDI, which may be potentially  
 253 due to a larger coverage of empirical loss observations representing the damage processes. Thus, a  
 254 smaller IS indicates a sharper distribution of the predictions with higher reliability. Most synthetic  
 255 models considered in this study are deterministic and hence, do not provide a distribution of loss  
 256 predictions. Thus, only MAE and MBE can be estimated for these synthetic models. However, if  
 257 uncertainty due to stochastic processes or missing variables are considered by the synthetic model as  
 258 it is the case for INSYDE (Dottori et al. 2016), the reliability of the synthetic and DDM models can be  
 259 compared using IS and HR estimates.

260

### 261.2.4. Case studies: Synthetic models, event description and empirical data

#### 262 2.4.1. **Cumbria, United Kingdom**

##### 263 2.4.1.1. Synthetic model: Multi Coloured Manual (MCM)

264 The Multi-Coloured Manual (MCM) (Penning-Rowsell et al., 2013) was initiated in 1977 and  
 265 incrementally improved thereafter and was developed for the purpose of benefit appraisal for flood  
 266 investment. It aims to represent national economic losses in sterling. Adopting a deterministic  
 267 approach, the MCM provides a range of synthetically-generated absolute depth-damage functions for  
 268 residential and non-residential properties of different types which have been developed to provide  
 269 national consistent values. The damage functions are generated for individual inventory items and

270 building contents per social grade based on the best ownership and economic values available from  
271 market-based surveys and synthetically generated susceptibility curves. For residential properties,  
272 unique damage functions are provided according to the type and duration of flooding, warning lead  
273 time, building type, year of construction and social class; and estimates of damage are provided for  
274 the building fabric and contents and the costs of drying and cleaning. Weighted average damage  
275 function curves are then obtained for the different properties considering the national distribution of  
276 properties in flood prone areas. For comparability, we utilize MCM loss data to only the residential  
277 building fabric and divide by reconstruction cost to obtain an estimate of relative loss. Since empirical  
278 data concerning social class was not available, an initial MCM assessment for building fabric losses was  
279 performed utilizing different damage functions based on type of flooding, water depth, duration of  
280 inundation, warning lead time, building type and year of building construction.

281

#### 282 2.4.1.2. Event description and empirical data: Cumbria 2015

283 The December 2015 flood event in Cumbria (Storm Desmond) was characterized by exceptionally high  
284 rainfall, temperature and soil moisture. This is the biggest recorded flooding in Cumbria in almost all  
285 the river basins. In comparison, the meteorological winter of 2015/2016 was the wettest on record  
286 across all of the UK. The December 2015 event with a return period of 800 to 1,000 years in some parts  
287 of Cumbria broke numerous climate records resulting in extreme flooding and strong winds. This event  
288 is estimated to have caused impacts between £520 and £662 Million (Szönyi et al. 2016). In most parts  
289 of Cumbria, the flooding occurred due to overtopping of the structural protection measures such as  
290 dikes and flood walls. In Cockermouth and Keswick, the improved flood protection reduced the impacts  
291 of the 2015 event. Further information on the event can be found in Szönyi et al. (2016) and Cumbria  
292 County Council (2018). The households reported up to 3 meters of inundation depth and the duration  
293 of inundation was between a few hours to almost 48 hours in many regions.

294

295 After the 2015 event, computer-aided telephone surveys were undertaken targeting the households  
296 that suffered damage during the 2015 flooding. A list of affected streets was obtained using the flood  
297 outlines published by the Environment Agency DEFRA (Environment Agency DEFRA, 2019) and the  
298 telephone numbers of households in these streets were obtained from public telephone directory. The  
299 survey locations were mainly spread over northern UK, mainly focused on the Cumbria region covering,  
300 Appleby, Keswick, Kendal, Carlisle and Cockermouth. The survey consisted of questions concerning the  
301 hazard (water depth, duration, velocity, contamination etc.), exposure (rebuilding cost and content  
302 value), vulnerability (building type, construction year, private precautionary measures, emergency  
303 measures, warning information etc.) and incurred damage to building structure and contents. The  
304 reconstruction costs for the houses were obtained from the Association of British Insurers  
305 (<https://www.abi.org.uk/>). The households that provided water depth and building loss information  
306 from the Cumbria region were selected for this analysis. This resulted in a dataset with 33 residential  
307 buildings. All of these households provided information pertaining to the initial appraisal of the MCM.  
308 The summary statistics of the responses from the households are provided in Table 1.

#### 309 2.4.2. **Meuse, Netherlands**

##### 310 2.4.2.1. Synthetic model: SSM

311 SSM is a flood loss model developed for the Dutch national government (De Bruijn et al., 2014). It is  
312 the standard model applied in all Dutch flood risk management studies for the national government.  
313 It is an update of an earlier model called Standard Damage and Fatality assessment model (HIS-SSM)  
314 (Kok et al., 2005). The damage function applied in this paper, for residential structural damage was  
315 first proposed in Duiser (1982). This damage function is based on a combination of information  
316 synthesized from empirical observations concerning flood damages from three events: the coastal  
317 floods in Zeeland in 1953, the Wieringermeer flood of 1945 from a large lake and a flood in Tuindorp-

318 Oostzaan in 1960 (canal dike breach), interviews from experts and damage functions from Penning-  
319 Rowsell et al. (1977).

#### 320 2.4.2.2. Event description and empirical data: Meuse 1993

321 This dataset is based on the 1993 flood of the Meuse River in the Dutch province of Limburg. It has  
322 been described in WL Delft (1994), Wind et al. (1999) and Wagenaar et al. (2017). The 1993 Meuse  
323 discharge was 3,120 m<sup>3</sup>/s, the highest recorded up to that point. 8% of the province was flooded  
324 causing about 180 Million Euro damage (price level 2016) (Wagenaar et al., 2017). Unlike most of the  
325 rest of Dutch rivers, in 1993 the Meuse River didn't have dikes yet in Limburg.

326 The data was collected to compensate affected households. Every flooded building was visited,  
327 resulting in a complete data set of 5,780 records. The data collection was carried out by insurance  
328 experts who visited the affected buildings weeks after the flood, often before restoration activities  
329 were completed. The experts also recorded the water depth in the buildings but this wasn't their  
330 primary objective and was sometimes difficult to assess because the flood had happened weeks prior.  
331 In Wagenaar et al. (2018) the recorded flood losses have been transferred to relative losses. The  
332 summary statistics of the survey responses are given in Table 1.

#### 333 2.4.3. **Adda, Caldogno and Secchia, Northern Italy**

##### 334 2.4.3.1. Synthetic model: INSYDE (Dottori et al, 2016)

335 INSYDE is an expert-based synthetic model, developed for the Italian context. The model is based on a  
336 what-if analysis, consisting in a virtual step-by-step inundation of a residential building and in the  
337 evaluation of the corresponding physical and monetary damage as a function of hazard and building  
338 characteristics. A mathematical function describes the damage mechanisms for each building  
339 subcomponent (walls, doors, etc.), and the associated cost for reparation, removal, and replacement;  
340 when the influence of hazard and building variables cannot be determined a priori, damage  
341 mechanisms are modelled using a probabilistic approach. In total, INSYDE adopts 23 input variables,  
342 six describing the flood event and 17 referring to building features. However, the model can be also  
343 applied when the available knowledge of the flood event and building characteristics is incomplete,  
344 given the possibility of automatically considering default values for unknown parameters and of  
345 expressing some of the variables as functions of other ones. The model supplies damage in absolute  
346 terms but an estimation of relative damage can be obtained.

347

##### 348 2.4.3.2. Event descriptions and empirical data: Adda 2002, Caldogno 2010, Secchia 2014

349 In this case study three flood events in the Po valley in Northern Italy are considered. The first one  
350 happened in November 2002 in the town of Lodi. The flood resulted from a most critical combination  
351 of events for the lower part of the Adda river, namely the simultaneous increase of the discharges from  
352 the Como lake and of the Brembo river, that is the largest tributary of the Adda upstream of Lodi.  
353 Between the 25<sup>th</sup> and 26<sup>th</sup> of November, the Adda reached the hydrometric height of 3.43 m above  
354 the reference level (68.28 m a.s.l.), corresponding to a discharge between 1,800 and 2,000 m<sup>3</sup>/s. The  
355 return period has been estimated as 100-200 years. Large portions of the town were flooded with  
356 water levels above 2 m in some neighbourhoods. The second flood event happened in the Veneto  
357 region, where from the 31<sup>st</sup> of October to the 2<sup>nd</sup> of November 2010, persistent rainfall affected the  
358 pre-Alpine and foothill areas, with peaks of more than 500 mm in some locations (ARPAV, 2010).  
359 Consequently, about 140 km<sup>2</sup> of land was inundated, involving 130 municipalities, some of which were  
360 particularly negatively affected. The situation of Bacchiglione River and its tributaries was especially  
361 critical, where hydrometric levels overcame historical records (water velocities in the river higher than  
362 330m<sup>3</sup>/s were registered; see Belcaro et al., 2011), causing the opening of a breach on the right levee  
363 of the river on the morning of the 1<sup>st</sup> of November. The countryside and the settlements of Caldogno,

364 Cresole and Rettorgole were flooded with an average water depth of 0.5 m (ARPAV, 2010) for about  
365 48 hours. The total damage, including residential properties, economic activities, agriculture and public  
366 infrastructures, was estimated to be about EUR 26 million, of which EUR 7.5 million relate to the  
367 residential sector (Scorzini and Frank, 2017). Finally, the last event happened in January 2014 in the  
368 central area of the Emilia–Romagna region (Modena province), where in the early morning of the 19<sup>th</sup>  
369 of January the water started to overtop the right levee of the Secchia River, flooding the countryside.  
370 The breach was not caused by an extreme river discharge (the return period of the event was estimated  
371 around 5 years), but by the collapse of the river embankment, weakened by animal burrows (D’Alpaos  
372 et al., 2014). Seven municipalities were affected with an inundated area of around 52 km<sup>2</sup> with the  
373 small towns of Bastiglia and Bomporto suffering the largest impacts remaining flooded for more than  
374 48 h. The total volume of overflowing water was estimated about 36x10<sup>6</sup> m<sup>3</sup>, with an average water  
375 depth of 1 m (D’Alpaos et al., 2014). The economic cost inflicted on residential properties, according  
376 to damage declaration, amounted to EUR 36 million.

377 After the three floods, public funding was made available by the national Civil Protection Authority. In  
378 order to be reimbursed, with similar procedures for all inundation events, citizens were requested to  
379 fill in pre-filled claim forms; the latter were then mostly collected by the affected municipalities and,  
380 in a small part, by the Regional Authorities. In total, our dataset includes 1,158 buildings in the flooded  
381 areas (Amadio et al. 2019). They include information on the owner, the address of the flooded building,  
382 its typology (e.g. apartment, single house), the number of affected floors, a description of the physical  
383 damage and its translation into monetary terms (distinguishing for the different rooms among damage  
384 to walls, windows and doors, floor and content). More information about the individual flood events,  
385 their hydrodynamic simulations and the data collection campaigns were published in Scorzini et al.  
386 (2018), Molinari et al. (2020), Scorzini and Frank (2017), Carisi et al (2018), Amadio et al. (2019).

387 The areas flooded in the three cases are characterized by similar exposure characteristics and  
388 economic well-being (Amadio et al. 2019). Previous studies compared the same cases and the findings  
389 sustain the opportunity to merge the dataset (Amadio et al. 2019). Hence, the three events are  
390 combined into one case study. The summary of empirical data from this case study is provided in Table  
391 1.

#### 392 2.4.4. Danube, Germany

##### 393 2.4.4.1. Synthetic model: Rhine Atlas Model (RAM) (ICPR, 2001)

394 The Rhine Atlas Model (RAM) was developed in 2001 in order to determine the regions with high flood  
395 risk in the Rhine catchment based on the 1995 floods and develop risk management strategies (ICPR,  
396 2001). Since, the RAM is intended for the Rhine catchment, an inherent transfer scenario exists when  
397 the RAM is generalized to the other catchments within Germany. However, given that a number of  
398 studies consider RAM as a standard synthetic flood loss model (Jongman et al. 2012), we use the model  
399 as the standard synthetic flood loss model for Germany. The RAM is mostly based on expert judgment  
400 as well as some information based on the HOWAS empirical flood damage data (Buck & Merkel. 1999).  
401 It is a stage-damage function using water depth as the only predictor. The RAM loss prediction is based  
402 on the resolution of land-use classes similar to that of the CORINE land use data (Jongman et al. 2012).  
403 We apply the stage-damage function corresponding to losses to building structure in the residential  
404 land-use class to estimate flood loss for each residential building.

##### 406 2.4.4.2. Event descriptions and empirical data: Danube 2002-2013

407 In this case study, three flood events that occurred between 2002 and 2013 in the Danube catchment  
408 is considered. Among the events, the 2013 flood was quite extreme with return period up to greater  
409 than 1000 years in some parts of the catchment. These were summer floods caused due to heavy  
410 rainfall resulting in surface water flooding and flash floods (Vogel et al. 2018). The 2013 floods were

411 characterized by high antecedent soil moisture combined with heavy precipitation resulting in large  
412 spatial extent of flood peaks with high magnitudes resulting in the most severe flooding in Germany  
413 over the past 6 decades (Merz et al., 2014, Schröter et al. 2015). Another distinguishing feature is the  
414 occurrence of dike breaches during the Danube 2013 event. Many properties were affected after dike  
415 breaches (e.g. at Deggendorf).

416

417 After these events, computer-aided cross-sectional telephone surveys of private households that had  
418 suffered from losses were undertaken using a standardized questionnaire. A list of affected streets was  
419 obtained using the flood masks derived from satellite data, (DLR, Center for Satellite Based Crisis  
420 information, <https://www.zki.dlr.de/>), and the telephone numbers of households in these streets were  
421 obtained from public telephone directory. The survey campaigns always focused on a single event.  
422 Depth of water within the house is determined using the reported water level in the highest affected  
423 storey by applying corrections based on the presence of a basement and height of the ground floor.  
424 Building reconstruction costs are adjusted for inflation to values as of 2013 using the building price  
425 index (DESTATIS, 2013). We consider all datasets which refer to households with basement (for  
426 unbiased measurements of water depth) and for which information on water depth and relative  
427 building loss were provided. Hence, the empirical data used in this study consists of 408 buildings from  
428 three events in the Danube catchment, that have a considerable number of completed surveys (sample  
429 size>25). The summary of empirical data from this case study is provided in Table 1.

430

#### 431 2.4.5. Elbe, Germany

##### 432 2.4.5.1. Synthetic model: Rhine Atlas Model (RAM) (ICPR, 2001)

433 The Rhine Atlas Model (RAM), described in section 2.4.4.1 is implemented for estimating losses in the  
434 Elbe catchment.

##### 435 2.4.5.2. Event descriptions and empirical data: Elbe 2002-2013

436 In the Elbe catchment, the 2002 and 2013 events were extreme with return periods greater than 100  
437 years. These events affected a large number of households. The 2002 event was characterized by a  
438 large number of dike breaches affecting households with low preparedness. However, after the 2002  
439 event, preparedness increased among households via implementation of private precautionary  
440 measures and emergency measures. Hence, a reduction in average losses is observed after the 2002  
441 event in the Elbe catchment. The other flood events (2006 and 2011) were smaller with return periods  
442 less than 50 years. They were caused due to rain-on-snow after the winter periods (Vogel et al. 2018).

443

444 Empirical damage data was collected from the affected households in the Elbe catchment during the  
445 same survey campaigns, explained in section 2.4.4.2. The study uses four events comprising of a total  
446 of 1,110 households, that provided information on water depth and relative building loss and have a  
447 considerable number of completed surveys (sample size>25). The summary of empirical data from this  
448 case study is provided in Table 1. More information about the individual flood events in the Elbe and  
449 Danube, the surveys and their results were published in Thielen et al. (2007), Kreibich et al. (2011,  
450 2017), Kienzler et al. (2015) and Vogel et al. (2018).

451

452 In this study, the Danube and Elbe catchments are considered as different case studies due to their  
453 strikingly different socio-economic and exposure characteristics which affect flood damage processes  
454 (Thielen et al. 2007). These regional differences have historical roots since the Danube catchment  
455 belonged to former West Germany and the Elbe catchment to the former East.

456

457 Table 1: Sample size, the summary (average) of water depth (wd) in meters, exposed building value  
458 (bv in EUR) , absolute and relative losses to residential buildings (bloss in EUR, rloss) for the five case  
459 studies.

Case study	Event	Sample size	wd	bv <sup>1</sup>	bloss <sup>1</sup>	rloss
Cumbria, United Kingdom (UK)	Cumbria 2015	33	0.6	390,320 <sup>2</sup>	32,640 <sup>2</sup>	0.08
Meuse, Netherlands (NL)	Meuse 1993	5780	0.4	138,000	4,307	0.03
Northern Italy (IT)	Adda 2002	270	0.9	197,356	10,592	0.05
	Caldogno 2010	294	0.4	268,175	18,398	0.07
	Secchia 2014	594	1.0	229,670	22,832	0.10
Danube, Germany (DE)	Danube 2002	225	1.7	360,107	6,352	0.02
	Danube 2005	104	2.0	412,102	7,992	0.02
	Danube 2013	79	3.0	580,109	45,675	0.08
Elbe, Germany (DE)	Elbe 2002	518	3.5	306,535	44,462	0.14
	Elbe 2006	42	2.9	312,417	7,066	0.02
	Elbe 2011	58	2.7	482,588	9,277	0.02
	Elbe 2013	492	2.7	434,095	23,599	0.05
Total		8489				

460

461 **Note:** <sup>1</sup> Values in € adjusted for inflation to values as of 2015; <sup>2</sup> Values in £ converted to € using  
462 conversion rate 1€ = 0.73£.

463

### 464 3. Results and Discussion - Comparison of predictions from synthetic loss models and BDDS models

465 The performance of the BDDS model is compared with the synthetic models from the respective  
466 regions. Since the development of BDDS models requires empirical data, the model is independently  
467 trained for each of the local 10-fold CV as well as temporal one-step-ahead CV and is validated on the  
468 left-out dataset. During both validation scenarios, there are no variations in definition and  
469 parameterization of the synthetic models. Point estimates are assessed via MAE and MBE and  
470 prediction uncertainty and reliability via IS and HR (section 2.3). Reliability and uncertainty of loss  
471 predictions are provided by all BDDS models, representing an enhancement over the deterministic  
472 synthetic models (4 out of 5 models). Among the synthetic models, INSYDE is the only synthetic model  
473 that provides distribution of loss estimates from which IS and HR can be determined. The model  
474 validation is performed by bootstrap sampling of the synthetic and BDDS model predictions with 1,000  
475 iterations with replacement, while preserving the sample size of the empirical data during each  
476 iteration.

477

#### 478 3.1. Local 10-fold CV

479 We perform a local 10-fold CV in order to validate the BDDS model predictions against the synthetic  
480 model predictions for the post-event scenario. The case studies with no empirical data from the region  
481 prior to the event are used for local 10-fold CV. This scenario (Equation 2) is applicable for the Cumbria  
482 2015, Meuse 1993, Adda 2002, Danube 2002 and Elbe 2002 flood events. These events are either the  
483 only available empirical data from the respective regions or the first event of the continuous empirical

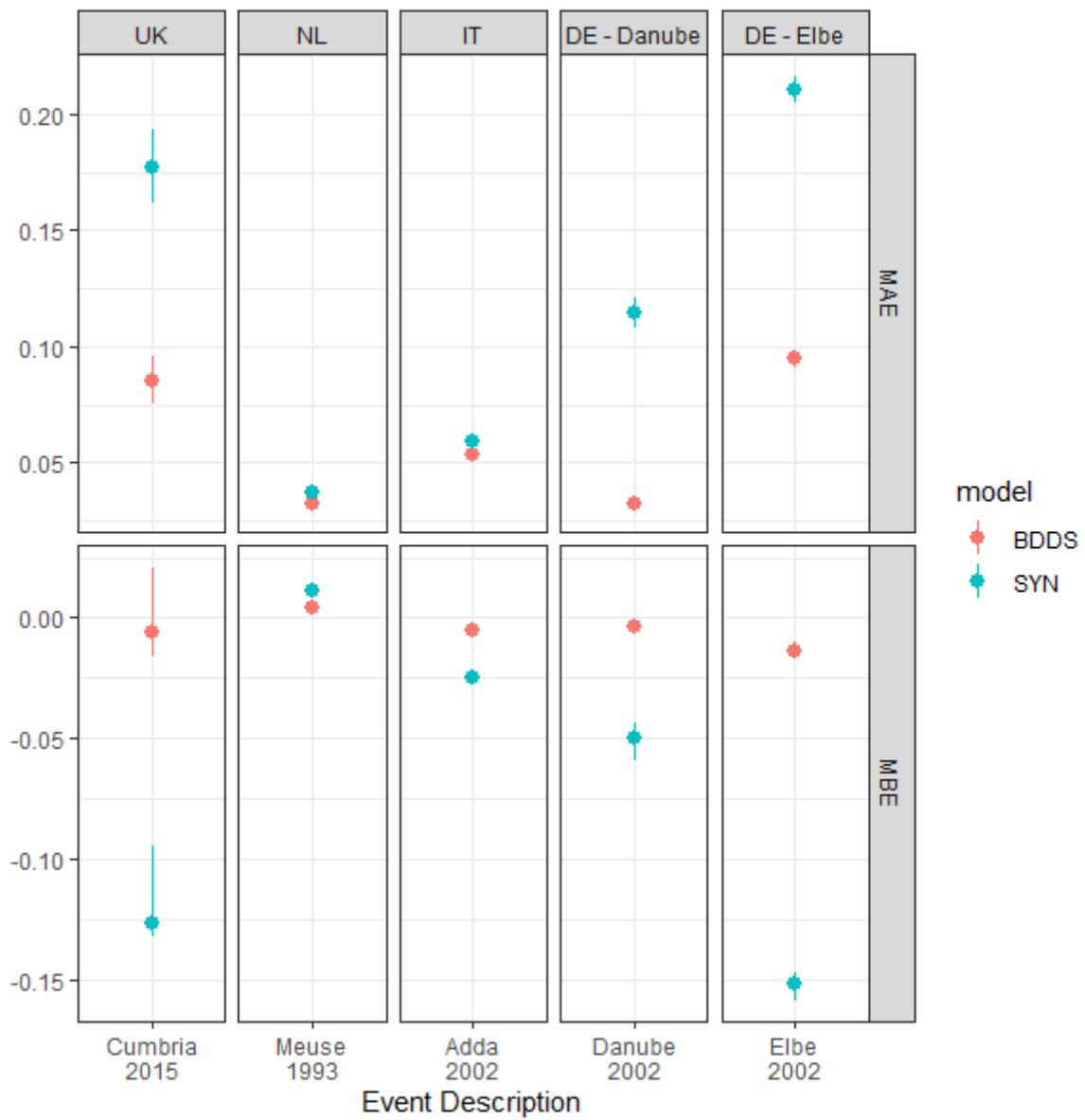
484 data collection campaigns. All synthetic models, except SSM, result in a negative MBE which indicates  
485 that on average, all these synthetic models over-estimate the building losses (see Figure 2a).

486

487 The prediction performance of the BDDS model with one event is compared against the performance  
488 of the synthetic models from the corresponding countries (Figure 2a). The BDDS model performs better  
489 than the synthetic model in terms of point estimates. As described in Equation 6, during the local 10-  
490 fold CV, the model is iteratively validated on residential buildings that are not used in the model  
491 development. Thus, the local 10-fold CV evaluates out-of-sample model performance of the BDDS  
492 model. The BDDS model with RAM and empirical data from the Elbe 2002 event results in the highest  
493 improvement in predictive performance in terms of MAE and MBE. Small improvement in predictive  
494 performance is exhibited by the BDDS models - SSM and empirical data from Meuse 1993 event and  
495 INSYDE with empirical data from the Adda 2002 event. However, among the tested synthetic models,  
496 the INSYDE and SSM models result in the smallest errors in the 10-fold CV. Among the two catchments  
497 in Germany, the RAM results in larger errors predicting losses for the Elbe 2002 event compared to the  
498 Danube 2002 event. The BDDS model consistently improves the predictions for the 2002 event in both  
499 catchments.

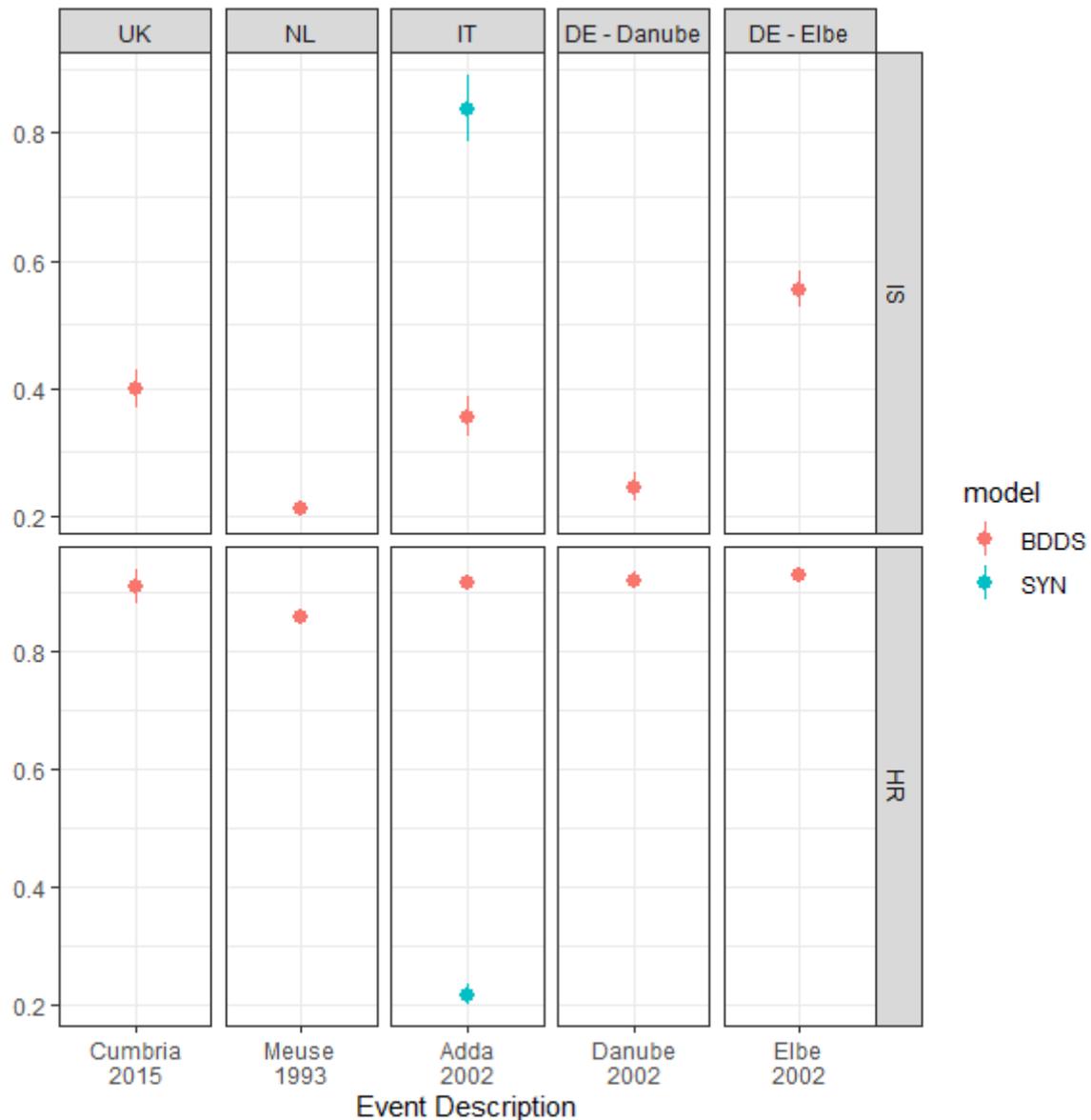
500

501 The uncertainty and reliability of the loss predictions is quantified using the IS and HR metrics. For the  
502 Adda 2002 event, the IS (HR) of the predictions from the INSYDE model is high (low) compared to the  
503 corresponding BDDS model. Hence, integrating empirical data with the INSYDE model using BDDS  
504 model reduces uncertainty and improves the reliability. The predictions from BDDS model with SSM  
505 and empirical data from the Meuse 1993 event have the least IS which represents a narrow prediction  
506 interval/ $HDI_{90}$ . The predictions from BDDS model with RAM and empirical data from Elbe 2002 event  
507 results in the highest HR with approximately 93% of the empirical loss data lying within the  $HDI_{90}$  of  
508 the predictions, representing high model reliability. However, the IS of these predictions is also high  
509 suggesting a large uncertainty. The predictions from BDDS model with empirical data from Danube  
510 2002 event show low IS and high HR representing a good balance between reliability and uncertainty.  
511 The  $HDI_{90}$  is narrow for these predictions and also a large percentage (92%) of the observed losses is  
512 captured within the  $HDI_{90}$  of the predictions.



(a)

513  
514



(b)

Figure 2 Model performances for local 10-fold CV using events and their corresponding synthetic loss models (shown in brackets) — Cumbria 2015 (MCM), Meuse 1993 (SSM), Adda 2022 (INSYDE), Danube 2022 (RAM) and Elbe 2022 (RAM). (a) MAE and MBE of flood loss predictions using synthetic models and BDDS models (b) IS and HR of loss predictions using BDDS models.

515  
516  
517  
518  
519  
520  
521

Among the tested synthetic models, the SSM and INSYDE models result in the least errors (see, Figure 2a). These models were developed after the occurrence of the respective events and may potentially capture flood damage processes based on recent events, which are comparable with the tested events. This may explain the better fit compared to the other models. Another plausible reason for the small errors from the SSM model is that the Meuse 1993 event resulted in small damage values (Table 1). This may lead to smaller errors in terms of MAE and MBE (Wagenaar et al. 2018). From the bootstrap iterations of MAE and MBE, the spread of the errors from the Cumbria 2015 event is the largest. This can be attributed to the low coverage (small sample) of empirical data from the Cumbria 2015 event. However, despite the limited availability of empirical data, the BDDS model enhances loss predictions from the MCM as well. The BDDS model reduces errors and provides predictive distributions indicating uncertainty and reliability of the predictions. In the case of Elbe 2022, the hit rate of the BDDS model is high and comparable with the performance of other BDDS models. However, the high IS indicates

533

534 that the loss distributions are not sharp. This high uncertainty may be attributed to variability in  
535 damage processes that are not adequately captured by the variables in the RAM (i.e. water depth  
536 only). This quantification of uncertainty and reliability from BDDS model is an enhancement over the  
537 established synthetic models, which is crucial for risk-based decision making (Polasky et al. 2011).

538

### 539 **3.2. Temporal One-step ahead CV**

540 In regions where, continuous empirical flood damage data is available, the predictions from synthetic  
541 models and BDDS models are compared using temporal one-step ahead CV. The losses suffered by  
542 residential buildings due to an event in the future is predicted from a BDDS model developed using the  
543 synthetic model and all available empirical data from the past events (Figure 1 and Equation 3). From  
544 our case studies, empirical damage data from northern Italy and Germany can be used to implement  
545 temporal one-step ahead CV.

546

547 Since we have empirical data from three events from Northern Italy, two BDDS models are developed,  
548 i.e. to predict losses from Caldogno 2010, the BDDS model is developed using INSYDE model and  
549 empirical data from Adda 2002, and to predict losses from Secchia 2014, the BDDS model is based on  
550 INSYDE model and empirical data from Adda 2002 and Caldogno 2010. Five BDDS models are  
551 developed for Germany using the RAM and empirical data from the past events to predict future losses.  
552 In the Danube catchment, to predict losses from the 2005 (2013) event, a BDDS model is developed  
553 using RAM and empirical data from 2002 (2002 and 2005). In the Elbe catchment, to predict losses  
554 from the 2006 (2011 / 2013) event, a BDDS model is developed using RAM and empirical data from  
555 2002 (2002 and 2006/ 2002, 2006 and 2011).

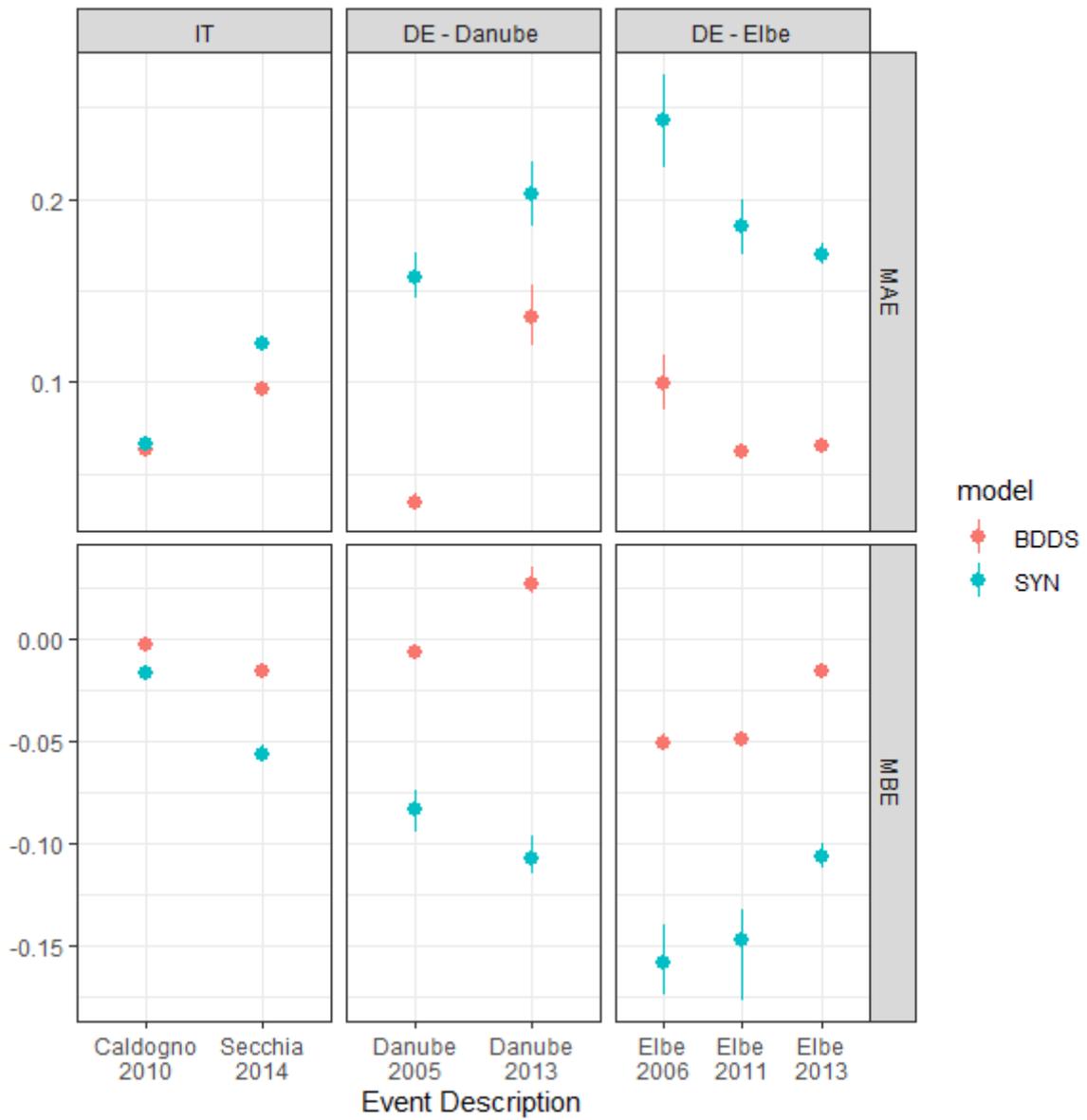
556

557 The results of the temporal one-step ahead CV are provided in Figure 3a. For all the case studies, the  
558 errors (MAE and MBE) from the BDDS model temporal one-step ahead prediction are smaller than the  
559 errors from the corresponding synthetic models. The results show that compared to the INSYDE model,  
560 the performance of the INSYDE model continuously integrated with empirical data from more events  
561 is higher. For the Elbe catchment, the BDDS model's improvement in predictive performance is  
562 observed for all future event predictions when integrated with a continuous collection of empirical  
563 data. These results suggest that, in these two regions, parameterizing the BDDS model with empirical  
564 data from events in the recent past improves the damage prediction for following events.

565

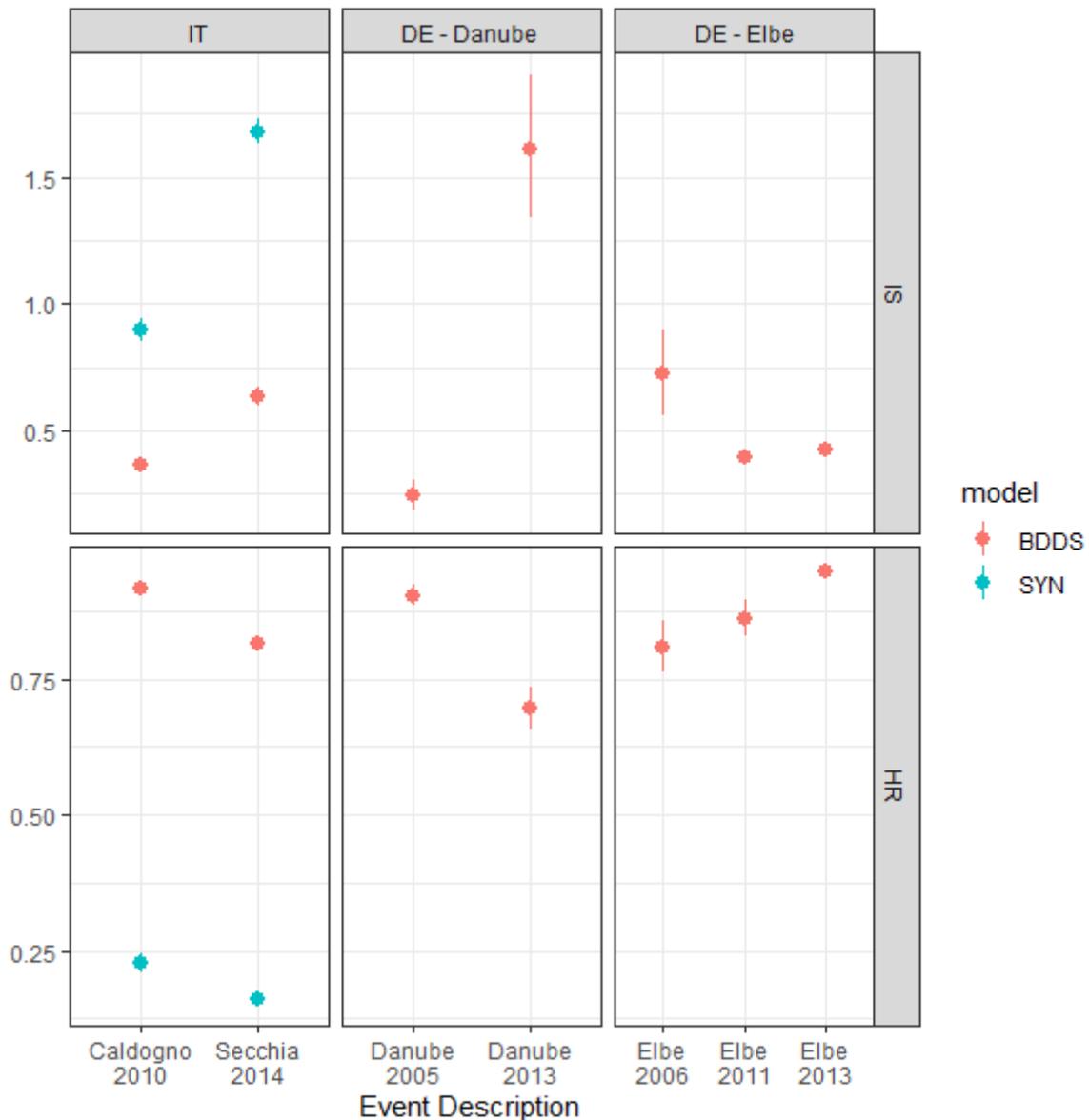
566 In the Danube catchment in Germany, the BDDS model outperforms the RAM for temporal one-step  
567 ahead predictions. However, the BDDS model shows a lower performance when data from an  
568 additional event is integrated. We also notice a change from negative to positive bias. This suggests  
569 that in the case of Danube 2013 event, the BDDS model developed by integrating RAM with empirical  
570 data from 2002 and 2005 events under-estimates the losses. The uncertainty and reliability estimates,  
571 i.e. IS and HR, from BDDS model one-step ahead temporal predictions are shown in Figure 3b. The two  
572 BDDS models developed for the case study in Northern Italy result in better HR and IS estimates  
573 compared with the INSYDE model. The BDDS model shows best reliability and least uncertainty for the  
574 Elbe 2013 event with a HR close to 100% and a relatively small IS, suggesting small uncertainty. On the  
575 other hand, loss predictions for the 2013 event in the Danube catchment from the BDDS model  
576 performs the worst with the least HR of 70% and a high IS, suggesting low reliability and large  
577 uncertainty.

578



(a)

579  
580



(b)

Figure 3: Model performances for temporal one-step ahead CV of events using empirical data from past events and their corresponding synthetic loss models (shown in brackets) — Caldogno 2010 (Adda 2002; INSYDE), Secchia 2014(Adda 2002, Caldogno 2010; INSYDE), Danube 2005 (Danube 2002; RAM), Danube 2013 (Danube 2002, 2005; RAM), Elbe 2006 (Elbe 2002; RAM), Elbe 2011 (Elbe 2002, 2006; RAM), Elbe 2011 (Elbe 2002, 2006, 2011; RAM). (a) MAE and MBE of flood loss predictions using synthetic models (SYN) and BDDS models (b) IS and HR of loss predictions using BDDS models.

581  
582  
583  
584  
585  
586  
587  
588  
589

During temporal one-step ahead CV, the BDDS model shows an overall improvement over the synthetic models. In the case of Danube 2013, integrating the RAM with Danube 2002 and 2005 events result in high IS and low HR (Figure 3b). This effect is also in agreement with the inferences from MBE for Danube 2013 estimated from the same model (Figure 3a). For all temporal one-step ahead CV cases, the synthetic models over-estimate the losses. However, when enhanced with empirical data from past events using BDDS model, the MBE is shifted towards zero. In the case of Danube 2013, the empirical data from past events reduces the overall bias, but leads to an underestimation of losses. This effect may result from some characteristics of the Danube 2013 event that differ from the other Danube events. For example, dike breaches that occurred during the Danube 2013 event inundated properties that were located away from the river with high water depths. These households had low

599

600 flood experience and were not prepared for flooding. Hence, high intensity flooding combined with  
601 low preparedness resulted in large damages (e.g. oil contamination from heating systems). Such  
602 effects are not sufficiently captured either by the uni-variable RAM or the empirical data from past  
603 events. Hence, it is important to evaluate if the empirical data is representative of the target event's  
604 damage processes. One example is the implementation of ensemble models based on the individual  
605 model characteristics and target case study (Figueiredo et al. 2018). A potential approach to capture  
606 the difference in damage processes between events is to introduce a multi-level model that allows  
607 both shared and separate parameters representing the similarities and differences between the  
608 damage processes exhibited by the different events (Sairam et al. 2019). The criteria for similarities in  
609 damage processes used by these studies were established on the basis of expert knowledge. To reduce  
610 the subjectivity in choice of models and relevance of empirical data, standardization of data for flood  
611 loss estimation along with a rigorous benchmarking of the loss models are important next steps.

612  
613 In order to interpret the importance of local empirical data, we discuss the performances of the BDDS  
614 model that is built with empirical data from the same event (local 10-fold CV) and past events  
615 (temporal one-step ahead CV). Local empirical data from the same event improves the overall  
616 reliability of the BDDS model and also results in low uncertainty, i.e. reduces IS and increases HR  
617 (Figures 2b and 3b). Hence, the use of empirical data from the same event is useful for post-event risk  
618 analysis and damage estimation. For risk-based decision making for future scenarios, we need accurate  
619 and reliable models, which can only be validated using empirical data from past events. Therefore, the  
620 IS and HR estimates obtained from the temporal one-step ahead loss predictions are more relevant.  
621 These metrics can be considered by decision makers and flood risk managers as the estimates of  
622 uncertainty and reliability of the damage model for future flood risk portfolios. In general, the BDDS  
623 model enhances synthetic models using local empirical data.

624

#### 625 **4. Conclusions**

626 Synthetic models are based on what-if analyses and are hardly validated and compared with  
627 observations. Models purely developed using empirical data require large samples of detailed object-  
628 level damage data, preferably from various events. By the presented approach it becomes possible to  
629 use the vast compendium of established synthetic damage functions in a harmonized probabilistic  
630 framework in order to improve damage estimation and quantify the reliability of the model  
631 predictions. We calibrate the synthetic models with local empirical damage data, for which not as many  
632 observations are necessary as for the development of empirical damage models.

633 We have performed 10-fold and temporal one-step ahead Cross Validation (CV) for assessing the  
634 model performances for post-event and future event scenarios, respectively. Some empirical damage  
635 data from the event is used in model training for 10-fold CV. Whereas, only empirical damage data  
636 from past events are used for model training for temporal one-step ahead CV. Our validation results  
637 show that empirical loss data from past events are valuable for enhancing the synthetic models to  
638 predict damage more accurately. From the tested case studies, on average, a reduction of 50% (51%)  
639 and 88% (74%) in mean absolute error and mean bias error were achieved by BDDS model for the  
640 post(future)-event scenarios, respectively. In respect to reliability, average hit rates of 90% and 85%  
641 were achieved for post and future event scenarios, respectively. Hence, for improving estimates of  
642 future risk, empirical data collection campaigns after flood events are crucial. However, the loss  
643 predictions from the post-event scenario show higher reliability compared to the future risk  
644 predictions. This suggests that flood damage processes vary across events and therefore dynamic  
645 damage models are required to capture this variability. Within the scope of this study, the models are  
646 not tested for regional (cross-country) transferability. This is considered as a follow-up research work  
647 for the future.

648 An important feature of the presented approach is the uncertainty quantification of the damage  
649 estimate, since this provides valuable information for improved decision making. In order to train a  
650 BDDS model for a new case study, availability of empirical damage data from past event(s) and ability  
651 to run the national standard synthetic loss model for the same event(s) are required. From the  
652 modelling perspective, knowledge concerning formulating regression equations in R (R Core Team,  
653 2019), interpretation of regression coefficients and understating probability distributions may help in  
654 customizing the presented model structure and parameter definitions, if needed. With respect to  
655 model application, no special skills are needed to use a trained BDDS model. The input data required  
656 to run the BDDS model are the same as that of the national standard synthetic model. The running  
657 time of the BDDS model is comparable to the national standard synthetic models for the samples in  
658 the tested case studies. Thus, the Bayesian Data-Driven approach is valuable for flood risk managers.

## 659 **Acknowledgements**

660 This research has received funding from the European Union's Horizon 2020 research and innovation  
661 program under Grant Agreement 676027 MSCA ETN System-Risk, as well as from the project DECIDER  
662 (BMBF, 01LZ1703G).

663

## 664 **Data and Software Availability**

665 The Multi Coloured Manual (MCM) database handbook is published by the Flood Hazard Research  
666 Centre (FHRC) at MiddleSex Univerisity, London, UK. The functions are proprietary and not publicly  
667 accessible. The SSM model is available from de Bruijn et al. (2014). INSYDE functions are available for  
668 download as R open source code, currently hosted on GitHub (<https://github.com/ruipcfig/insyde/>).  
669 The Rhine Atlas Model (RAM) is available from ICPR (2001).

670

671 The data implemented in the Cumbria 2015 case study is currently not publicly accessible. The dataset  
672 may be obtained upon request. The data used in the Meuse 1993 case study is available from  
673 Wagenaar et al. 2017. The dataset used in the Northern Italy case study are not publicly accessible.  
674 The first reason behind this is that some data come from private sources (i.e., businesses, utilities  
675 companies) that agreed on sharing their data only for research objectives, including sensitive  
676 information. The dataset may be obtained upon request. For the Danube and Elbe case studies, flood  
677 damage data of the 2005, 2006, 2010, 2011, and 2013 events along with instructions on how to access  
678 the data are available via the German flood damage database, HOWAS21 ([http://howas21.gfz-  
679 potsdam.de/howas21/](http://howas21.gfz-potsdam.de/howas21/)). Flood damage data of the 2002 event was partly funded by the reinsurance  
680 company Deutsche Rückversicherung ([www.deutscherueck.de](http://www.deutscherueck.de)) and may be obtained upon request.  
681 The surveys were supported by the German Research Network Natural Disasters (German Ministry of  
682 Education and Research (BMBF), 01SFR9969/5), the MEDIS project (BMBF; 0330688) the project  
683 "Hochwasser 2013" (BMBF; 13N13017), and by a joint venture between the German Research Centre  
684 for Geosciences GFZ, the University of Potsdam, and the Deutsche Ruckversicherung AG, Dusseldorf.

685

686 The models presented in this paper are implemented in the stan modeling language (Carpenter et al.,  
687 2017) using the brms package version 3.3.2 (Bürkner, 2017) in R (R Core Team, 2019).

688

## 689 **References**

- 690 1. Amadio, M., Scorzini, A. R., Carisi, F., Essenfelder, A. H., Domeneghetti, A., Mysiak, J., &  
691 Castellarin, A. (2019). Testing empirical and synthetic flood damage models: the case of  
692 Italy. *Natural Hazards and Earth System Sciences*, 19(3), 661-678.

- 693 2. ARPAV: Scheda Evento "Idro" 31 Ottobre–5 Novembre 2010, available  
694 at: <http://www.regione.veneto.it/> (last access: 21 March 2019), 2010.
- 695 3. Barendrecht, M. H., Viglione, A., Kreibich, H., Merz, B., Vorogushyn, S., & Blöschl, G. (2019). The  
696 value of empirical data for estimating the parameters of a sociohydrological flood risk  
697 model. *Water Resources Research*, 55, 1312–1336. <https://doi.org/10.1029/2018WR024128>
- 698 4. Barredo JI (2009) Normalised flood losses in Europe: 1970–2006. *Nat Hazards Earth Syst Sci* 9:97–  
699 104. doi: 10.5194/nhess-9-97-2009
- 700 5. Belcaro, P., Gasparini, D., and Baldessari, M.: 31 ottobre–2 novembre 2010: l'alluvione dei Santi,  
701 Regione Veneto, available at: <http://statistica.regione.veneto.it/> (last access: 21 March 2019),  
702 2011.
- 703 6. Buck, W., & Merkel, U. (1999). *Auswertung der HOWAS-Schadendatenbank* (No. HY98/15).  
704 Institut für Wasserwirtschaft und Kulturtechnik: Universität Karlsruhe.
- 705 7. Bürkner, P.-C. (2017), 'brms: An R Package for Bayesian Multilevel Models Using Stan', *Journal of*  
706 *Statistical Software* 80(1), 1–28.
- 707 8. Carisi, F., Schröter, K., Domeneghetti, A., Kreibich, H., and Castellarin, A.: Development and  
708 assessment of uni- and multivariable flood loss models for Emilia-Romagna (Italy), *Nat. Hazards*  
709 *Earth Syst. Sci.*, 18, 2057–2079, <https://doi.org/10.5194/nhess-18-2057-2018>, 2018.
- 710 9. Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., ... & Riddell, A.  
711 (2017). Stan: A probabilistic programming language. *Journal of statistical software*, 76(1).
- 712 10. Cumbria County Council. (2018). Flooding in Cumbria, December 2015 Impact Assessment,  
713 Carlisle, UK: Cumbria County Council – Performance and Intelligence Team. URL:  
714 [https://www.cumbria.gov.uk/eLibrary/Content/Internet/536/671/4674/17217/17225/43312152](https://www.cumbria.gov.uk/eLibrary/Content/Internet/536/671/4674/17217/17225/43312152830.pdf)  
715 [830.pdf](https://www.cumbria.gov.uk/eLibrary/Content/Internet/536/671/4674/17217/17225/43312152830.pdf) (last accessed 9 May 2019).
- 716 11. D'Alpaos, L., Brath, A., and Fioravante, V.: Relazione tecnoscienza sulle cause del collasso  
717 dell' argine del fiume Secchia avvenuto il giorno 19 gennaio 2014 presso la frazione San Matteo,  
718 available at: <http://www.comune.bastiglia.mo.it/> (last access: 21 March 2019), 2014.
- 719 12. da Costa, R. T., Manfreda, S., Luzzi, V., Samela, C., Mazzoli, P., Castellarin, A., & Bagli, S. (2019). A  
720 web application for hydrogeomorphic flood hazard mapping. *Environmental Modelling &*  
721 *Software*, 118, 172-186.
- 722 13. De Bruijn, K. M., Wagenaar, D. J., Slager, K., De Bel, M., and Burzel, A.: Proposal for the new flood  
723 damage assessment method: SSM2015, Deltares, 2014
- 724 14. DESTATIS (Federal Statistical Office): Statistisches Jahrbuch – Deutschland und Internationales  
725 (2013). Available at:  
726 <https://www.destatis.de/DE/Publikationen/StatistischesJahrbuch/StatistischesJahrbuch2013.pdf>  
727 (last access: 20 May 2018), 2013 (in German).
- 728 15. Domeneghetti, A., Carisi, F., Castellarin, A., Brath, A. Evolution of flood risk over large areas:  
729 quantitative assessment for the Po River. *Journal of Hydrology* 527 (2015) 809–823, 2015;  
730 doi:10.1016/j.jhydrol.2015.05.043
- 731 16. Dottori, F., Figueiredo, R., Martina, M. L., Molinari, D., & Scorzini, A. (2016). INSYDE: a synthetic,  
732 probabilistic flood damage model based on explicit cost analysis.
- 733 17. Duiser, J. A.: Een verkennend onderzoek naar methoden ter bepaling van inundatieschade bij  
734 dijkdoorbraak, 82-0644 TNO, Delft, the Netherlands, 1982
- 735 18. Elmer, F., Thielen, A. H., Pech, I., & Kreibich, H. (2010). Influence of flood frequency on  
736 residential building losses. *Natural Hazards and Earth System Sciences*, 10(10), 2145–2159.
- 737 19. Environment Agency DEFRA. (2019): Recorded Flood Outlines. URL:  
738 <https://environment.data.gov.uk/dataset/8c75e700-d465-11e4-8b5b-f0def148f590>

- 739 20. Figueiredo, R., Schröter, K., Weiss-Motz, A., Martina, M. L. V., and Kreibich, H.: Multi-model  
740 ensembles for assessment of flood losses and associated uncertainty, *Nat. Hazards Earth Syst.*  
741 *Sci.*, 18, 1297–1314, <https://doi.org/10.5194/nhess-18-1297-2018>, 2018.
- 742 21. Gerl, T., Kreibich, H., Franco, G., Marechal, D., & Schröter, K. (2016). A review of flood loss  
743 models as basis for harmonization and benchmarking. *PLoS one*, 11(7), e0159791.
- 744 22. Gneiting, T. and Raftery, A. E.: Strictly Proper Scoring Rules, Prediction, and Estimation, *Journal of*  
745 *the American Statistical Association*, 102(477), 359–378, doi:[10.1198/016214506000001437](https://doi.org/10.1198/016214506000001437),  
746 2007.
- 747 23. ICPR: Atlas of flood danger and potential damage due to extreme floods of the Rhine,  
748 International Commission for the Protection of the Rhine, Koblenz, 2001.
- 749 24. Jongman, B., Kreibich, H., Apel, H., Barredo, JI, Bates, PD, Feyen, L., ... & Ward, PJ  
750 (2012). Comparative flood damage model assessment: towards a European approach.
- 751 25. Kienzler, S., Pech, I., Kreibich, H., Müller, M., & Thieken, A. H. (2015). After the extreme flood in  
752 2002: Changes in preparedness, response and recovery of flood-affected residents in Germany  
753 between 2005 and 2011. *Natural Hazards and Earth System Sciences*, 15(3), 505–526.
- 754 26. Kok, M., Huizinga, H. J., Vrouwenvelder, A. C. W. M., and van den Braak, W. E. W.:  
755 Standaardmethode 2005, Schade en Slachtoffers als gevolg van overstroming, HKV, TNO bouw,  
756 Rijkswaterstaat DWV, 2005
- 757 27. Kreibich, H., Di Baldassarre, G., Vorogushyn, S., Aerts, J. C., Apel, H., Aronica, G. T., et al. (2017).  
758 Adaptation to flood risk: Results of international paired flood event studies. *Earth's Future*, 5,  
759 953–965. <https://doi.org/10.1002/2017EF000606>
- 760 28. Kreibich, H., Seifert, I., Thieken, A. H., Lindquist, E., Wagner, K., & Merz, B. (2011). Recent  
761 changes in flood preparedness of private households and businesses in Germany. *Regional*  
762 *Environmental Change*, 11(1), 59–71.
- 763 29. Lüdtke, S., Schröter, K., Steinhausen, M., Weise, L., Figueiredo, R., & Kreibich, H. (2019). A  
764 consistent approach for probabilistic residential flood loss modeling in Europe. *Water Resources*  
765 *Research*, 55. <https://doi.org/10.1029/2019WR026213>
- 766 30. Merz B, Elmer F, Kunz M, Mühr B, Schröter K, Uhlemann-Elmer S. The extreme flood in June 2013  
767 in Germany. *La Houille Blanche*. 2014; 1(June 2013):5–10. <https://doi.org/10.1051/lhb/2014001>
- 768 31. Merz B, Kundzewicz ZW, Delgado J, Hundedcha Y, Kreibich H (2012) Detection and attribution of  
769 changes in flood hazard and risk. In: Kundzewicz ZW (ed) Changes in flood risk in Europe. IAHS  
770 Special Publication 10:435–458
- 771 32. Merz, B., Kreibich, H., Schwarze, R., Thieken, A. (2010): Review article 'Assessment of economic  
772 flood damage'. - *Natural Hazards and Earth System Sciences (NHES)*, 10, 8, pp. 1697-1724. doi:  
773 <http://doi.org/10.5194/nhess-10-1697-2010>
- 774 33. Merz, B., Vorogushyn, S., Lall, U., Viglione, A., Blöschl, G. (2015): Charting unknown waters - On  
775 the role of surprise in flood risk assessment and management. - *Water Resources Research*, 51, 8,  
776 6399-6416. <https://doi.org/10.1002/2015WR017464>
- 777 34. Molinari, D., Scorzini, A. R., Arrighi, C., Carisi, F., Castelli, F., Domeneghetti, A., Gallazzi, A.,  
778 Galliani, M., Grelot, F., Kellermann, P., Kreibich, H., Mohor, G. S., Mosimann, M., Natho, S.,  
779 Richert, C., Schroeter, K., Thieken, A. H., Zischg, A. P., and Ballio, F.: Are flood damage models  
780 converging to reality? Lessons learnt from a blind test, *Nat. Hazards Earth Syst. Sci. Discuss.*,  
781 <https://doi.org/10.5194/nhess-2020-40>, in review, 2020.
- 782 35. Penning-Rowsell, E. C. and Chatterton, J. B.: The benefits of flood alleviation: A manual of  
783 assessment techniques, Gower Technical Press, Aldershot, 1977
- 784 36. Penning-Rowsell, Edmund & Priest, Sally & Parker, Dennis & Morris, Joe & Tunstall, Sylvia &  
785 Viavattene, Christophe & Chatterton, John & Owen, Damon. (2013). *Flood and Coastal Erosion*  
786 *Risk Management A Manual for Economic Appraisal*.

- 787 37. Polasky, S., Carpenter, S. R., Folke, C. and Keeler, B.: Decision-making under great uncertainty:  
788 environmental management in an era of global change, *Trends in Ecology & Evolution*, 26(8),  
789 398–404, doi:[10.1016/j.tree.2011.04.007](https://doi.org/10.1016/j.tree.2011.04.007), 2011.
- 790 38. R Core Team (2019). R: A language and environment for statistical computing. R Foundation for  
791 Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- 792 39. Rözer, V., Kreibich, H., Schröter, K., Müller, M., Sairam, N., Doss-Gollin, J., ... & Merz, B. (2019).  
793 Probabilistic Models Significantly Reduce Uncertainty in Hurricane Harvey Pluvial Flood Loss  
794 Estimates. *Earth's Future*, 7(4), 384-394.
- 795 40. Sairam, N., Schröter, K., Rözer, V., Merz, B., & Kreibich, H. (2019). Hierarchical Bayesian approach  
796 for modeling spatiotemporal variability in flood damage processes. *Water Resources Research*,  
797 55. <https://doi.org/10.1029/2019WR025068>
- 798 41. Schröter, K., Kreibich, H., Vogel, K., Riggelsen, C., Scherbaum, F., & Merz, B. (2014). How useful  
799 are complex flood damage models?. *Water Resources Research*, 50(4), 3378-3395.
- 800 42. Schröter, K., Kunz, M., Elmer, F., Mühr, B., & Merz, B. (2015). What made the June 2013 flood in  
801 Germany an exceptional event? A hydrometeorological evaluation. *Hydrology and Earth System  
802 Sciences*, 19(1), 309–327. <http://doi.org/10.5194/hess-19-309-2015>
- 803 43. Scorzini A.R., Radice A., Molinari D., A New Tool to Estimate Inundation Depths by Spatial  
804 Interpolation (RAPIDE): Design, Application and Impact on Quantitative Assessment of Flood  
805 Damage, *Water* 2018, 10, 1805; doi:10.3390/w10121805
- 806 44. Scorzini, A. R. and Frank, E.: Flood damage curves: new insights from the 2010 flood in Veneto,  
807 Italy, *J. Flood Risk Manage.*, 10, 381–392, <https://doi.org/10.1111/jfr3.12163>, 2017.
- 808 45. Smith, D. I.: Flood damage estimation – A review of urban stagedamage curves and loss  
809 functions, *Water SA*, 20(3), 231–238, 1994.
- 810 46. Szönyi, M., May, P. and Lamb, R. (2016). Flooding after Storm Desmond. UK: Zurich Insurance  
811 Group. [https://www.jbatrust.org/wp-content/uploads/2016/08/flooding-after-storm-desmond-  
812 PUBLISHED-24-August-2016.pdf](https://www.jbatrust.org/wp-content/uploads/2016/08/flooding-after-storm-desmond-PUBLISHED-24-August-2016.pdf) (last accessed 9 May 2019).
- 813 47. Teng, J., Jakeman, A. J., Vaze, J., Croke, B. F., Dutta, D., & Kim, S. (2017). Flood inundation  
814 modelling: A review of methods, recent advances and uncertainty analysis. *Environmental  
815 modelling & software*, 90, 201-216.
- 816 48. Thielen, A. H., Kreibich, H., Muller, M., & Merz, B. (2007). Coping with floods: Preparedness,  
817 response and recovery of flood-affected residents in Germany in 2002. *Hydrological Sciences  
818 Journal*, 52(5), 1016–1037.
- 819 49. Vogel, K., Weise, L., Schröter, K., & Thielen, A. H. (2018). Identifying driving factors in flood-  
820 damaging processes using graphical models. *Water Resources Research*, 54, 8864–8889.  
821 <https://doi.org/10.1029/2018WR022858>
- 822 50. Wagenaar, D., de Jong, J., and Bouwer, L. M.: Multi-variable flood damage modelling with limited  
823 data using supervised learning approaches, *Nat. Hazards Earth Syst. Sci.*, 17, 1683–1696,  
824 <https://doi.org/10.5194/nhess-17-1683-2017>, 2017.
- 825 51. Wagenaar, D., Lüdtke, S., Schröter, K., Bouwer, L. M., & Kreibich, H. (2018). Regional and  
826 temporal transferability of multivariable flood damage models. *Water Resources Research*, 54,  
827 3688– 3703. <https://doi.org/10.1029/2017WR022233>
- 828 52. Ward, P. J., Blauhut, V., Bloemendaal, N., Daniell, J. E., de Ruiter, M. C., Duncan, M., Emberson,  
829 R., Jenkins, S. F., Kirschbaum, D., Kunz, M., Mohr, S., Muis, S., Riddell, G., Schäfer, A., Stanley, T.,  
830 Veldkamp, T. I. E., and Winsemius, H. C.: Review article: Natural hazard risk assessments at the  
831 global scale, *Nat. Hazards Earth Syst. Sci. Discuss.*, <https://doi.org/10.5194/nhess-2019-403>, in  
832 review, 2019.
- 833 53. Wind, H. G., Nierop, T. M., de Blois, C. J., and de Kok, J. L.: Analysis of flood damages from the  
834 1993 and 1995 Meuse floods, *Water Resour. Res.*, 35, 3459–3466, 1999.

- 835 54. Winter, B., Schneeberger, K., Huttenlau, M. et al. *Nat Hazards* (2018) 91: 431.  
836 <https://doi.org/10.1007/s11069-017-3135-5>
- 837 55. WL Delft: Onderzoek watersnood Maas, Deelrapport 1: Wateroverlast December, Deelrapport 9:  
838 Schade, WL Delft, Delft, Appendix A, 1994 (in Dutch).
- 839 56. Zischg, A.P., Felder, G., Mosimann, M., Röthlisberger, V., Weingartner, R., Extending coupled  
840 hydrological-hydraulic model chains with a surrogate model for the estimation of flood losses,  
841 *Environmental Modelling and Software* (2018), doi: 10.1016/j.envsoft.2018.08.009