Humboldt-Universität zu Berlin - Geographisches Institut

Cumulative Dissertation

Bayesian Approaches for Modelling Flood Damage Processes

Zur Erlangung des akademischen Grades
doctor rerum naturalium
(Dr. rer. nat.)
eingereicht an der
Mathematisch-Naturwissenschaftlichen Fakultät
der Humboldt-Universität zu Berlin
von

Nivedita Sairam

Dekan der Mathematisch-Naturwissenschaftlichen Fakultät
Prof. Dr. Elmar Kulke

Gutachter/in: 1. Prof. Tobias Krüger

2. Prof. Dagmar Haase

3. Prof. Alberto Viglione

Datum der Einreichung: 15/03/2020

Datum der Promotion: 3/11/2020

Declaration of Originality

I, Nivedita Sairam, declare that I have completed the thesis independently using only the aids, tools and sources specified. I have not applied for a doctor's degree in the doctoral subject elsewhere and do not hold a corresponding doctor's degree. I have taken due note of the Faculty of Mathematics and Natural Sciences PhD Regulations, published in the Official Gazette of Humboldt-Universität zu Berlin no. 42/2018 on 06/03/2020.

Location, Date

Nivedita Sairam

Acknowledgements

I would like to express my gratitude and thanks to a number of individuals who helped me directly and indirectly in making this thesis a reality. Firstly, I am very grateful to my advisor, Heidi Kreibich for supporting me at every step through my PhD. Heidi always encouraged discussions about new ideas and provided valuable guidance to improve my research abilities, writing and presentation skills and career planning.

I would like to thank my mentor, Kai Schröter, for the brainstorming discussions without which this thesis would not have been possible. My special thanks go to Bruno Merz, who gave me the opportunity to work at the section Hydrology at the GFZ. His support and recommendations were paramount in the success of this thesis.

My sincere thanks go to my thesis committee members, Prof. Doris Dransch, Prof. Tobias Krüger, Prof. Dagmar Haase and Dr. Sebastian Scheuer and external reviewer Prof. Alberto Viglione, for agreeing to dedicate their expertise and precious time to review this thesis.

I am very thankful for having been part of the System Risk ITN, Marie Curie fellowship. I would like to thank the supervisors, mentors and fellow ESRs from the project for organizing training programs, secondments and successful collaborations. I would also like to thank colleagues from section Hydrology for creating a nice and friendly work environment. Also, I would like to thank my peers Duha Metin, Max Steinhausen, Viktor Rözer, Fabio Brill, Stefan Lüdtke and Marlies Barendrecht for the numerous discussions that contributed to the thesis.

I would like to express my heartfelt thanks to my friends and family for supporting me throughout my PhD. Special thanks go to my father, for being my role model for continuous learning. Finally, I would like to thank my best friend and husband, Kaushik, for always believing in me and for being my go-to support system during the last three years.

Zusammenfassung

Hochwasser betreffen Menschen weltweit und verursachen Verluste in Höhe von mehr als 100 Milliarden US-Dollar (UNISDR, 2015). Mit dem zunehmenden globalen Wandel – beispielsweise steigenden Lebensstandards, Landnutzungsänderungen und Klimawandel – wird ein weiterer Anstieg der hochwasserbedingten Schäden erwartet (IPCC, 2012). Um dieser Entwicklung entgegenzuwirken, müssen starke Anstrengungen zur Reduzierung des Hochwasserrisikos unternommen werden. Hierzu bedarf es verlässlicher Methoden der Hochwasserschadensschäzung.

Hochwasserschadensprozesse werden von den drei Komponenten des Hochwasserrisikos bestimmt – der Gefahr, der Exposition und der Vulnerabilität. Im Vergleich zu Gefahr und Exposition wird die Vulnerabilität, obwohl von gleicher Bedeutung, häufig nur stark vereinfacht durch eine Wasserstand-Schadens-Funktion abgebildet. Dabei bleiben wichtige Einflussgrößen auf die Vulnerabilität, wie die private Hochwasservorsorge aufgrund fehlender quantitativer Informationen unberücksichtigt. Diese Arbeit entwickelt daher eine robuste statistische Methode zur Quantifizierung des Einflusses von privater Hochwasservorsorge auf die Reduzierung der Vulnerabilität von Haushalten bei Hochwasser. Es konnte gezeigt werden, dass in Deutschland private Hochwasservorsorgemaßnahmen den durchschnittlichen Hochwasserschaden pro Wohngebäude um 11.000 bis 15.000 Euro reduzieren. Hochwasserschadensmodelle mit Expertenwissen und datengestützten Methoden sind dabei am besten in der Lage Unterschiede in der Vulnerabilität durch private Hochwasservorsorge zu erkennen.

Hochwasserschadenprozesse sind stochastische Prozesse. Die über sie erhobenen Daten und Modellannahmen sind von Unsicherheit geprägt und so sind auch Schätzungen mit Hochwasserschadensmodellen immer unsicherheitsbehaftet. Trotz dessen nutzt eine Vielzahl von Studien einfache, deterministische Modelle zur Schadensschätzung. Dies kann zu einem unvollständigem Verständnis möglicher Schadensszenarien und verzehrten Risikobewertungen fürhern. Die Bayesschen Modelle, die in dieser Arbeit entwickelt und angewandt werden, nutzen Annahmen über Schadensprozesse als Prior und empirische Daten zur Aktualisierung der Wahrscheinlischkeitsverteilungen. Die Modelle bieten Hochwasserschadensschätzungen als Verteilung, welche die Bandbreite der Variabilität der Schadensprozesse und die Unsicherheit der Modellannahmen abbilden. Daher sind sie eine Verbesserung der etablierten

Hochwasserschadensmodelle, hinsichtlich der Prognoseerstellung und Anwendbarkeit. Ins Besondere verbessert die Verwendung einer Beta–Verteilung die Zuverlässigkeit der Modellergebnisse im Vergleich zu den häufig genutzten Gaußschen oder nicht parametrischen Verteilungen. Die datengestützten Bayesschen Modelle quantifizieren Unsicherheiten in etablierten deterministischen Modellen mit synthetischer Datenbasis; der hierarchische Bayessche Ansatz schafft eine verbesserte Parametrisierung von Wasserstand-Schadens-Funktionen und ersetzt so die Notwendigkeit empirischer Daten durch regional- und Ereignis-spezifisches Expertenwissen. Auf diese Weise kann die Vorhersage bei einer zeitlich und räumlichen Übertragung des Models verbessert werden.

Summary

Floods affect people worldwide and account for more than USD 100 billion losses (UNISDR, 2015). Through rapidly changing climate and increasing settlements in flood plains, risk of flooding has risen globally and it is also expected to increase in the future (IPCC, 2012). In order to develop effective Flood Risk Management (FRM) strategies, reliable prediction of flood losses is important.

Flood damage processes are influenced by the three components of flood risk - hazard, exposure and vulnerability. In comparison to hazard and exposure, the vulnerability component, though equally important is often generalized in many flood risk assessments by a simple depth-damage curve. Hence, this thesis developed a robust statistical method to quantify the role of private precaution in reducing flood vulnerability of households. In Germany, the role of private precaution was found to be very significant in reducing flood damage (11 - 15 thousand euros, per household). Also, flood loss models with structure, parameterization and choice of explanatory variables based on expert knowledge and data-driven methods were successful in capturing changes in vulnerability, which makes them suitable for future risk assessments.

Flood damage processes are stochastic and significant uncertainty in the underlying data and model assumptions exists. Hence, flood loss models always carry uncertainty around their predictions. However, many risk assessment studies use simple, deterministic models for flood loss predictions that may result in incomplete understanding of the possible loss scenarios and biased risk estimations. This thesis develops Bayesian approaches for flood loss modelling using assumptions regarding damage processes as priors and available empirical data as evidence for updating. Thus, these models provide flood loss predictions as a distribution, that potentially accounts for variability in damage processes and uncertainty in model assumptions.

The models presented in this thesis are an improvement over the state-of-the-art flood loss models in terms of prediction capability and model applicability. In particular, the choice of the response (Beta) distribution improved the reliability of loss predictions compared to the popular Gaussian or non-parametric distributions; the Bayesian Data-Driven approach quantified uncertainty in established deterministic synthetic models; the Hierarchical Bayesian approach resulted in an improved parameterization of the common stage damage functions that replaces empirical data requirements with region and event-specific expert knowledge, thereby, enhancing its predictive capabilities during spatiotemporal transfer.

Contents

List of Figures

List of Tables

List of Abbreviations

ABI	Association of British Insurers
ACS	American Community Survey
AIC	Akaike Information Criterion
ATE	Average Treatment Effect
BDDS	Bayesian Data-Driven Synthetic
BIC	Bayesian Information Criterion
BN	Bayesian Network
BT	Bagging decision Tree
CATI	Computer-Aided Telephone Interview
CV	Cross-Validation
DE	Deutschland (Germany)
DEFRA	Department for Environment, Food & Rural Affairs
DESTASIS	Federal Statistical Office of Germany
DLR	German Aerospace Center
ELPD	Expected Log-pointwise Predictive Density
EU	European Union
EUR	Euro (currency)
FEMA	Federal Emergency Management Agency
FLEMOps	Flood Loss Estimation Model for private sector
FRM	Flood Risk Management
GBP	British Pound Sterling
GDP	Gross Domestic Product
HBM	Hierarchical Bayesian Model
HCAD	Harris County Appraisal District
HDI	Highest Density Interval
HR	Hit Rate
ICPR	International Commission for the Protection of the Rhine
IPCC	Intergovernmental Panel on Climate Change
INSYDE	In-depth Synthetic Model for Flood Damage Estimation
IS	Interval Score
IT	Italy
LOGO	Leave-One-Group-Out
LOO	Leave-One-Out
MAE	Mean Absolute Error
MBE	Mean Bias Error
MCM	Multi Coloured Manual
MCMC	Markov Chain Monte Carlo
MDM	Mahalanobis Distance Matching
NL	Netherlands
NN	Nearest Neighbor
NRC	National Response Center
NUTS	No U-Turn Sampler
PLR	Property Level Resilience
PNNL	Pacific Northwest National Laboratory
PS	Propensity Score

PSIS	Pareto Smoothed Importance Sampling
PSM	Propensity Score Matching
RAM	Rhine Atlas Model
RT	Regression Tree
SDF	Stage-Damage Function
SE	Standard Error
SSM	Standard Damage and Fatality assessment model
TX	Texas
UK	United Kingdom
USA	United States of America
USD	US Dollar (currency)
RCNLD	Replacement Cost New Less Depreciation
RMSE	Root Mean Squared Error
ZKI	Zentrum für satellitengestützte Kriseninformation (Center for Satellite-based Crisis Information)

1 Motivation and Objectives

1.1 Flood Risk

Natural Hazards include severe weather and climatic events. Natural hazards become disasters when they severely impact people and their livelihoods. The impacts of natural disasters are non-uniformly distributed over regions. Hence, the geographic and socio-economic contexts of the regions are essential for developing appropriate Flood Risk Management (FRM) strategies (IPCC, 2012). Among several natural disasters, floods affect more people than other disasters. Worldwide, damage due to floods are increasing (IPCC, 2012; Barredo, 2009; Neumayer & Barthel, 2011). In order to reduce the impacts of flooding, many countries have recognized the importance of establishing a systems approach towards flood risk. A systems approach requires understanding and modelling the drivers of flood risk and their feedbacks (Vorogushyn et al., 2018). Flood risk is influenced by hazard, exposure and vulnerability components. (IPCC, 2012). In the context of flood risk, hazard is the characteristic of the physical event, which includes aspects such as inundation depth, duration, velocity, etc. Exposure generally refers to the people, infrastructure, and assets in regions prone to flooding. Vulnerability is the susceptibility of the exposed people and assets to be affected by flooding.

A comprehensive flood risk assessment considers the entire risk chain including the dynamics of hazard, exposure and vulnerability components (Merz et al., 2014b; Vorogushyn et al., 2018). Though this is recommended for developing FRM strategies and analyzing risk portfolio for (re-) insurance markets, very few studies consider the interactions and feedbacks between the different components of flood risk chain (Vorogushyn et al., 2018). The advancements in the hydrology and hydraulics along with the increase in computational capabilities have led to continuous simulation of the flood risk chain (Falter et al., 2015) and large-scale flood inundation maps for varying return periods (Rojas et al., 2013; Ward et al., 2013; Paprotny et al., 2017). A recent study confirmed the influence of climate change in consistently shifting the spatial and temporal patterns of floods in Europe (Blöschl et al., 2017). Though the influence of climate change on flood risk is extensively addressed by a number of studies, they often resulted in fragmented conclusions depending on case studies, modelling approaches and scenarios (Metin et al., 2018).

In addition to the hazard component, it is important to consider the influence of human activities (exposure and vulnerability) within the risk chain, for assessing the consequences of flooding in the future (IPCC 2012). In the case of regional or small-scale case studies, normalized asset values adjusted for inflation based on building typologies (Lüdtke et al., 2019) along with land use classes (Kreibich et al., 2017a) were used as indicators of exposure. In large-scale (global) studies, exposure to Natural hazards was generally represented in terms of GDP and population (Ward et al., 2019). According to a recent study that analyzed flood impacts in Europe over a period of 150 years, it is inferred that though there is an increase in inundated area and number of people affected, there is no significant increase in damage when the values are normalized based on the GDP (Paprotny et al., 2018). Complementing this inference, several studies (Winsemius et al., 2016; Güneralp et al., 2015) also substantiate the importance of quantifying changes in exposure (urban development) as a significant driver of flood risk.

Vulnerability of the exposed population and assets to flooding strongly depends on their adaptation capabilities. In conventional FRM, adaptation strategies were limited to large-scale structural protection measures such as dikes, flood walls and retention basins. However, a systems approach takes into consideration where the flood defenses might fail, for example, dike breech or overtopping. Hence, this approach complements structural flood protection with nonstructural or soft solutions, for example, raised awareness, private precaution, land use planning, organizational emergency management and insurance (Bubeck et al., 2017; Kreibich et al., 2015; Kunreuther et al., 2009). Effective adaptation strategies can reduce the vulnerability and compensate for the adverse effects of climate change (Di Baldassarre et al., 2013; Jongman et al., 2015; Mechler and Bouwer, 2015; Metin et al., 2018). However, the feasibility of implementation of adaptation measures is highly contextual. For example, in some cases, the floodplain management is able to prohibit settlements in the flood prone areas (Burby et al., 1988), while in other cases, this could not be implemented (Crichton, 2012). According to section 5 of the German Federal Water Resource Act, it is the obligation of every person who is prone to flood risk to undertake appropriate actions that are reasonable and within one's means (Rolfsen, 2009). Whereas, in some parts of England, the environmental agency undertook pilot campaigns implementing Property level resilience (PLR) measures in Appleby region spending approximately £5,000, per household (Defra/Environment Agency, 2009). Vulnerability reduction via successful adaptation leading to a reduction in incurred flood damage is evident from a paired event study, where the damage from the second event always caused lower damage than the first event due to better adaptation practices (Kreibich et al., 2017b). A strong interplay is persistent across changes in hazard, exposure and vulnerability, resulting in complex feedbacks within the system (Di Baldassarre et al., 2013). Therefore, a systems approach toward flood risk is required to identify the hazard, exposure and vulnerability factors that influence flood risk and develop modelling methods to quantify their influence on flood damage.

1.2 Flood vulnerability of private households

Often, a large number of private households or residential buildings are impacted by floods resulting in damage to the building structure and contents (Merz et al., 2010b). During the extreme floods in Germany in 2013, the private households suffered almost 1/4th out of the total damage of 6.67 Billion Euros (Bundesministerium des Innern, 2013). Managing flood risk to private households requires identification of the factors influencing and resisting damage to private buildings and modelling the damage processes. The focus of this thesis is the flood damage processes that occur in private household buildings. In the reminder of the thesis, flood loss or flood damage corresponds to private household buildings.

Vulnerability of private households are influenced by the household or building characteristics and adaptation/private precautionary measures undertaken by the household (Few, 2003). Commonly implemented private precautionary measures may be engineering or non-structural/soft measures. Engineering measures require applying alterations to building structures. Some examples include, sealing basements, elevating the house and so on. Non-structural or soft measures mainly include avoiding expensive fittings and low-value usage on in the flood prone floors, securing oil tanks, installing and flood barriers/gates (Kreibich et al., 2005, 2015). In general, soft measures are more widely implemented by households as they do not have a direct financial implication on the owners (Kreibich et al., 2011). Some studies also consider a third group of measures that influence the flood vulnerability under private precautionary measures. These are non-primary measures and mostly influence the uptake of the other measures (Cumiskey et al., 2018). The non-primary measures often include awareness increase, knowledge concerning preparedness and motivation to protect from floods. These are often achieved using awareness campaigns, regulations, financial incentives, financial disincentives etc. In some contexts, flood insurance is also considered as a primary adaptation measure as they reduce vulnerability of private households by transferring risk.

In addition to the non-primary measures, the uptake or implementation of private precaution is also motivated by influencing variables. These include socio-economic profile of the household, building characteristics. The socio-economic variables have low explanatory capability to predict implementation of precaution, on their own. However, they result in valuable interpersonal characteristics when combined with threat and coping appraisals and developed into a socio-psychological model (Bubeck et al., 2018; Grothmann et al., 2006). These studies are majorly based on Protection Motivation Theory (PMT) (Grothmann et al., 2006; Bubeck et al., 2012; Bubeck et al., 2017; Poussin et al., 2015). Households with low risk perception are more inclined towards not implementing precaution, especially when they are reliant on large scale protection measures (Babcicky & Seebauer, 2019). Self and response efficacy also influence implementation of private precaution (Botzen et al., 2019). Since the threat appraisal peaks after a devastating flood, many residents implement precautionary measures (Bubeck et al., 2012). Government policies and financial subsidies are also highly relevant in motivating private households to implement precautionary measures (Defra/Environment Agency, 2009). However, in all the cases, communication of risk and appropriate adaptation measures was found to be very important for maximizing the effectiveness of private precaution (Richert et al., 2017; Babcicky & Seebauer, 2019).

Most FRM strategies consider that private precautionary measures are effective in preventing or reducing flood losses. However, in order to encourage implementation of private precautionary measures, it is required to accurately estimate their loss reducing capability (Richert et al., 2017). Some studies estimate the effectiveness of implementing private precaution. These studies may be based on empirical data (e.g., Hudson et al., 2014; Kreibich et al., 2005; Poussin et al., 2015) or practical studies based on expert judgment and/or a rather not transparent database (e.g., ABI, 2003; Defra, 2008; ICPR, 2002). The estimates of effectiveness are contextual and vary across different hazard and exposure scenarios. For example, in France, engineering measures such as elevating the ground floor is reported as the most effective measure in reducing the damage to buildings by up to 5,500€ and to home contents by up to 6,500€ (Poussin et al., 2015). Non-structural/soft measures such as flood adapted use and flood adapted interior fitting are reported as the most effective precautionary measures with building loss reductions of about 50% or in terms of absolute loss reductions of over 10,000€ in Germany (Hudson et al., 2014; Kreibich et al., 2005). This is also in agreement with the results of cost-benefit analyses that revealed that low-priced measures like elevating the boiler and securing the oil tank as the most cost-effective ones (Kreibich et al., 2011, 2012; Poussin et al., 2015). These are especially beneficial for households in <50-year return period regions. Whereas, more expensive structural measures are only suitable for high-risk regions (Kreibich et al., 2011).

Recent studies combined the propensity of private households to undertake precaution along with other drivers of flood risk to better model the role of private precaution (Haer et al., 2016; Yang et al., 2018). While these models are valuable for determining future adaptation scenarios, for quantifying future risk in terms of monetary damage, it is crucial to combine the potential adaptation scenarios along with the effectiveness of the adaptation measures. In this respect, it is crucial that flood loss models are capable of capturing the loss reducing effect of private precautionary measures. Thus, quantifying the effectiveness of private precaution not only helps in motivating implementation of measures, but also, if captured by flood loss models, supports comprehensive risk assessment based on different adaptation scenarios.

1.3 Advancements in flood loss modelling

In flood risk assessment, the vulnerability component is generally quantified using flood loss estimation models, which are also called vulnerability functions (Ward et al., 2019). Flood loss models are an essential component of the risk chain as they quantify flood risk in terms of economic losses (Bubeck & Kreibich, 2011; Merz et al., 2010b). Many studies use the so-called stage-damage functions for estimating economic losses (Muis et al., 2015; Arnell & Gosling, 2016). The stage damage functions are uni-variable models with water depth as the only predictor of flood damage (NRC, 2000). For accurate and reliable prediction of flood losses in private households, it is important that the flood loss models closely represent the building damage processes. Flood loss models are developed mainly based on the synthetic approach (e.g. Penning-Rowsell and Chatterton, 1977; Parker et al., 1987; Smith, 1994; Klaus et al., 1994) and the empirical approach (e.g. Nicholas et al., 2001; Zhai et al., 2005; Thieken et al., 2008; Elmer et al., 2010; Kreibich et al., 2010 and Carisi et al., 2018).

Synthetic models are developed using What-If scenarios synthesized using multiple sources of information concerning damage processes (Penning-Roswell and Chatterton, 1977). These may include expert knowledge and engineering analysis. These models are not generally fit to observed loss values and are also rarely validated against observed losses except for a few models such as the INSYDE (Dottori et al., 2016). Thus, it is possible to develop synthetic models for regions with no recorded flood event and/or no available empirical loss data. However, high effort is required to develop detailed scenarios of damage processes comprising of all influencing variables. Also, the relationships between loss influencing factors are subjective to the expert’s opinion and may lead to high uncertainty, which remains obscure.

Empirical models are developed mostly using relationships between observed loss and its influencing variables derived from empirical data (Merz et al., 2010b). Several sources of empirical data include face-to-face surveys, telephone surveys, insurance claims etc. These models are mostly data-driven. That is, the damage processes are defined completely using attributes in the empirical data and their statistical relationship to the incurred loss. With improvements in computation capability and machine learning algorithms, more complex multi-variable empirical models were developed to predict flood losses (Merz et al., 2013; Kreibich et al., 2017a; Nafari et al., 2016). These models improved loss prediction by explicitly taking into consideration the effect of factors influencing flood damage. With a comprehensive empirical dataset, the probabilistic models quantify uncertainty due to hydrological, early warning, preparedness and socio-economic attributes which provide reliability of loss predictions. Despite these advantages, the empirical models generally do not perform well when transferred to a different region or event type than the one they are intended for (Schröter et al., 2014; Cammerer et al., 2013; Jongman et al., 2012). Hence, it is not advisable to directly use the models for assessing potential risk for a region with no available local empirical loss data.

In addition to the empirical and synthetic approaches, a few models are based on both empirical data and expert knowledge. FLEMOps (Thieken et al., 2008; Elmer et al., 2010) and Bayesian Network based BN-FLEMOps (Wagenaar et al., 2018) are some examples of such models. These models combine causal inferences from expert knowledge along with the statistical relationships based on data-driven methods. In comparison to the purely data-driven models such as RT-FLEMOps (Merz et al., 2013) and BT-FLEMOps (Merz et al., 2013; Kreibich et al., 2017a), the Bayesian Network based models are adept in generalizing over different damage processes (Schröter et al., 2014; Wagenaar et al., 2018).

The data-driven models developed using local data from specific regions tend to implicitly account for regional characteristics which are common to all households in the region. However, these characteristics tend to be different for households from a different region, experiencing a different event. Hence, transferring the local model may result in biased loss estimates for the new event. For example, most households affected by the 2013 event in the Danube catchment in Germany had low preparedness and experienced oil contamination which resulted in large damage even for small water depths (Thieken et al., 2016). A data-driven loss model developed only with empirical data from Danube 2013 may over-estimate damage for a different event where households experienced no contamination. The models that include variables and relationships based on expert knowledge explicitly define these characteristics that are generally implicitly assumed by a data-driven model. While the BN-FLEMOps is advantageous during regional and temporal transfer (Wagenaar et al., 2018), the model was found to be less suitable in the absence of object-level data concerning the detailed indicators of vulnerability (Lüdtke et al., 2019). Thus, for performing large-scale risk assessments, there is a need to develop modelling approaches that improve the representation of flood damage processes when there is limited or no empirical data available.

1.4 Objectives

Accurate and reliable risk assessments are crucial for developing FRM strategies to counteract the increasing risk due to flooding. Flood risk is commonly quantified in terms of expected monetary damage, which are estimated using flood loss models. Currently, a number of flood loss models of varying complexity are available from the research community and insurance industry. However, little attention is given regarding the models' capability to represent the influence of changes in vulnerability on flood damage processes. However, for risk-based decision making, flood loss models are required to account for changes in vulnerability, especially when human activities alter damage processes.

In addition, the applicability of the developed models to different types of flood events and regions remains unclear. This is mainly because most of the existing models are not transparent regarding their underlying assumptions. They are often localized and not validated against observed flood losses. These concerns may result biased loss estimates when transferring the flood loss model to a different region or type of flood event characterized by damage processes that are not captured by the model. While implementing an existing flood loss model, often, data concerning influencing factors of damage are not available from the target region, at the resolution required by the model. Hence, data requirements also challenge the applicability of flood loss models.

Therefore, the main objective of this thesis is to improve the overall representation of damage processes using Bayesian modelling approaches by capturing the changes in vulnerability due to implementation of private precaution; reducing uncertainty of loss predictions; enhancing synthetic model predictions and improving spatiotemporal transferability of empirical flood loss models. Deriving from the main objective, this thesis addresses the following research questions in chapters given in the brackets.

1. What is the role of private precaution in reducing flood vulnerability of households? (Chapter 2)

2. Which of the state-of-the-art flood loss models account for the effect of this vulnerability reduction? (Chapter 2)

3. What influences the uncertainty and reliability of flood loss models? (Chapters 3, 4)

4. Can reliability of synthetic flood loss models be quantified? (Chapter 4)

5. How to improve spatiotemporal transfer of flood loss models? (Chapters 3, 4, 5)

The chapters in this thesis are organized into five studies that contribute to improved representation of flood damage processes. This cumulative thesis is organized into six chapters, as shown in figure 1.1. Chapters 2 - 5 are in the form of research manuscripts. Chapter 2 quantifies the role of private precaution in reducing flood vulnerability on the basis of empirical damage data and identifies modelling methods that are capable of capturing such changes in vulnerability.

In order to reduce the uncertainty in flood loss predictions, Chapter 3 proposes an appropriate response distribution for regression-based flood loss models. Chapter 4 proposes a Bayesian Data-driven approach for enhancing the loss predictions from established synthetic models by reducing errors in the loss predictions and quantifying their uncertainty and reliability. Chapter 5 proposes a hierarchical Bayesian approach to parameterize a transferable flood loss model that captures spatiotemporal variability in damage processes using expert knowledge and improves loss predictions when no empirical data or relevant synthetic model is available. This is followed by chapter 6 where important findings from the thesis are reported and discussed along with scope for future research and conclusions.

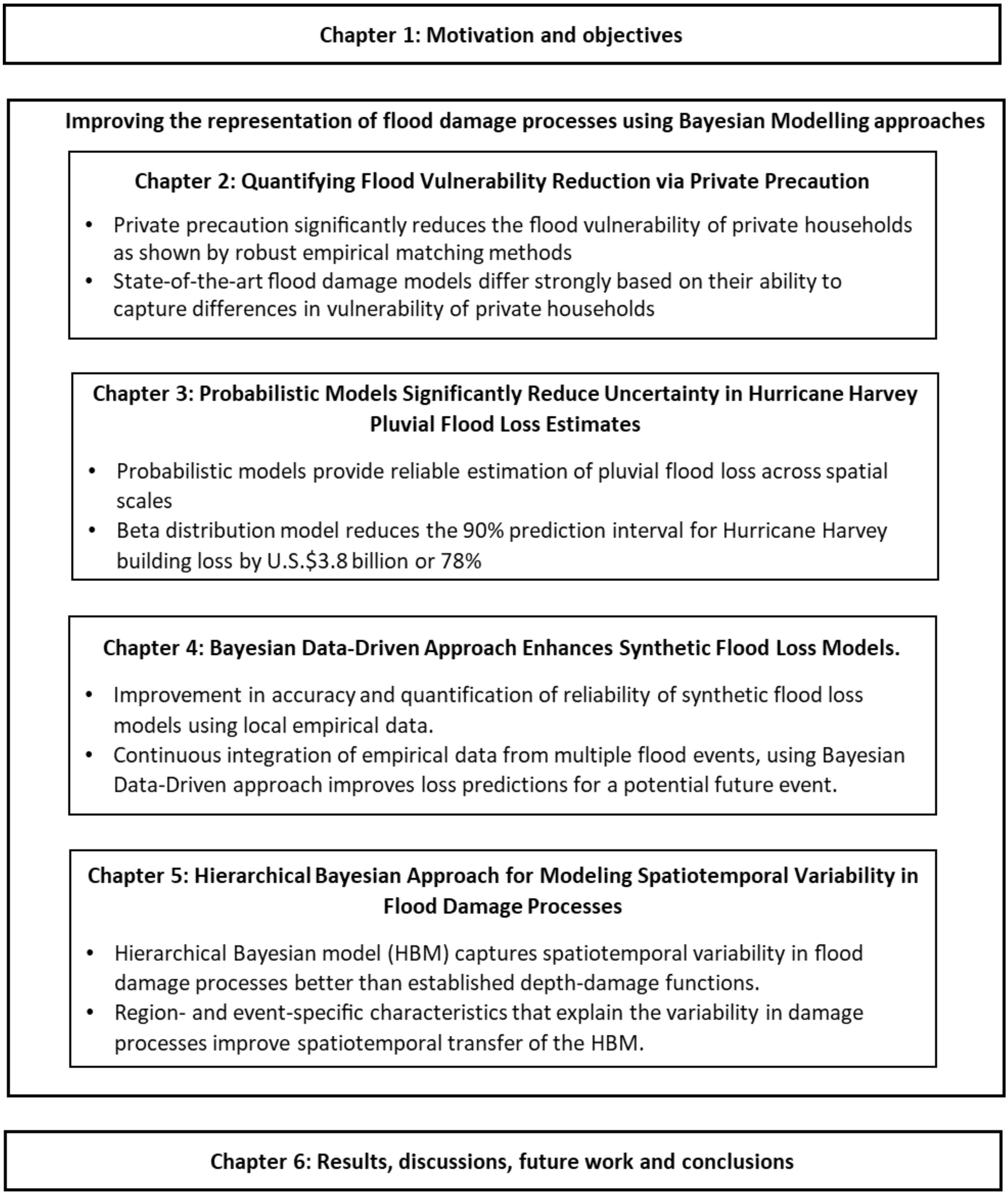


Figure 1.1: Organization of chapters in the thesis

1.5 Author Contributions

This thesis is composed of four manuscripts which have been published, or are submitted and are intended to be published in international peer-reviewed journals. These manuscripts are the result of a collaboration between the author of this thesis (N.S.) and several co-authors. In this section, the contributions of N.S. and all co-authors (initials) are outlined:

Chapter 2: Research Design: N.S., H.K.; Data Analysis and model development: N.S.; Visualization: N.S., S.L; Interpretation of results: N.S., H.K., B.M., K.S.; Writing - Original Draft: N.S.; Writing - review & editing: N.S., K.S., B.M., H.K. and S.L.

Chapter 3: Research Design: V.R., H.K., K.S., U.L. and B.M.; Data collection and processing: M.M.; Data Analysis and model development: V.R., K.S., J.D.-G. and N.S; Visualization: V.R., N.S.; Interpretation of results: V.R., H.K., B.M., K.S., U.L.; Writing - Original Draft: V.R.; Writing - review & editing: V.R., N.S., K.S., M.M., B.M., H.K., J.D.-G and U.L.

Chapter 4: Research Design: N.S., K.S., H.K.; Case study contribution: F.B., S.P., C.V., A.D., F.C., D.M., D.W.; Data Analysis and model development: N.S.; Visualization: N.S.; Interpretation of results: N.S., H.K., B.M., K.S.; Writing - Original Draft: N.S.; Writing - review & editing: N.S., K.S., B.M., H.K., F.B., S.P., C.V., A.D., F.C., D.M., D.W.

Chapter 5: Research Design: N.S., K.S., B.M.; Data Analysis and model development: N.S.; Visualization: N.S.; Interpretation of results: N.S., H.K., B.M., K.S.; Writing - Original Draft: N.S.; Writing - review & editing: N.S., K.S., B.M., H.K. and V.R.

2 Quantifying Flood Vulnerability Reduction via Private Precaution

Summary

Private precaution is an important component in contemporary FRM and climate adaptation. However, quantitative knowledge about vulnerability reduction via private precautionary measures is scarce and their effects are hardly considered in loss modelling and risk assessments. However, this is a prerequisite to enable temporally dynamic flood damage and risk modelling, and thus the evaluation of FRM and adaptation strategies. To quantify the average reduction in vulnerability of residential buildings via private precaution empirical vulnerability data (n=948) is used. Households with and without precautionary measures undertaken before the flood event are classified into treatment and non-treatment groups and matched. Post-matching regression is used to quantify the treatment effect. Additionally, we test state-of-the-art flood loss models regarding their capability to capture this difference in vulnerability. The estimated average treatment effect of implementing private precaution is between 11 and 15 thousand EUR per household, confirming the significant effectiveness of private precautionary measures in reducing flood vulnerability. From all tested flood loss models, the expert Bayesian Network based model BN-FLEMOps and the rule-based loss model FLEMOps perform best in capturing the difference in vulnerability due to private precaution. Thus, the use of such loss models is suggested for flood risk assessments to effectively support evaluations and decision making for adaptable FRM.

Published as: 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

2.1 Introduction

An integrated approach towards FRM is conceptualized and accepted in many countries worldwide (Merz et al., 2010a). These concepts consider that flood defences might fail, and thus complement flood protection with non-structural solutions, e.g. private precaution, land-use planning and insurance (Bubeck et al., 2017; Kreibich et al., 2015; Kunreuther et al., 2009). Burby et al., (1988) revealed that floodplain management is able to divert development away from floodplains and reduce potential flood damage. According to § 5 of the German Federal Water Resource Act, it is the obligation of every person who is prone to flood risk to undertake appropriate actions that are reasonable and within one's means (Rolfsen, 2009). Reliable flood risk and cost-benefit analyses are essential for efficient FRM, since they support optimum investments in adaption measures. Cost-benefit analyses need to consider all suitable risk mitigation measures, associated costs and expected flood losses, since incomplete accounting of costs and benefits, for example, only structural measures considered will lead to a deviation from the global optimum in the analyses (Kreibich et al., 2014). The economic damage from floods has been increasing over the last decades, mostly due to societal factors such as increased standard of living, real per capita wealth and population increase (Barredo, 2009; Mechler & Bouwer, 2015), and this trend is likely to continue (IPCC, 2012; Jongman et al., 2014). Assessments need to account for this dynamic nature of risk to be able to detect relevant changes in risk and initiate appropriate adaptation to changes (Kreibich et al., 2014). Thus, there is a need to accurately estimate flood risks over long time-periods. To be able to capture temporal dynamics in flood loss and risk, which is also a prerequisite to enable evaluations of FRM and climate adaptation strategies, loss models that are able to account for differences in vulnerability, e.g. due to private precaution, are necessary.

Flood risk is influenced by a broad range of characteristics and processes, which can be categorized into hazard, exposure and vulnerability (IPCC, 2012). Understanding the role of these components for changes in risk is essential for effective adaptation. Few studies are available which investigate the role of vulnerability, using modelling (Jongman et al., 2015; Mechler and Bouwer, 2015) or empirical (Kreibich et al., 2017a) approaches. There are various definitions of "vulnerability" and many vulnerability concepts consider a quite broad context (e.g. Brooks et al., 2005; Turner et al., 2003; Kelly & Adger, 2000). Additionally, there are suggestions to complement the concept of vulnerability with resilience, which adds considerations of recovery (Fekete et al., 2014; de Bruijn, 2004; Bruneau et al., 2003). For our study, we follow the natural sciences-oriented approach, which defines vulnerability as the characteristic of a system that describes its potential to be harmed (Gouldby et al., 2005; IPCC 2012). Thus, vulnerability is the susceptibility of a household to flooding which is altered by precautionary measures as well as by changes in household or building characteristics (Few, 2003).

Precautionary measures that are commonly implemented amongst private households to reduce residential building loss include waterproof sealing, flood adapted use and flood adapted interior fitting (Kreibich et al., 2005; 2015). It is generally assumed, that precautionary measures are effective in mitigating flood losses (Dutta et al., 2003; Holub and Fuchs 2008; De Moel et al., 2014), and also some empirically based quantitative information is available: The positive effect of private precautionary measures was revealed by loss reductions of 35% and up to 50% between two similar flood events in 1993 and 1995 at the Meuse and the Rhine Rivers, respectively, where many households had undertaken precautionary measures after the flood in 1993 (Wind et al., 1999; Bubeck et al., 2012). Some studies quantified the damage-reducing effect of individual precautionary measures and identified the most effective ones: These include scientific studies based on empirical damage data (e.g. Kreibich et al., 2005; Hudson et al., 2014; Poussin et al., 2015) as well as practical studies based on expert judgment and/or a rather not transparent database (e.g. ICPR 2002; ABI 2003; DEFRA 2008). A study in France identified "elevating the ground floor" to be the most effective measure in reducing the damage to buildings by up to 5500€ and to home contents by up to 6500 € (Poussin et al., 2015). Studies in Germany identified the measures "flood adapted use" and "flood adapted interior fitting" as the most effective precautionary measures with building loss reductions of about 50% or in terms of absolute loss reductions of over 10000 € (Kreibich et al., 2005; Hudson et al., 2014). On the other hand, cost-benefit analyses revealed low priced measures like "elevating the boiler" and "securing the oil tank" as the most cost-effective ones (Poussin et al., 2015; Kreibich et al., 2011, 2012). Depth-damage curves were developed for different types of flood proofing adaptations through flood and exposure simulations (Dawson et al., 2011).

Estimating the damage-reducing effect of precautionary measures from observed flood loss data should consider the possible bias due to confounding variables. One approach was to estimate the difference in average flood loss experienced by households with precaution and households with no precaution, while controlling for similar inundation depth (Kreibich & Thieken, 2009) or inundation depth and building characteristics (Kreibich et al., 2011). This approach to analyze controlled household groups faces two challenges: (1) Controlling for hazard and building variables results in small samples that can be used for further analysis, (2) Controlling the influence of a large number of variables is not feasible. In order to overcome these challenges, Poussin et al., (2015) developed a regression-based method to determine the effectiveness of individual precautionary measures by controlling for the effects of potential flood risk variables. Another suitable approach to control for confounding variables is matching techniques, since they test causal inference with fewer assumptions than typical regression models, using a smaller, pre-processed dataset (Rosenbaum and Rubin, 1983). For instance, Aldrich (2012) used five different methods of matching on propensity scores, i.e. kernel, radius, nearest neighbour, nearest neighbour with replacement and Mahalanobis matching, to investigate the influence of social capital on the pace of population recovery following the 1923 Tokyo earthquake. Allaire (2016) tested the effectiveness of online information and social media in enabling households to reduce disaster losses using propensity score matching, i.e. nearest neighbor and kernel matching was undertaken followed by a post matching regression analysis. That is, the average treatment effect was estimated using the matched sample to run post-matching regression of the outcome on covariates that are associated with flood losses, but not necessarily the likelihood of using social media. Hudson et al., (2014) implemented expert-selected Propensity Score Matching (PSM) to quantify the treatment effect of different precautionary measures on building and content losses. This method is able to control for an extensive set of variables, i.e. all variables that are likely to introduce selection bias. Our study builds on these approaches to determine the average treatment effect of private precaution in general (not focused on individual measures) by matching based on confounders of private precaution and applying post-matching regression controlling for variables describing flood hazard, warning and emergency measures.

Various flood loss models have been developed for estimating direct economic loss to buildings (Smith, 1994; Merz et al., 2010b; Schröter et al., 2014; Carisi et al., 2018). Many models represent the loss in terms of relative loss, which is the ratio between costs of loss and the value of asset at the time of the event. A standard approach are depth-damage functions, that model the loss as a function of one variable, i.e. inundation depth, commonly differentiated according to the building type or use (White, 1964; Grigg & Helweg, 1975; Smith, 1994; Penning-Rowsell et al., 2005). Recently, multi-variable flood loss models have been developed. For instance, FLEMOps+r (Elmer et al., 2010) is a rule-based model to estimate flood loss to residential buildings based on five different classes of water depth, three individual building types, two classes of building quality, three classes of flood frequency, three classes of contamination and three classes of private precaution. Further, more complex models, based on machine learning algorithms and covering various aspects of flood damage processes, are being developed. Examples are multi-variable tree-based models (Merz et al., 2013; Hasanzadeh Nafari et al., 2016; Kreibich et al., 2017a). They do not require any special treatment for discrete and continuous variables and no specific prior assumptions about the distributions of variables. Bagging Decision trees is an ensemble approach with a number of individual trees. The loss estimate is then determined using the mean as the prediction of the ensemble of trees. Also, Bayesian Networks are used in flood loss estimation (Vogel et al., 2014; Schröter et al., 2014; Wagenaar et al., 2018). Bayesian Networks are Directed Acyclic Graphs constructed from assertion of dependencies and principle of conditional independence (Heckerman, 1998). They have the advantage of inherently quantifying uncertainty associated with the loss estimation. Thus, a variety of models with varying complexities and working concepts are currently available, and it is not trivial to decide which one to use for a specific application (Apel et al., 2009; de Moel et al., 2015; Figueiredo et al., 2018). Several studies have tested and compared various flood loss models in respect to their predictive accuracy and reliability (e.g. Cammerer et al., 2013; Jongman et al., 2012; Gerl et al., 2016; Hasanzadeh Nafari et al., 2016). In contrast, to the best of our knowledge, no study so far has examined the ability of loss models in capturing differences in vulnerability due to private precaution. However, loss models with this ability are necessary to enable temporally dynamic flood damage and risk modelling and thus the evaluation of FRM and adaptation strategies.

Hence, our study aims at quantifying the average loss-reducing effect of private precaution, by considering possible biases due to confounding variables, and to assess how well different types of flood loss estimation models are able to represent this difference in vulnerability.

2.2 Data and methods

2.2.1 Description of dataset

The dataset contains flood loss data collected via cross-sectional telephone surveys of private households that had suffered from losses due to floods in 2002, 2005, 2006, 2010, 2011 or 2013 mainly in the Elbe and Danube catchments in Germany (Table 2.1). On basis of flood reports, press releases and flood masks derived from satellite data (www.zki.dlr.de), lists of inundated streets were compiled separately after one or two flood events. On basis of these lists, property-specific random samples of potentially affected households, i.e. their telephone numbers were selected from the public telephone directory. Property-specific means that only one household was interviewed per address. Computer-aided telephone interviews were undertaken in independent campaigns in April/May 2003, in November/December 2006, in February/March 2012 and in February/March 2014 (Table 2.1). In 2003, households from the list of telephone numbers were sampled randomly. In the following campaigns, comprehensive surveys were conducted, i.e. all the researched telephone numbers were contacted. Each interview was focused on one specific flood event. At the beginning of the interview, it was asked if the household had suffered losses due to the specific flood event, the interview was only continued if this was the case. Thus, the dataset does not contain cases, where the precautionary measures fully prevented loss. This limitation of the dataset is considered when interpreting the results. At the beginning of the telephone call, the person on the phone was asked who in the household has the best knowledge about the flood event and the incurred economic losses. Then the interview was undertaken with this person. The questionnaires for all the survey campaigns contained about 180 questions including aspects of hazard (e.g. inundation depth, duration and velocity), flood experience and awareness, early warning, emergency and precautionary measures, building and socio-economic characteristics, building and content losses. Building loss includes all costs (e.g. costs of wages and material) that are associated with repairing the damage caused by floods to the building structure. Damage may be due to moisture penetration as well as cracks, pushed in doors and windows, subsidence or deformation of walls and ceilings, etc. Repair works may include, for instance, plastering, laying bricks, replacing construction components or broken windows. Building losses are adjusted to prices as of 2013 (inflation) by adjusting the reported loss estimates given at the time of the events by the building cost index (DESTATIS, 2013). The losses reported by the surveyed households were believed to be reliable, since most people had restored their building by the time of the survey (except for after the 2002 flood; Kienzler et al., 2015) and had claimed their losses either from government funds or from their insurers. The responses from the survey after the 2002 flood was confirmed by comparing it with official loss data provided by the Saxon Bank of Reconstruction, which looked after providing governmental disaster assistance after the 2002 flood in the federal state of Saxony (Thieken et al., 2005). Nevertheless, data collected via surveys is associated with uncertainty, which is however difficult to quantify since hardly alternative means to measure these variables exist. The building loss ratio was calculated consistently for all surveys as follows: the absolute losses reported by the surveyed households are divided by the building values as at the time of the flood event. Actuarial valuation method

VdS guideline 772 1988-10 (Dietz, 1999), which is commonly used in the insurance sector for Germany was used to estimate absolute values of residential building in terms of replacement costs (in contrast to depreciated values). In order to apply this valuation, some attributes from the survey responses such as total floor space, basement area, number of storeys, roof type, etc. are used. In respect to precautionary measures people were asked about the kind of measure (check list and additional open answers possible, multiple answers possible) and the time of realisation (check list: "undertaken before the flood", "after the flood", "planned within the next six months", "not intended"). The check list contained among others the following building precautionary measures: adapt interior fitting, adapt use and adapt building structure. Adapting interior fitting involves using less expensive fittings that are easily replaceable or preferably water proof fittings in lower floors; Adapting usage to floods means for instance to use the flood endangered floors in a low-value way; Adapting building structure to floods include structural measures like sealing the basement. These measures are also sometimes referred to as passive preparedness measures (Cumiskey et al., 2017) undertaken often after flood events during the reconstruction phase, however always much before an event. Thus, precautionary measures are not dependent on event forecast and early warning information, in contrast to emergency measures. The questionnaire included also questions that reveal the perception of the interviewee regarding aspects like effectiveness of precautionary measures, usefulness of early warning information, and the quality of their building. People were asked to assess these qualitative variables on a scale from 1 to 6, the meanings of the end points of the scales were given to the interviewee. Indicators were developed for some variables such as flood experience, emergency measures and warning information. Variables used in this study are described in Table 2.2 and 2.3. The corresponding questions, possible options for answers and score computation for indicators are included in the appendix (Section 2.5.1). Further details about the development and calculation of indicators are given in Thieken et al., (2005) and Elmer et al., (2010). More information about the individual flood events, the surveys and their results were published in Thieken et al., (2007), Kreibich et al., (2011, 2017(b)) and Kienzler et al., (2015). A total of 4468 interviews were completed, of which 2671 interviews furnished building loss in EUR. If one or more of the precautionary measures are not practically applicable for a particular household, this dataset is not included in the analysis. For example, households with no basement/cellar are not potential candidates for all structural adaptation measures (e.g. sealing the basement). Hence, the households with no basement are removed from the analysis. Since the methodology does not deal with missing variables, households with incomplete data are removed. Thus, data consisting of 974 households with complete observations are available for the analyses.

Table 2.1: Flood surveys: computer-aided telephone interviews with private households who suffered flood loss

Characteristics	**Surveys**			
Date of survey:	April/May 2003	November/December 2006	February/March 2012	February/March 2014
Flood(s):	August 2002	August 2005, April 2006	August 2010, January 2011	June 2013
Affected regions	Elbe and Danube catchments	Elbe and Danube catchments	Elbe, Oder and Rhine catchments	Elbe, Danube, Rhine and Weser catchments
Number of households interviewed:	1697	461	658	1652
Response rate	15%	18%	16%	17%
Sampling type	random	comprehensive	comprehensive	comprehensive
References	Thieken et al., 2007	Kreibich et al., 2011	Kienzler et al., 2015	Kreibich et al., 2017b

2.2.2 Difference in vulnerability between households with respect to private precaution

2.2.2.1 Average Treatment Effect considering selection bias

A dichotomous indicator (0/1) is used to distinguish private households into low/high vulnerability with respect to implementation of precautionary measures (pre). Private precautionary measures considered are: adapt interior fitting, adapt use and adapt building structure. The indicator for private precaution takes a value of 1 for households with one or more of these precautionary measures implemented before the flood (treatment group) and 0 for households with none of these measures implemented before the flood (control group). Actually, many of the households have undertaken several precautionary measures, which differ in their way how they mitigate flood damage to the building structure and function jointly in the case of a flood event.

The average effect of private precaution in reducing building structure losses in EUR, referred to as the Average Treatment Effect (ATE) contributes to the differences in vulnerability between the two groups. ATE is estimated using the Roy-Rubin model (Rubin, 1974; Roy, 1951):

$$ATE(T) = E[Y(T = 1) - Y(T = 0)] \qquad \textit{Equation 2.1}$$

Where T is the Treatment – implementation of one or more private precautionary measures (1/0). Y is the outcome that is influenced by the treatment, i.e. the reported building loss in EUR.

Considering the heterogeneity among the households with respect to building characteristics, socio-economic attributes, flood experience and awareness, the observed difference in losses between the two groups may not be necessarily only due to the effect of private precaution. This is due to the fact that the households from treatment and control groups have different probabilities of undertaking private precaution. The attributes that influence a household to undertake private precaution are the confounding variables or confounders of private precaution (Table 2.2). The bias in ATE caused due to the effect of confounding is called selection bias. Matching households from treatment and control groups based on the sufficient set of confounders provides an appropriate solution to get rid of selection bias. It is important to only include pre-treatment variables to the list of confounders, whose measurement is not altered by the implementation of private precaution (Pearl, 2009). Equation 2.1 is altered to Equation 2.2, where the building loss estimate is conditioned on the treatment variable, i.e. private precaution, as well as the set of confounding variables.

$$ATE(pre) = E(loss|pre > 0, X) - E(loss|pre = 0, X) \qquad \textit{Equation 2.2}$$

Where $ATE(pre)$ is the treatment effect of implementing private precaution; $loss$ is the reported building loss of households (EUR) and X is the set of confounding variables.

From direct answers to interview questions and derived indicators described in section 2.2.1, we choose 16 attributes that potentially influence whether a household undertakes private precaution (Table 2.2). These attributes are potential pre-treatment confounders. They are categorized into building characteristics, socio-economic attributes, flood experience and awareness. In order to remove hidden bias due to unaccounted variation in the characteristics of different flood events that lead to selection bias, we include 'event' as a nominal covariate in the set of confounders.

Table 2.2: List of potential confounders of private precaution

Categories	Attributes	Type	Attribute Explanation - Range, Unit
Building characteristics Building characteristics may induce limitations or technical feasibility to be able to undertake some precautionary measures. (Cumiskey et al., 2017)	Building quality (bq)	ordinal	1 – very good; 6 – very bad
	Building area (ba)	continuous	[24,299997] sq. metres
	Single-family house (bt1)	dichotomous	0 – no, 1 - yes
	Multi-family house (bt2)	dichotomous	0 – no, 1 - yes
	Building value corrected for inflation 2013 (bv)	continuous	[98496, 10411183] EUR
	Number of flats in the building (nfb)	continuous	(1,45) flats
Socio-economic attributes People from varying socio-economic groups vary in aspects like sense of responsibility, willingness to respond and ability to invest in mitigation measures. (Bubeck et al., 2012; Cumiskey et al., 2017)	Ownership – Apartment (own_1)	dichotomous	0 - tenant, 1 – apartment owner
	Ownership – building (own_2)	dichotomous	0 – not building owner, 1 – building owner
	Age of the interviewee (age)	continuous	[16,99] years
	Household size (hs)	continuous	[1,20] persons
	Household monthly net income indicator (inc)	ordinal	11=below 500 EUR to 16=3000 EUR and more
flood experience and awareness Flood Experience and strong social networks improve awareness about hazard and coping appraisal. (Parker et al., 2007; Kreibich et al., 2005; Bubeck	Knowledge about flood hazard (kh)	dichotomous	1 – Has sufficient knowledge, 0 – Has no knowledge
	Flood experience (fe)	ordinal	0 – no flood experience, 9 – recent flood experience
	Neighbourhood preparedness programmes (neigh_ind)	dichotomous	1 – participated in neighbourhood programs, 0 – not participated in neighbourhood programs
	Flood insurance (ins_ind)	dichotomous	1 – Has flood insurance, 0 – Has no flood insurance

et al., 2013, Atreya et al., 2017)	Perceived effectiveness of private precaution (epre)	ordinal	1=very efficient to 6=not efficient at all
	Event	nominal	Flood events in 2002, 2005, 2006, 2011, 2013

2.2.2.2 Matching distances and methods

There are a number of matching methods and distance estimates that can be used to eliminate selection bias and obtain a matched dataset. We test the suitability of two distance estimates: (1) Propensity Score Matching (PSM), and (2) Mahalanobis Distance Matching (MDM).

PSM has been used widely in socio-economics and medical studies (Dehja & Wahba, 1999; Vincent et al., 2002). Propensity score is the probability that a particular household will undertake precautionary measures, conditioned on the set of confounding variables (Equation 2.3).

$$p_i \equiv P(T_i = 1|X) = \frac{1}{(1+e^{X_i\beta})} \qquad \textit{Equation 2.3}$$

Where p_i is the propensity score of i^{th} household in the dataset obtained through linear logistic regression, T is the private precaution indicator (Treatment), X is the set of confounding variables, β is the set of regression coefficients and base e denotes the exponential function. The distance between matched households from the two groups is estimated as the scalar difference between their propensity scores. The common support for propensity scores is determined using equation 2.4. Only households that lie in the range of common support are considered for matching.

$$Common\ Support = [\max(\min(P_t, P_c)), \min(\max(P_t, P_c))] \qquad \textit{Equation 2.4}$$

Where, $P_t\ and\ P_c$ are propensity scores of households with private precaution and with no precaution, respectively. MDM is a covariate matching method. It uses the Mahalanobis distance as the distance estimate. The Mahalanobis distance matrix is furnished using the distance estimates between pairs of households from the two groups, with the set of confounders as covariates (Equation 2.5).

$$M(X_i, X_j) = \sqrt{\left[(X_i - X_j)^T \mathcal{E}^{-1}(X_i - X_j)\right]} \qquad \textit{Equation 2.5}$$

Where $M(X_i, X_j)$ is the Mahalanobis distance estimate between two households i and j based on the set of confounders X, X_i and X_j represent column matrices of values of confounders from treatment and control households, $(X_i - X_j)^T$ denotes the transpose of the matrix $(X_i - X_j)$ resulting in a row matrix and $\mathcal{E}^{-1}$ is the inverse covariance matrix. This results in $M \times N$ Mahalanobis distance matrix (where M and N are the numbers of households in treatment and control groups from the original population). Once the distance estimates are obtained, different methods of matching (Ho et al., 2007) are tested,

1. Nearest neighbourhood (NN) with/without replacement; with/without caliper
2. Inverse Probability Treatment Weighting (IPTW)
3. Genetic matching algorithm (Diamond and Sekhon, 2006)

Small pruning threshold/caliper are required to reduce bias while matching. We consider 1/4th standard deviation of the PSM and Mahalanobis distances as the caliper to remove unsuitable matches, since it reduces the imbalance by at least 90% (Rosenbaum and Rubin, 1985). From the two distance estimates and six matching methods, twelve potential matched datasets are obtained.

2.2.2.3 Quality of matching

Two tests are performed to assess imbalance in individual confounders after matching: the two-sample weighted t-test and the standardised differences test. The potential matched datasets that pass the two tests for all confounders are chosen for the estimation of ATE.

The two-sample weighted t-test evaluates whether the distributions of confounders belonging to treatment and control households are significantly different (Rosenbaum and Rubin, 1985). In the standardised differences test (Equations 2.6 a, b), an absolute value of standardised difference less than 10% for each of the confounders belonging to treatment and control households is considered to be an accurate match (Austin et al., 2006). It is a point estimate with no significance limits attached to it.

$$\text{Standardised difference (continuous)} = \frac{(\bar{x}_T - \bar{x}_C)}{\sqrt{\frac{(s_T^2 + s_C^2)}{2}}} \qquad \textit{Equation 2.6(a)}$$

Where $\bar{x}$ is the covariate mean of treatment (T) and control (C) groups; s is the covariate standard deviation of treatment (T) and control (C) groups.

Standardised difference (dichotomous) = $\frac{(p_T - p_C)}{\sqrt{\frac{[p_T(1-p_T)+p_C(1-p_C)]}{2}}}$ *Equation2.6(b)*

p is the sample prevalence (proportion of TRUE (value = 1) in the sample of a dichotomous variable) of the covariate in treatment (T) and control (C) groups.

2.2.2.4 Post-matching regression and sensitivity analysis

Post-matching regression/model-fitting is performed in order to control for the bias in treatment effect introduced by aspects that influence the outcome (building loss), but do not potentially influence the treatment (implementation of private precaution). Varying flooding intensities across different households influence the degree of damage experienced. Further, emergency response measures such as pump out water, use sandbags/barriers and switch-off electricity and gas also potentially reduce flood damage. Unlike, precautionary measures, the implementation of emergency measures are highly dependent on event forecast and early warning. Hence, in addition to the matching procedure, which controls for the pre-treatment variables (table 2.2), potential bias due to flooding situation, emergency measures and warning information (table 2.3) is removed via post-matching regression. The choice of post-matching regression model depends on the ability of the model to account for the influence of flooding scenario, early warning and emergency measures on incurred loss. A standard linear regression model is commonly used to remove bias in post-matching. In addition to linear regression, bagging decision trees (ensemble of 1000 regression trees) are used as the post-matching regression model due to its ability to predict losses with least errors compared to standard linear regression models (Merz et al., 2013). Bagging decision trees are an ensemble of regression trees built on bootstrapped samples of the data such that, model dependency and over-fitting are reduced. Bagging decision trees approximate non-linear regression to heterogeneous data. Using the matched samples that pass the quality check, regression models (linear and bagging decision trees), are built for predicting building loss (in EUR) using the treatment variable (private precaution - pre) and predictors from table 2.3. Two intervention scenarios – treatment (pre=1); control (pre=0) are applied to the model and the loss estimates (in EUR) are determined for each scenario. The difference between the two groups of model estimates result in Average Treatment Effect of private precaution. The survey questions and score computations corresponding to the variables for post-matching regression are included in the appendix section 2.5.1.4.

Table 2.3: Attributes for post-matching regression

Attributes	Attribute Type	Attribute Explanation - Range, Unit
Inundation depth (wst)	continuous	[-245,674] cm
Duration of inundation (d)	continuous	[0,1440] hours
Contamination (con)	ordinal	0 – no contamination to 2– heavy contamination
Velocity of water (v)	dichotomous	0: v=0, 1: v>0
Emergency measures (em)	ordinal	1=no measures undertaken to 17=many measures undertaken
Warning information (wi)	ordinal	0=no helpful information to 12=many helpful information

The ATE estimate may still be sensitive to the choice of confounders (Caliendo & Kopeinig, 2005). Since the list of confounders are chosen through expert knowledge, there is a possibility that some aspects of confounding may be missing or unmeasured. The matching methodology cannot eliminate potential bias due to unobserved or missing confounders. Potential unobserved or missing confounders that the analysis does not control for may be specific building or contents characteristics which may favour or hamper certain building precautionary measures, or differences in the ability of households to undertake measures. Rosenbaum's Sensitivity Analysis (Rosenbaum, 2002) using Hodges-Lehmann point estimate quantifies the robustness of the causal relationship between treatment (precaution) and outcome (building loss) to the presence of bias introduced by missing confounders (DiPrete & Gangl, 2004).

Two households which are matched based on the set of confounders may vary in the probability of undertaking private precaution by at most a factor of Γ (sensitivity parameter). If $\Gamma = 1$, the two groups of matched households have the same probability of undertaking precaution (no hidden bias). If $\Gamma = 2$, the matched households in the treatment group may have at most twice the probability of undertaking precautionary measures when compared to the households in the control group. When Γ is increased from 1.0, the bounds of Hodges-Lehmann point estimate (Rosenbaum's bounds) widen and the certainty with which we estimate the treatment effect decreases. The robustness of the estimate is represented by the value of Γ, when the Rosenbaum's bounds extend further from the positive effect of treatment and bracket to zero (Keele, 2010).

2.2.3 Ability of Flood Loss Estimation Models to capture differences in vulnerability due to private precaution

2.2.3.1 Flood Loss Estimation Models

A range of flood loss estimation models are applied to the matched dataset to test to which extent the models are able to capture differences in vulnerability due to private precaution. The models are of varying complexities from deterministic rule-based models to probabilistic Bayesian Network based models. All these models estimate the relative building loss (brloss) for private household buildings using multi-variable predictors from the surveys. The brloss values range between 0 (no loss) and 1 (total loss). From the brloss estimates, the absolute losses are computed by multiplying with the building value of the respective private buildings (in EUR) corrected to 2013 inflation.

FLEMOps+r (Elmer et al., 2010; Thieken et al., 2008) estimates relative building losses based on defined rules associating seven input variables (Table 2.4) to relative building loss. FLEMOps+r works in two steps: first, relative flood loss is estimated on basis of water level and building characteristics (i.e. building type and building quality); second, the estimate is refined by a scaling factor which considers contamination (in 3 classes, i.e. no, medium, heavy), precaution (in 3 classes, i.e. little, medium, strong), and recurrence interval (in 3 classes, i.e. 1-9 years, 10-99 years, from 100 years onwards). In this model, private precaution takes a value of 0 for little precaution, 1 for medium precaution and 2 for strong precaution.

Tree based models (Merz et al., 2013; Kreibich et al., 2017a), i.e. regression trees (RT-FLEMOps) and bagging decision trees (BT-FLEMOps), are grown with seven variables (Table 2.4). RT-FLEMOps is grown with a minimum of 60 households in each leaf, resulting in 25 leaves (Figure 2.1 (a)). In RT-FLEMOps, the precautionary measure indicator appears only once in the bottom of the tree, and hence, the variable does not hold an important role in estimating relative building losses. BT-FLEMOps is an ensemble approach consisting of 1000 trees. The variable importance plot (Figure 2.1(b)) shows that private precaution has a relatively low importance. The tree-based algorithms are developed using Statistics and Machine Learning toolbox (MATLAB 2015b).

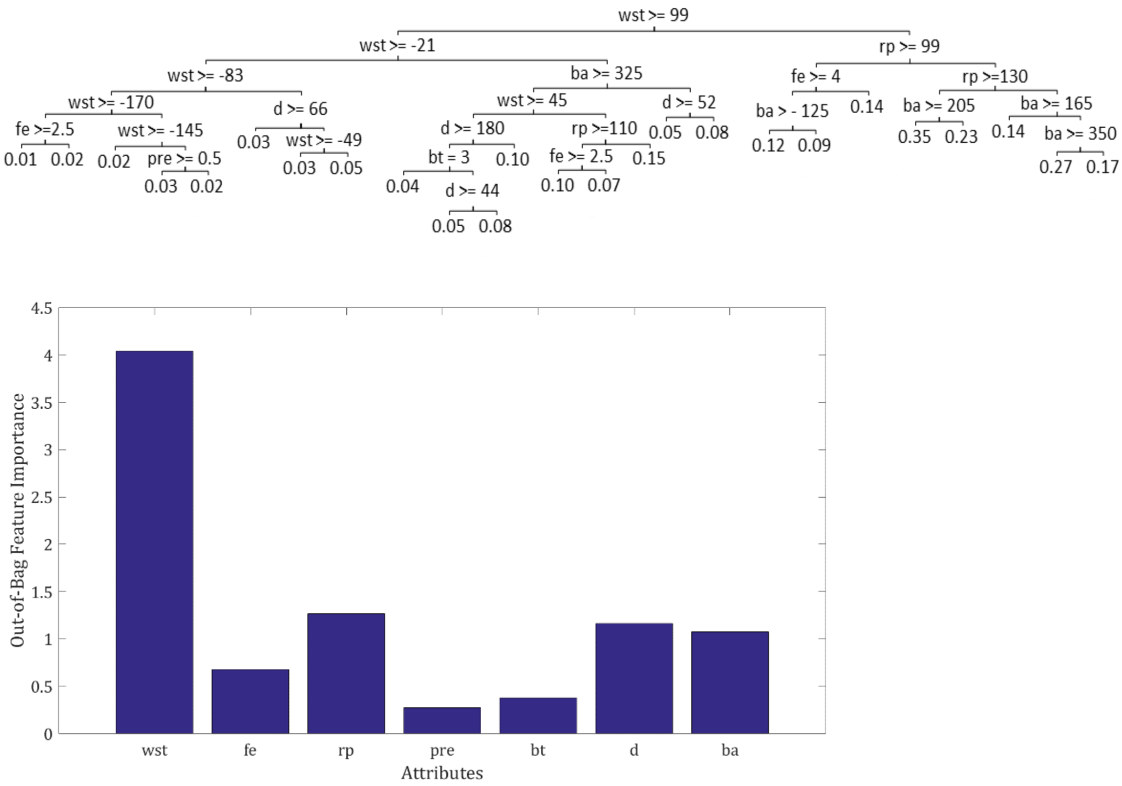


Figure 2.1: (a) Regression tree with 7 variables and 25 leaves (RT1), (b) Feature importance of flood loss predictors for Bagging decision trees BT with 7 variables and an ensemble of 1000 trees

BN-FLEMOps (Wagenaar et al., 2018) is a discrete Bayesian Network model, which is constructed with seven variables (Table 2.4). The continuous variables in the model were discretized on the basis of bins with equal frequency with inundation depth (wst) and relative building loss (brloss) in 10 classes, return period (rp) and inundation duration (d) in 5 classes and building area (ba) in 3 class. The network structure (Figure 2.2) describing the conditional dependencies between the variables is learnt using 500 iterations of score-based local search algorithms - Fast-IAMB (Tsamardinos et al., 2003) and a hill-climbing approach using the Bayesian Dirichlet Equivalent (Heckerman et al., 1995). The set of network structures and all arcs that occurred at least in 80% of all iterations provided the basis to define the network used. Relative building losses are estimated as the medians of the conditional probability distributions of the *brloss* node in the network. The discrete Bayesian Network is derived using *bnlearn package, R version 3.3.1* (Scutari, 2009).

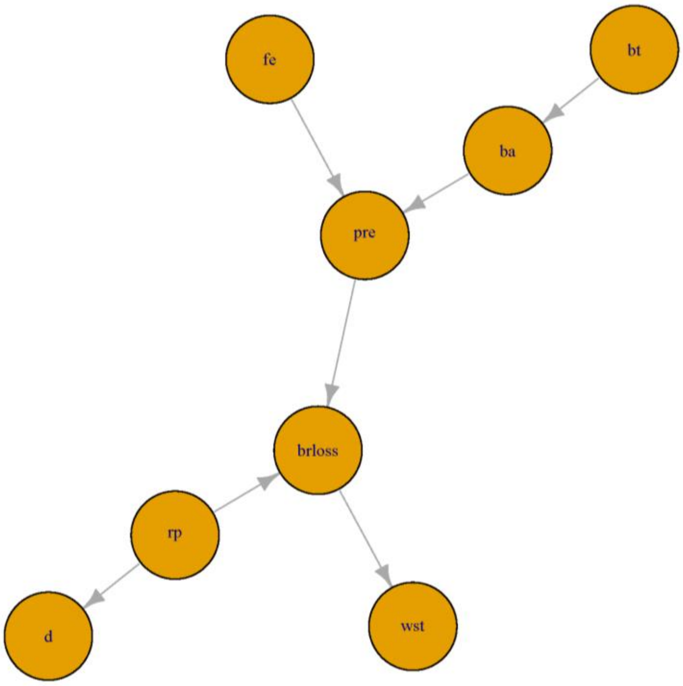


Figure 2.2: Structure of the Bayesian Network: BN-FLEMOps (Wagenaar et al., 2018)

Table 2.4: Summary of the flood loss estimation models

Model	Variables	Type
FLEMOps+r (Elmer et al., 2010)	Inundation depth, return period, building value, building type, building quality, precautionary and contamination indicators.	Point estimate
Regression Trees: RT-FLEMOps (Merz et al., 2013)	inundation depth, return period, duration of inundation, flood experience, precautionary measure, building area and building type	Point estimate
Bagging Decision Trees: BT-FLEMOps (Kreibich et al., 2017a)	inundation depth, return period, duration of inundation, flood experience, precautionary measure, building area and building type	Point estimate (Ensemble approach)
Bayesian Networks: BN-FLEMOps (Wagenaar et al., 2018)	inundation depth, return period, duration of inundation, flood experience, precautionary measure, building area and building type	Distribution function

2.2.3.2 Model performance – loss estimation and vulnerability differences

The performance of the tested Flood Loss Estimation Models is evaluated using,

1. Accuracy of the models in estimating flood losses to buildings,
2. Vulnerability differences due to private precaution accounted by the models.

Two, point estimate accuracy indictors – RMSE (Root Mean Square Error) and MBE (Mean Bias Error) from 1000 bootstrap iterations of the overall sample of households are used for assessing accuracy in loss estimation. The influence of vulnerability differences due to private precaution on the model outcome is captured by introducing an intervention for private precaution, i.e. forcing the model to consider two scenarios: (1) pre>0: all households have implemented one or more private precaution measures (treatment), (2) pre=0: all households have no precaution (control). The scenarios are applied to determine the model loss estimates for the matched households. The differences between the averages of the loss estimates obtained from the two scenarios is the differences in vulnerability due to private precaution, captured by the models.

$$\text{Difference in vulnerability accounted by loss models (pre)} = \mathrm{E}(\text{loss estimate} \mid \text{pre} > 0) - \mathrm{E}(\text{loss estimate} \mid \text{pre} = 0) \qquad \textit{Equation 2.7}$$

Where, pre represents the precautionary measure indicator of respective models.

2.3 Results and Discussion

2.3.1 Matching households with and without private precaution

In order to determine the effectiveness of private precaution in mitigating building loss, the data is controlled for heterogeneity due to potential confounding variables. We use pre-treatment variables pertaining to the households (Table 2.2) for removing selection bias from the survey dataset and then perform post-matching regression using variables pertaining to the flooding and response scenarios (Table 2.3). 948 households with no missing confounding variables undergo the matching procedure. Households with propensity scores in the common support region (equation 2.4) [0.07, 0.93] between treatment and control groups are considered for matching. This results in 32 households outside the common support and 916 households within the common support, which are considered for propensity score matching.

PSM and MDM distance estimates for selection bias combined with six different methods of matching (section 2.2.2.2.) result in twelve potential matched datasets. The following three datasets pass the quality checks for suitable matches (<10% standard error and insignificant bias, as described in section 2.2.2.3) –

1. PSM – NN with caliper and no replacement
2. PSM – NN with caliper and with replacement
3. PSM – Genetic matching

The summary statistics of Propensity Scores of households from the overall dataset, common support and suitable matched datasets are provided in table 2.5. In the appendix, section 2.5.2 summarizes the imbalance in covariates before and after matching.

Table 2.5: PS - summary of overall and matched datasets

Dataset	**Criterion** / **Precaution**	**Sample size**	**Min**	**Median**	**Mean**	**Max**
Overall	Treatment	454	0.07	0.64	0.61	0.99
	Control	494	0.05	0.31	0.36	0.93
Households in Common Support	Treatment	425	0.07	0.61	0.58	0.92
	Control	491	0.08	0.32	0.36	0.93
PSM – NN with caliper (Households in matched dataset)	Treatment	248	0.07	0.46	0.46	0.92
	Control	248	0.09	0.45	0.46	0.93
*PSM – NN with caliper and with replacement (Households in matched dataset)	Treatment	352	0.07	0.56	0.56	0.92
	Control	203	0.08	0.55	0.56	0.93
*PSM – Genetic Matching	Treatment	425	0.07	0.61	0.58	0.92

(Households in matched dataset)	Control	197	0.08	0.59	0.56	0.90

*Estimates are adjusted for weights created during matching

2.3.2 Differences in vulnerability due to Private Precaution – Empirical estimate

Vulnerability reduction of households due to private precaution is estimated as the ATE of private precaution undertaken. The ATE estimates with standard deviation in brackets for all three suitable matched datasets are provided in table 2.6. A scenario with no post-matching regression and a simple linear model are also included for reference. It is evident that Bagging Decision Trees provide a better estimate of ATE with least standard deviations than using a linear model for post-matching regression or no post-matching regression at all. Thus, our best estimate of ATE of private precaution is between 11238 and 15053 EUR. Detailed results of post-matching regression along with the regression tables and feature importance plots from Bagging Decision Trees are provided in appendix - section 2.5.3. Despite the fact, that building loss of households with and without private precaution and ATE estimates are based on empirical data controlled for pre-treatment confounders and post-treatment loss influencing variables, there might still be alternative explanations for precautionary measures being associated with reduced building losses. Building loss may differ due to damage that a frequently affected household had not repaired after a previous flood. Also, bias may still be present due to specific building characteristics for which the approach has not controlled for. However, Rosenbaum's sensitivity bounds for robustness of the estimated ATEs confirm that ATE of private precaution is unlikely to be sensitive to unobserved confounders (Table 2.7). The monetary loss reduction of 11238-15053 EUR is equal to approximately 27% of the average losses across all the households (47769 EUR). An average 27% loss reduction due to general private precaution is lower than the reported 50% reduction in median residential building loss comparing the 1993 and 1995 Rhine floods, which was attributed to a considerable general increase in the implementation of private precautionary measures (Bubeck et al., 2012). It is also in the lower range of loss reduction due to wet and dry flood proofing presented in the review of Kreibich et al., (2015). However, these studies hardly controled for confounding variables. The generalized effectiveness of private precaution of 11238-15053 EUR is comparable with the avarage treatment effects of individual private precautionary measures reported by Hudson et al., (2014): 14385 EUR for flood adapted use and 11302 EUR for flood adapted interior fitting. However, due to the survey design, we do not have households, where the precautionary measures fully prevented loss, e.g. water barriers, which hindered water to reach the building. Hence, the contribution of flood barriers to the generalized effectiveness of private precaution is not quantified in the analysis.

Table 2.6: ATE estimates from matched datasets

Post-matching regression models	**ATE estimate from matched datasets in EUR**		
	PSM – NN with caliper	**PSM – NN with caliper and with replacement**	**PSM – genetic matching algorithm**
No post-matching regression model	-26097 (6372)	-29305 (6639)	-16474 (5304)
Linear regression	-17025 (5713)	-21850 (5750)	-14330 (4541)
Bagging Decision Trees	**-12217 (2608)**	**-15053 (2947)**	**-11238 (2348)**

Table 2.7: Rosenbaum's bounds for ATE of private precaution

Matched Dataset	Γ where ATE becomes statistically insignificant (p-value > 0.05)	Γ where ATE brackets to zero (treatment effect=0)
PSM – NN with caliper	2.4	2.0
PSM – NN with caliper and with replacement	2.4	2.0
PSM – genetic matching algorithm	2.0	1.8

2.3.3 Assessment of Flood Loss Models

The comparison of flood loss models described in section 2.2.3.1 is provided in table 2.8. All tested models perform relatively similar in predicting building loss, with the Bagging Decision Tree model (BT-FLEMOps) showing the lowest RMSE and MBE and the Bayesian Network model BN-FLEMOps showing the highest errors. To test the ability of the models to capture differences in vulnerability, we evaluate how close the model-based ATE estimates are to the empirical ATE estimate by comparing the model results obtained for both the vulnerability groups using Equation 2.7.

Only two of the models result in a significant ATE for the implementation of private precaution. The ATE estimates from FLEMOps+r and BN-FLEMOps are 12185 EUR and 14687 EUR, respectively (Table 2.8), which fall within the range of the empirical estimates (11238-15053 EUR). Both models have been developed through a combination of expert knowledge and analysis of empirical data and explicitly take into consideration the direct influence of private precaution (Elmer et al., 2010; Wagenaar et al., 2018). The rule-based model FLEMOps+r considers precaution in the second model step together with contamination and recurrence interval. BN-FLEMOps has private precaution indicator (pre) in the Markov blanket of relative building loss (brloss) (Figure 2.2). This implies that, in BN-FLEMOps, predictions of building relative loss are directly influenced by private precaution. The treatment and control interventions of private precaution in tree-based models do not result in significant differences in loss estimations between vulnerability scenarios for households. The tree-based models are developed exclusively based on association inferences from empirical data, not using expert knowledge. In RT-FLEMOps, the precautionary measure indicator appears only once in the bottom of the tree and also the variable importance plot of BT-FLEMOps reveals a low importance of precaution. Hence, the influence of private precaution on the estimation of building loss is superseded by the effect of other, more important variables. Thus, the building loss estimates of the tree-based models result in an insignificant difference in losses between the two groups of households.

Table 2.8: Comparison of Flood Loss Models

Flood Loss Models	**ATE estimate from matched datasets in EUR** (Vulnerability reduction due to private precaution)**				**Relative loss estimation accuracy**	
	PSM – NN with caliper	PSM – NN with caliper; with replacement	PSM – genetic matching	Mean	RMSE	MBE
FLEMOps+r	-13497 (4202)	-11997 (4188)	-11060 (4307)	**-12185***	0.122	0.001
RT-FLEMOps	-742 (3806)	-763 (3486)	-657 (4127)	-721	0.122	0.000
BT-FLEMOps	-3816 (3448)	-3619 (3149)	-3557 (3867)	-3664	**0.116**	**0.000**
BN-FLEMOps	-14502 (3744)	-15142 (3609)	-14416 (4239)	**-14687***	0.130	0.002

* Significant ATE estimate (p-value <=0.05)

** Standard errors corresponding to the ATE estimate are provided in brackets

2.4 Conclusions

We provide robust evidence from a rigorous statistical analysis of a large empirical dataset that implementation of private precaution reduces residential building loss with an Average Treatment Effect of 11238-15053 EUR currently in Germany. More generally, this confirms previous results that undertaking private precaution is an effective means to reduce vulnerability of households against floods. Our methodology implements matching confounders of private precaution using two distance estimates and six matching methods. From this, three matched datasets are obtained with no significant bias between covariates of households with/without private precaution. Each step in the implemented methodology is customized and tested for its appropriateness for matching flood loss predictors influencing private precaution.

Dynamic risk assessments that account for the differences in vulnerability are necessary for efficient climate-based adaptation in FRM. Since, flood loss estimation models are crucial to quantify risk, it is important that these models appropriately capture differences in vulnerability, including private precaution. Only two of the tested models are able to capture these differences, these are the rule-based FLEMOps and the expert Bayesian Network based BN-FLEMOps models. In comparison with the tree-based data mining models, the accuracy with which these models predict flood losses are lower. The estimate of ATE and model performances are limited to Germany. Hence, one direction for further research could be assessing data- and model-based quantification of vulnerability due to private precaution in a spatial transferability scenario. It is also evident from the assessment of flood loss models that further research to account for the aspects of dynamic risk without compromising on prediction accuracy is required. Possible other directions in research would include developing better graphical models based on expert knowledge complemented by machine learning algorithms (Hugh et al., 2010). These models should represent causal relationships amongst potential flood loss estimators and also provide model-based scenarios of flood vulnerability.

2.5 Appendix

2.5.1 Questionnaire

The Questionnaire has been translated from German to English. Questions from the survey that are relevant to the analysis are presented here along with some statements/previous questions providing the required context.

2.5.1.1 Treatment – private precaution

Q: Now we would like to ask you about action you have taken or will take to prevent flood damage or recover from it: Which of the following precautionary and recovery measures did you implement before the flood, after the flood, are you planning to implement within the next 6 month or are you not planning to implement at the moment? Please answer with either “before the flood”, “after or during the flood”, “soon (within 6 month)” or “not planned”. INT: Measures that are planned to be implemented later than 6 months from now are counted as “not planned”.

a) Using the floors at risk (including the basement) for purposes with low values, to reduce the damage in case of a flood event

b) Avoiding expensive, built-in interior furnishing on the floors at risk. Instead using water-resistant or easy replaceable materials.

c) Sealing the basement walls

d) Installing mobile and built-in water barriers to avoid water intrusion into the building

2.5.1.2 Outcome – building loss

Context (previous question): Now we get to the building damage, directly caused by the flood. Please amplify the damage to the building structure on basis of the following list. I will read the list, please state all damage types incurred.

a) Moisture penetration;

b) Small cracks, pushed in doors and windows, or replacement of construction components required;

c) Big cracks, subsidence or deformation of walls and ceilings;

d) Collapse of building parts (Walls, Ceilings)

e) Collapse of the building

f) Demolition required

Q: If you sum up all costs of wages and material for all necessary repair work at the building. How big was the total damage, which had occurred at this building? INT: Please note the total amount. Make sure that the amount is in EURO! Important: this question asks for the total sum of ALL repairs, this includes costs of materials, hiring equipment, etc.

2.5.1.3 Confounders of private precaution

1. *Building Characteristics*

Q: How would you value the quality standard of your house BEFORE the flood? Please use a scale from 1 to 6, where 1 means “very good condition, luxury” and 6 means: ”very bad, in need of redevelopment”

Q: What is the floor area of your building? INT: This means the amount of area taken up by a building.. Can be entered as square meters or l x w.

Q: What description fits your building type best?

a) detached single-/two-family house (i.e. relatives living together)

b) semi-detached house (i.e. just two neighbours)

c) row home, flat

Q: How many flats are in the building?

2. <u>Socioeconomic Characteristics</u>

Q: Are you the tenant or owner of your house/apartment? INT: People who are partly owning the house/apartment are counted as owners (e.g. joint ownership by couples, community association etc.)

a) I am renting

b) I am the owner of the apartment

c) I am the owner of the house

Q: How old are you?

Q: How many people are permanently living in your household, including you and your children?

Q: What is the approximate monthly total net income of your household in EURO? This means the sum of all incomes after deduction of taxes, social insurance etc. INT: For farmers, freelances etc.: How high is your net income of your household after deduction of your overheads

a) less than 500€

b) 500€ to 1000€

c) 1000 € to 1500 €

d) 1500 € to 2000 €

e) 2000 € to 3000 €

f) 3000 € and more

<u>Score computation for income:</u> households falling in categories (a) to (f) are coded as class 11 to 16.

3. *Flood experience and awareness*

Q: Did you have knowledge about the flood hazard? Yes/No

Q: How often was the house in which you were living affected by flood damage? This can also be a different house then the one he/she is currently living in.

a) never before
b) once
c) twice
d) thrice
e) four times
f) more than four times

Q: When was the last flood that you experienced?

a) More than 25 years ago
b) 10 to 25 years ago
c) 5 to 10 years ago
d) 2 to 5 years ago
e) 2 or less years ago

Q: What was your amount of damage of previous flood event?

Flood experience:

a) Points for number of floods experienced:

No flood experience - 0

One previous flood – 1

Two previous floods - 2

Three previous floods - 3

Four previous floods - 4

More than four previous floods - 5

b) Points for time since last flood:

No flood experience - 0

More than 25 years ago – 1

10 to 25 years ago - 2

5 to 10 years ago - 3

2 to 5 years ago - 4

2 or less years ago - 5

c) Amount of damage of previous flood event

EUR 1000 or more – 1

Less than EUR 1000 – 0.7

<u>Score computation for flood experience:</u> $(c) \times (a + b)$

Q: Do you involve yourself in neighbourhood network activities for flood preparedness? Yes/No

Q: Do you have a flood insurance? Yes/No

Q: Private precautionary measure can considerably reduce flood damage: On a scale from 1 to 6 whether you agree or disagree with this statement. 1 means "I completely agree", 6 means "I completely disagree".

2.5.1.4 Variables for Post-matching regression

Q: How high was the water level in highest affected floor? Water level is measured based on the flooring? INT: Please enter value in centimetres; which floors of your house were affected at the maximum water level? Outhouses and detached garages are not taken into consideration here; how many steps lead to the ground floor?

Q: For how many hours in total did you have water in your house? (This means the time from the moment the water entered the building until the water drained or was pumped out of the house.

Q: Did the flood water have one or more of the contaminations:

a) Oil

b) Sewage

c) Chemical

d) Gasoline

<u>*Score computation for contamination:*</u> *0: No contamination; 1- Sewage contamination; 2 – Oil, chemical and/or gasoline contamination*

Q: Please characterize the flow velocity in close proximity to your house. Please choose a number between 1: “slow-flowing” to 6 “wild and rapidly”.

Context: “The following questions refer to the activities undertaken after receiving the first warning or becoming aware of the hazard”

Q: Which of the following emergency measures did you implement before the flood?

a) Saving documents and valuables

b) Put movable content upstairs

c) Safeguard oil tanks

d) Pump out water

e) Safeguard domestics animals and pets

f) Protect building against inflowing water

g) Redirect water on the property

h) External help (fire brigade etc.)

i) unplug electronic devices/ secure power outlets

j) Secure heating, fuse box, doors etc.

k) switch off gas/electricity

l) gas/electricity switched off by municipality

<u>*Score computation for emergency measures:*</u> $2 \times (b) + 2 \times (c) + 5 \times (d) + 5 \times (f) + 2 \times (g) + 1 \times (j)$

Q: Which of the following information was included in the warning? INT: Please read all answers and check the correct ones. Multiple answers possible.

a) Information regarding severe weather: time and region
b) Information regarding severe weather: expected volume of rainfall
c) Information regarding flood warning: water levels, time and depth of max water level
d) Information regarding flood warning: affected areas
e) Rules of behaviour and suggestion for self-protection (e.g. switch of electricity in the house, move valuables to higher floors, close all windows and doors, use sandbags, drive away your car, bring your garden furniture in a safe place)
f) Information regarding evacuation
g) Information regarding bursting dikes
h) Information regarding detours, closed roads

<u>Score computation for warning information:</u> $2 \times (a) + 2 \times (b) + 2 \times (c) + 2 \times (d) + 4 \times (e)$

2.5.2 Matching Quality

Table 2.9: Summary statistics of matching quality

Dataset	**Precaution**	**Treatment**			**Control**			**Standardized differences**	**p-value of two-sided weighted t-test**
	Covariates	***min***	***mean***	***max***	***min***	***mean***	***max***		
Overall sample	bq	1	2.23	5	1	2.24	6	-1.97	0.72
	bv	98496	523006	7354839	115450	503277	10411183	4.54	0.47
	inc	11	14.36	16	11	14.27	16	6.4	0.25
	own_1	0	0.75	1	0	0.84	1	**-13**	**0.01**
	own_2	0	0.45	1	0	0.8	1	**-12**	**0.01**
	age	20	55	89	16	53	91	**11**	**0.05**
	hs	1	2.8	10	1	2.88	11	-0.76	0.88
	kh	0	0.8	1	0	0.56	1	**60.2**	**0**
	fe	0	2.76	10	0	1.13	10	**56.5**	**0**
	neigh_ind	0	0.44	1	0	0.22	1	**43.1**	**0**
	nfb	1	1.84	23	1	1.72	45	6.9	0.29
	ins_ind	0	0.44	1	0	0.44	1	**22.6**	**0**
	ba	24	773.76	3000	30	4278	3800	**-28.6**	**0.004**
	bt1	0	0.22	1	0	0.19	1	7.6	0.17

	bt2	0	0.2	1	0	0.21	1	-5.1	0.37
	epre	1	2.65	6	1	3.17	6	**-33**	**0**
PSM – NN with caliper (Matched dataset)	bq	1	2.3	4	1	2.28	4	8.63	0.34
	bv	147097	624184	7354839	122110	461367	2004750	-0.53	0.95
	inc	12	14.37	16	11	14.25	16	1.87	0.83
	own_1	0	0.93	1	0	0.92	1	3.77	0.68
	own_2	0	0.93	1	0	0.91	1	8.26	0.39
	age	25	55	87	21	51	75	3.79	0.67
	hs	1	3	8	1	2.89	6	0	1
	kh	0	0.85	1	0	0.54	1	0.87	0.92
	fe	0	2.76	10	0	1.25	7	-5.69	0.53
	neigh_ind	0	0.42	1	0	0.17	1	3.48	0.70
	nfb	1	2.42	23	1	1.56	7	2.60	0.78
	ins_ind	0	0.60	1	0	0.44	1	9.35	0.21
	ba	30	270	3000	48	2178	3700	1.88	0.81
	bt1	0	0.27	1	0	0.18	1	7.62	0.38
	bt2	0	0.25	1	0	0.22	1	-0.96	0.91
	epre	1	2.62	6	1	3.17	6	4.10	0.66
PSM – NN with caliper and with replacement (Matched dataset)	bq	1	2.34	6	1	2.23	4	1.03	0.91
	bv	147097	625453	7354839	122110	498688	3127410	-7.81	0.08
	inc	12	14.44	16	11	14.2	16	-8.21	0.36
	own_1	0	0.90	1	0	0.90	1	0.10	0.91
	own_2	0	0.89	1	0	0.89	1	2.98	0.74
	age	25	54.75	87	21	52.19	91	6.46	0.46
	hs	1	3.05	8	1	2.81	6	-8.02	0.41
	kh	0	0.84	1	0	0.53	1	-5.39	0.53
	fe	0	2.84	10	0	0.91	7	-5.84	0.50
	neigh_ind	0	0.39	1	0	0.14	1	9.93	0.13
	nfb	1	2.40	23	1	1.93	25	-9.77	0.38
	ins_ind	0	0.54	1	0	0.45	1	-2.87	0.75
	ba	30	233	3000	48	225	3000	7.38	0.17
	bt1	0	0.27	1	0	0.19	1	-5.57	0.54
	bt2	0	0.25	1	0	0.28	1	0.69	0.94
	epre	1	2.64	6	1	3.22	6	0.71	0.94
PSM – Genetic Matching (Matched dataset)	bq	1	2.23	4	1	2.30	5	7.64	0.37
	bv	147420	656043	7354839	122110	480178	3127410	9.40	0.15
	inc	12	14.38	16	12	14.24	16	-9.35	0.19
	own_1	0	0.89	1	0	0.90	1	-1.91	0.82
	own_2	0	0.88	1	0	0.90	1	-5.62	0.49
	age	25	54.4	78	21	52.55	80	1.54	0.85
	hs	1	3.08	7	1	2.73	6	5.05	0.51
	kh	0	0.83	1	0	0.53	1	9.62	0.24
	fe	0	2.85	10	0	1.15	7	9.71	0.06
	neigh_ind	0	0.49	1	0	0.16	1	8.11	0.34
	nfb	1	2.29	23	1	1.60	9	9.89	0.06
	ins_ind	0	0.61	1	0	0.43	1	-5.69	0.50

	ba	30	230	1000	48	303	1416	0.25	0.98
	bt1	0	0.27	1	0	0.20	1	9.89	0.06
	bt2	0	0.27	1	0	0.18	1	-2.285	0.79
	epre	1	2.57	6	1	3.09	6	3.85	0.65

Significant standardized errors (>10%) and p-value estimates (<=0.05) are represented in bold. In the overall sample, out of 17 covariates, significant imbalance is observed in 9 covariates. Three matched datasets are obtained with zero imbalance among all covariates.

2.5.3 Post-matching regression

2.5.3.1 Regression Tables and feature importance plots

Table 2.10: Coefficients of post-matching regression (linear) for the three matched datasets: (a) PSM - NN with caliper; (b) PSM - NN with caliper and with replacement; (c) PSM – Genetic Matching

Coefficients	Estimate	Std. Error	t value	Pr(>|t|)
Intercept	-328.1	8636.3	-0.038	0.96971
wst	166909.2	18052.9	9.246	<2.00E-16
pre	-17607.2	5937.6	-2.965	0.00317
con	19384.4	9350.7	2.073	0.03869
d	32500.6	20942	1.552	0.12133
v_ind	-4958.6	5947.6	-0.834	0.40485
em	-12409.7	12484.7	-0.994	0.32072
wi	17723	14960	1.185	0.23672

Coefficients	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-4206	8581	-0.49	0.624173
wst	172966	16371	10.565	<2.00E-16
pre	-21616	5769	-3.747	0.000198
con	39288	8626	4.555	6.47E-06
d	22842	19291	1.184	0.236895
v_ind	-1369	5562	-0.246	0.805709
em	-13539	11554	-1.172	0.241791
wi	20218	12816	1.578	0.115258

Coefficients	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	917	6680	0.137	0.89086
wst	146167	13381	10.924	<2.00E-16
pre	-14292	4586	-3.116	0.00192

con	36633	6882	5.323	1.43E-07
d	22152	16991	1.304	0.1928
v_ind	-4057	4347	-0.933	0.35098
em	-23215	8786	-2.642	0.00844
wi	18880	9947	1.898	0.05817

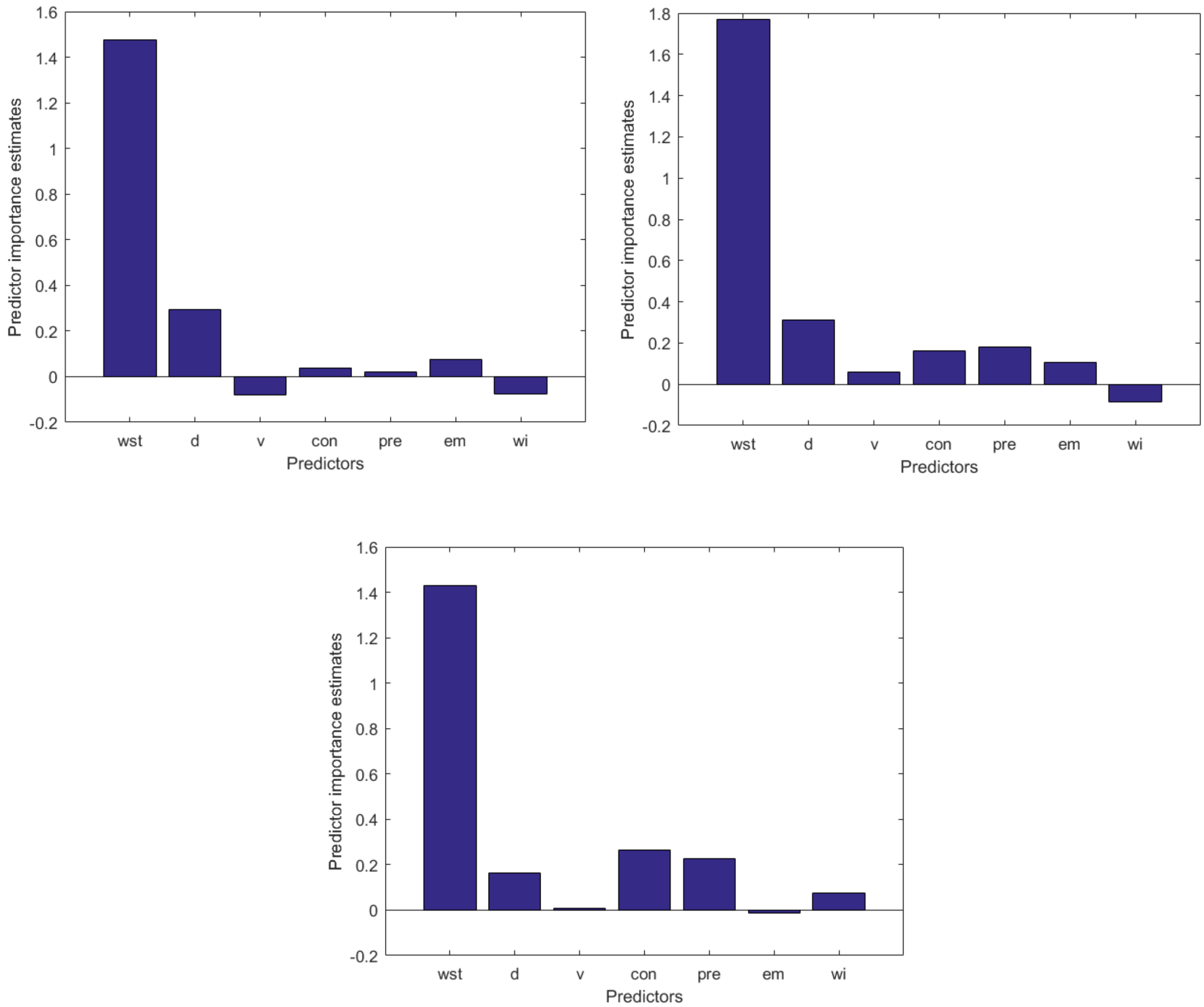


Figure 2.3: Feature importance plots for the three matched datasets. (a) PSM - NN with caliper; (b) PSM – NN with caliper and with replacement; (c) PSM – Genetic Matching (clockwise)

2.5.3.2 Results

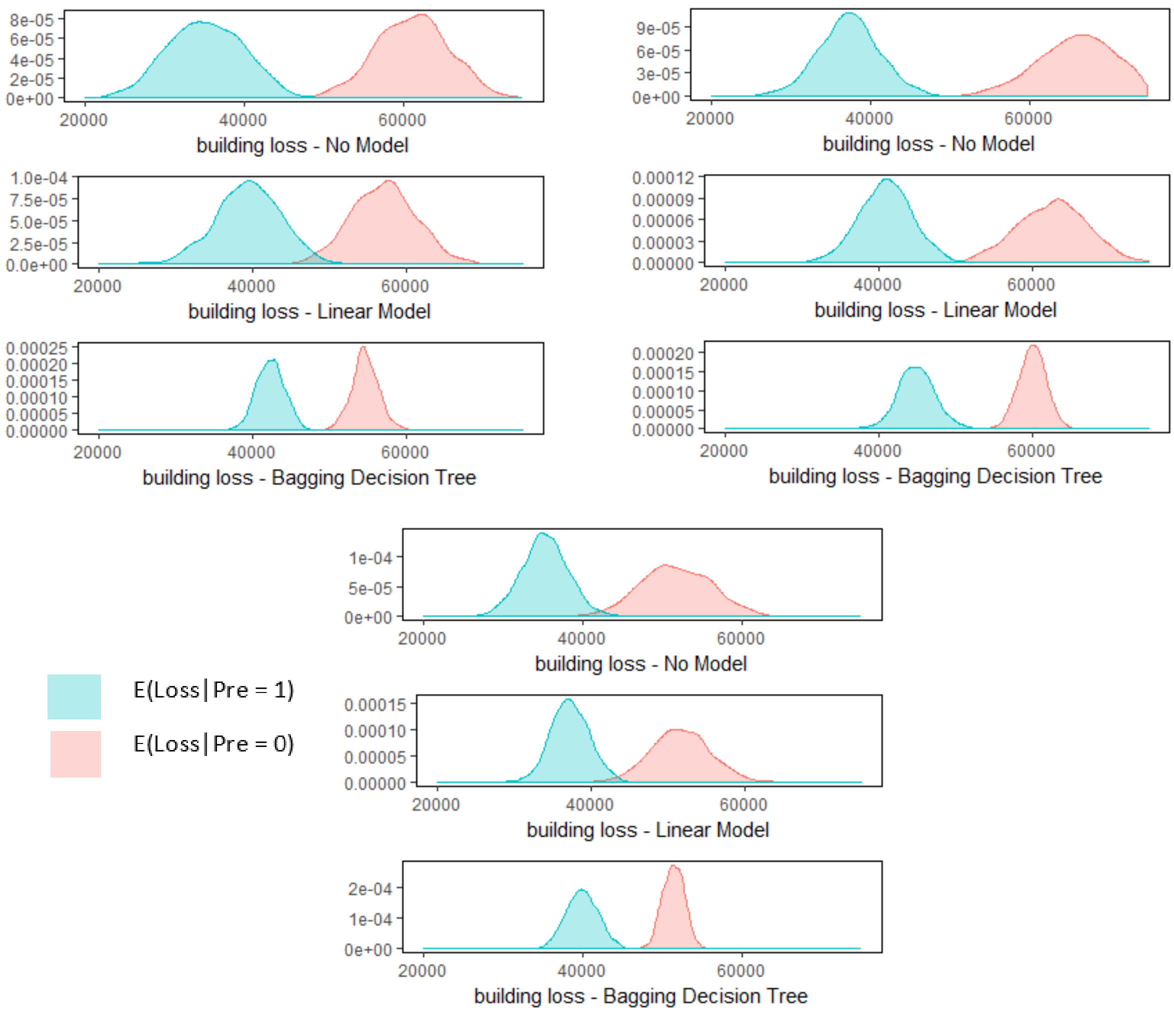


Figure 2.4: Results of post-matching regression: Expected building loss of households (EUR) from (a) PSM - NN with caliper; (b) PSM – NN with caliper and with replacement; (c) PSM – Genetic Matching (clockwise)

Acknowledgements: 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 were 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), no. 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.

3 Probabilistic Models Significantly Reduce Uncertainty in Hurricane Harvey Pluvial Flood Loss Estimates

Summary

Pluvial flood risk is mostly excluded in urban flood risk assessment. However, the risk of pluvial flooding is a growing challenge with a projected increase of extreme rainstorms compounding with an ongoing global urbanization. Considered as a flood type with minimal impacts when rainfall rates exceed the capacity of urban drainage systems, the aftermath of rainfall-triggered flooding during Hurricane Harvey and other events show the urgent need to assess the risk of pluvial flooding. Due to the local extent and small-scale variations, the quantification of pluvial flood risk requires risk assessments on high spatial resolutions. While flood hazard and exposure information are becoming increasingly accurate, the estimation of losses is still a poorly understood component of pluvial flood risk quantification. We use a new probabilistic multivariable modelling approach to estimate pluvial flood losses of individual buildings, explicitly accounting for the associated uncertainties. Except for the water depth as the common most important predictor, we identified the drivers for having loss or not and for the degree of loss to be different. Applying this approach to estimate and validate building structure losses during Hurricane Harvey using a property level data set, we find that the reliability and dispersion of predictive loss distributions vary widely depending on the model and aggregation level of property level loss estimates. Our results show that the use of multivariable zero-inflated beta models reduce the 90% prediction intervals for Hurricane Harvey building structure loss estimates on average by 78% (totalling U.S.$3.8 billion) compared to commonly used models.

Published as: 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

3.1 Introduction

Quantifying the future economic risk of pluvial flooding is critical for climate change adaptation of an increasing urban population. Pluvial, or often referred to as surface water flooding, is directly caused by extreme rainstorms with rainfall rates exceeding the capacity of the urban drainage system. Cities around the globe have been impacted by recent pluvial flood events. Large-scale pluvial flooding in the Houston area in Texas during Hurricane Harvey has led to 68 deaths and estimated total damage in the range of U.S.$90 to 160 billion, making it the second most expensive natural disaster in the history of the United States (Blake & Zelinsky, 2018). Other examples include flooding after a rainstorm in Copenhagen 2011 causing total economic damage of U.S.$1 billion (Wojcik et al., 2013) or in Beijing 2012 causing total economic damage of U.S.$1.86 billion and 79 fatalities (Wang et al., 2013). An increasing pluvial flood risk caused by an expected increase of intensity and frequency of heavy precipitation events (Donat et al., 2016; Kundzewicz et al., 2014) combined with an ongoing urbanization with a concentration of population and assets in cities (Skougaard Kaspersen et al., 2015) motivates the need to assess the current and future risk of pluvial flooding. A review by Rosenzweig et al., (2018) identified the lack of knowledge in the quantification of present and future pluvial flood impacts as one of three key research areas for the development of flood resilient cities. However, pluvial flood risk is mostly excluded or neglected in flood risk analysis, although there is evidence that the high frequency of these events leads to long-term cumulative losses comparable to less frequent but severe flood events (Veldhuis, 2011). This lack of knowledge includes FRM and mitigation plans. With few exceptions, official flood hazard maps are exclusively focused on fluvial and coastal flood risk. For the conterminous United States, Wing et al., (2018) found that the poor coverage of urban catchments in flood hazard maps produced by the Federal Emergency Management Agency (FEMA) has led to an underestimation of the population affected by pluvial and fluvial flooding by a factor of 2.6–3.1. With scarce information on the hazard, only few loss estimation models for pluvial floods have been developed. Existing approaches include adapting water depth-damage functions (also known as stage-damage models) from river floods (Freni et al., 2010; Olsen et al., 2015; Zhou et al., 2012), using multiple linear regression models (Van Ootegem et al., 2015), or by correlating rainfall measurements with insurance claims or survey data (Spekkers et al., 2014; Van Ootegem et al., 2018). However, the lack of data, the complex nature of the hazard and impact as well as the lack of a consistent quantification of the associated uncertainties, has so far hampered an extensive estimation of expected pluvial flood losses needed to decide on adaptation strategies in cities. Van Ootegem et al., (2015, 2018) construct different multivariate pluvial flood damage models from survey data of a study in Belgium based on water depth-damage and rainfall-damage relationships. Key findings of their study include the importance of additional non-hazard variables such as risk awareness and the effect of reported zero loss cases. However, the results do not provide information as to whether additional variables can also improve loss estimates.

In this study, we use probabilistic high-resolution loss models to estimate pluvial flood losses on different spatial scales. Unlike widely used deterministic stage-damage functions, these probabilistic univariable and multivariable loss models provide a consistent approach to quantify how certain a loss prediction is by providing predictive distributions instead of point estimates. Application and validation of different high-resolution probabilistic loss models in Harris County, Texas, reveal significant differences in the dispersion and reliability of property and county level pluvial flood loss predictions for Hurricane Harvey. Only two out of the six tested models reliably predicted the reported loss with a difference of 78% in the 90% prediction intervals between the two models equalling to an absolute difference of U.S.$3.8 billion for pluvial flood building structure loss in Harris County. These results have major implications for cost-benefit analysis of FRM and adaptation decisions in cities.

3.2 Background

With the need to adapt cities to an expected increase in pluvial flood risk, decision makers face the challenge to take appropriate decisions under the uncertainty of how the risk of pluvial flooding evolves in the future including the expected losses. As uncertainties in flood losses estimates are usually high, probabilistic loss models could greatly aid a comprehensive pluvial FRM (Todini, 2018). Unlike deterministic estimates, probabilistic predictions provide continuous predictive distributions where the dispersion of the distribution can provide the range an expected loss would fall in with a certain probability (e.g., 90%). The reliability of a probabilistic prediction can be expressed as the ability of the predictive distribution to cover the actual observed loss. Although probabilistic loss models have been developed for river floods (Dottori et al., 2016; Kreibich et al., 2017a; Schröter et al., 2014), these models are the exception and deterministic estimates based on empirical or synthetic relationships between the water depth and the absolute or relative building loss are still widely used for loss estimations for all types of flooding (Gerl et al., 2016; Merz et al., 2010b; Scawthorn et al., 2006). The resulting loss estimates in these stage-damage functions are commonly expressed as point estimates for the repair and/or replacement costs in monetary values (i.e., U.S.$) or percentage of the depreciated value of a building. Instead of a direct quantification of uncertainty inherent to probabilistic predictions, uncertainty in stage-damage functions is often based on expert judgment and/or by calculating a range of possible outcomes using different loss functions (Dittes et al., 2018). Missing information, and/or a lack of consistency in quantifying how certain a loss estimate is, makes it challenging for decision makers to, for example, evaluate the potential of an adaptation measure to reduce future losses. While the deviations of point estimates for deterministic loss models are often shown to be reasonably small for loss estimates on large spatial scales typical for river or coastal flooding, loss predictions become highly uncertain on smaller scales (i.e., individual buildings; Merz et al., 2004; Scawthorn et al., 2006). However, due to the local extent and small-scale variations, reliable small-scale loss models are required to quantify pluvial flood risk for a specific location. In this context, we use machine learning as well as different univariable and multivariable probabilistic approaches to investigate three main research objectives: we (i) identify important loss influencing variables and their effect on the uncertainty of loss predictions; (ii) analyze the potential of parametric and nonparametric probabilistic approaches on reducing the dispersion and increasing the reliability of building-level loss estimates; and (iii) evaluate the applicability of probabilistic multivariable loss models in the context of new sensors and data sources for pluvial flood loss estimation on different spatial scales (Ford et al., 2016; Schröter et al., 2018).

3.3 Materials and Methods

3.3.1 Data

We construct a data set that consists of self-reported pluvial flood losses and related information of private households. The data were obtained through a standardized questionnaire using computer-aided telephone surveys after pluvial flood events in five cities in Germany between 2005 and 2014 (Rözer et al., 2016; Spekkers et al., 2017). Based on 120 items in the questionnaire, a data set with 56 predictors and two loss variables is constructed covering eight groups: reported loss, hazard, warning, emergency response, precaution, experience, building information, and social-economic information. The loss variables are represented as relative loss (rloss) and a variable with binary information if a building suffered from structural damage or not (dam). rloss is on the scale from 0 (no loss) to 1 (total loss), normalizing the reported direct building loss in Euro [EUR] with the total replacement cost value less depreciation of the respective building. We exclude observations where rloss could not be derived due to missing information on the building replacement value or the reported loss itself resulting in a total of 431 observations. Out of 56 predictors in the data set, 12 are excluded from the analysis, because of their zero or near-zero variance, resulting in 44 variables to be considered for further analysis. To address the issue of censoring zero loss observations, pluvial flood affected households without direct building loss are included in the data set if water intrusion into the building was reported (9% of observations; see Van Ootegem et al., 2015). Missing values in other variables were imputed using complementary information available in the questionnaire (i.e., missing information of the total living area through building footprint and number of habitable floors). In few cases where causal inference was not possible, missing values are imputed using nearest neighbor imputation. A more detailed description of the data including a table describing all 56 predictors, the two loss variables, the variables excluded from the analysis, and the percentage of imputed missing values is provided in the supporting information (SI; Data section).

3.3.2 Detection of Important Loss Influencing Variables

Prior to the actual model development, we screen the previously described data set for variables with the highest predictive power given the complex correlations and interdependencies in the data set using machine learning. A reduced set of variables out of the full 44 variables is then used to develop the multivariable probabilistic models described in the following section. The most important loss influencing variables are detected by using an ensemble of variable importance measures of two tree-based (Bagging [cRF; Strobl et al., 2007] and Boosting [GBM; Friedman, 2001]) and two linear regression-based (Ridge [Hoerl & Kennard, 1970 and LASSO [Tibshirani, 1996]) machine learning algorithms. The four different types of algorithms are used in two different settings: a binary classification between loss/no loss (dam) and a regression analysis modeling the degree of loss (rloss) of a building. Based on the variable importance score of each variable, its rank within each ensemble member as well as its overall rank is determined. The top five variables with the highest overall rank for rloss and dam are further considered in the model development process. For details on the variable selection procedure, see SI (Materials and Methods section).

3.3.3 Probabilistic Loss Estimation Models

Bayesian zero-inflated beta regression (Ospina & Ferrari, 2010) is used to predict the relative loss to a building by pluvial flooding (rloss) using the previously selected important loss influencing variables. The probabilistic prediction y for rloss on the interval [0,1) is modeled as follows: We define z_i to be a binary variable for the occurrence of flooding in the i^{th} observation and estimate it with a logistic regression:

$$z_i \sim Bernoulli(\gamma X_i) \qquad \textit{Equation – 3.1}$$

where X_i is the vector of predictors for the i^{th} observation, γ is the vector of coefficients, and Bernoulli(θ) indicates a Bernoulli trial with probability θ. Once z_i is known, then we can calculate y_i following a zero-inflated Beta regression model

$$y_i = \begin{cases} Beta(\alpha_i, \beta_i), & z_i = 1 \\ 0, & z_i = 0 \end{cases} \qquad \textit{Equation – 3.2}$$

where $\alpha_i > 0$ and $\beta_i > 0$ are the shape and scale parameters, respectively, of the Beta distribution. To estimate these parameters, we define

$$\alpha_i = \mu_i \varphi \qquad \textit{Equation – 3.3}$$

$$\beta_i = (1 - \mu_i)\varphi \qquad \textit{Equation – 3.4}$$

following Ferrari and Cribari-Neto (2004). This parameterization allows us to define

$$\mu_i = X_i \beta \qquad \textit{Equation – 3.5}$$

where β is the coefficient vector for the Beta regression. In summation, our zero-inflated beta regression model conducts simultaneous inference on the vector γ, the vector β, and the scalar ϕ, given observations of flood occurrence z, flood damage y (i.e., the variable *rloss*), and predictive variables X.

The probabilistic predictions of *rloss* from the Bayesian zero-inflated beta model (*Beta*) are compared with probabilistic predictions of two additional model types used for empirical flood loss estimation in previous studies. A simple Bayesian parametric model based on a *Gaussian* response distribution is used as a probabilistic representation of a model type widely used in flood loss estimation (Gerl et al., 2016; Van Ootegem et al., 2015) and a nonparametric model based on the *RandomForest* algorithm, used for probabilistic flood loss estimation in previous studies (Schröter et al., 2016). The three model types (*Beta*, *Gaussian*, and *RandomForest*) are fit as univariable and multivariable models (i.e., with a single predictor in X or with multiple predictors) to investigate the effect of additional variables on the predictive performance, resulting in six different models in total. The univariable models are fit using water depth *wd* as their only predictor, reflecting the current standard in flood loss estimation (Gerl et al., 2016; Merz et al., 2010b). The univariable parametric models (*Beta* and *Gaussian*) are fit with the square root of the water depth to be comparable with reference functions in previous studies (Merz et al., 2013; Schröter et al., 2014; Wagenaar et al., 2017). All multivariable models use the set of predictors shown in Table 3.1. For more details on the models including details on the priors of the Bayesian models, see SI (Materials and Methods section).

Table 3.1: Mean Variable Importance Scores of the Five Most Important Predictors for rloss and dam on the Scale (0, 100) for Each Ensemble Member (Tree-Based Bagging [cRF] and Boosting [GBM]; Penalized Regression with L1 [LASSO] and L2 [Ridge] Regularization)

Name	Variable	cRF	GBM	LASSO	Ridge	Avg. rank	Corr
Degree of loss (*rloss*)							
Water depth	wd	100^{1}	100^{1}	94^{1}	97^{1}	1	+
Duration	d	38^{2}	50^{2}	81^{3}	90^{2}	2	+
Basement [Y/N]†	bu	12^{9}	11^{13}	84^{2}	85^{3}	6	+
Contamination [Y/N]	con	15^{8}	9^{17}	77^{4}	81^{4}	6	+
Household size†	hs	17^{4}	17^{8}	45^{7}	64^{5}	6	-
Loss/no loss (*dam*)							
Water depth	wd	99^{1}	100^{1}	89^{1}	90^{2}	1	+
Household size	hs	84^{2}	14^{2}	67^{3}	93^{1}	2	-
Knowledge hazard	pre1	72^{3}	6^{4}	48^{7}	81^{3}	3.5	-
Age of respondent†	age	69^{4}	13^{3}	3^{32}a	42^{9}	6.5	+
Multifamily home [Y/N]	bt	49^{7}	1^{11}a	50^{6}	51^{6}	6.5	-

Note. Corr indicates direction of the trend: "+" increasing; "-" decreasing. Superscript numbers indicate rank within each ensemble member. Avg. rank indicates the overall rank based on the median rank of each ensemble member. Variables marked with a "†" showed no improvement in the predictive performance of the probabilistic loss models and were therefore not considered in the final models. [a] Importance scores not stable.

3.3.4 Model validation and comparison

We validate the probabilistic loss predictions on the building level for the previously described models and data using 10-fold cross validation. For determining the error of the point estimate (median of the predictive distribution), the root-mean-square error (RMSE) and the mean bias error (MBE) are used. For validating and comparing the reliability of the loss estimate, we calculate the hit rate (HR), meaning the percentage of cases where the observed value lies within the 90% highest density interval (HDI) of the predictive distribution. We use the width of the 90% HDI to evaluate the dispersion of the predictive distribution. In addition, we calculate the interval score, a combined dispersion and reliability score, penalizing predictions based on the width of the 90% HDI and the percentage of observations that are outside the 90% HDI of the respective predictive distributions (Gneiting and Raftery, 2007). To evaluate the effect of including the option to have no building loss in the model, we validate and compare the different models for three scenarios: one where zero-loss observations are removed from the data set prior to fitting the model, one where the zero-loss observations are kept in the data set (zero-loss proportion 9%), and one where the proportion of zero-loss observations is up sampled to 20%. Details on the validation procedure and the different scores used to compare the models are provided in SI 3.6.2.4 - Model comparison and scoring methods.

3.3.5 Application Harris County, TX

We apply the previously trained probabilistic loss models in Harris County, TX, to analyze the potential for reducing the dispersion and improving the reliability of probabilistic loss estimates for direct building damage of private households caused by pluvial flooding during Hurricane Harvey. To demonstrate the feasibility of probabilistic building-level loss estimation, we construct a high-resolution data set from publicly available data sources for Harris County, TX.

Based on refined pluvial flood inundation maps for Hurricane Harvey provided by JBA Risk Management (JBA Risk Management, 2017), detailed information of affected properties are gathered from the Harris County Appraisal District Real & Personal Property Database including the type and value of each affected building (HCAD, 2018). In addition, census information is used to derive the average household size on the block level (U.S. Census

Bureau, 2016). Besides this information, the constructed data set contains data on the knowledge about the flood hazard based on if a property is within the 100-year flood zone derived by FEMA (Zone A) and the probability of a property being affected by contamination. The contamination data was created by spatially interpolating reported point sources of contamination from the National Response Center of United States Coast Guard and volunteered geographic information using 2-D kernel density interpolation (NRC, 2018; Sierra Club, 2017). The resulting dataset for Harris County contains information of more than 304 000 individual buildings affected by pluvial flooding during Hurricane Harvey. For validation and visualization the property level loss distributions of each model are aggregated on the zip code as well as on the county level. The aggregated loss estimates are validated using the sum and average total building damage from FEMA's Housing Assistance Program available on the zip code level as well as for the entire county for Hurricane Harvey (FEMA, 2018a). Details on the data sets and models used in Harris County including the validation data are provided in SI 3.6.1.3 - Harris County data.

3.4 Results

3.4.1 Important Loss Influencing Variables

Screening the high-dimensional data set for the most important loss influencing variables to be considered in the probabilistic loss models, we find that the drivers for having loss or not having any loss (*dam*) and the drivers for the degree of loss (*rloss*) to a building are different, indicating different damaging mechanisms. While both cases share the water depth as the most important predictor, other important predictors hardly overlap. Looking at the second to fifth most important predictors for *dam*, the resistance of a building and its inhabitants is decisive. Given a low inundation depth, larger households, multifamily buildings, younger residents, and residents who previously informed themselves about pluvial flooding have a lower probability of having any loss. In contrast, the second and fourth most important predictors influencing *rloss* are directly related to the flood intensity. Higher inundation depths, longer flood duration, and contamination of the flood water lead to higher losses. The variable importance scores of the five most important predictors of the four machine learning algorithms their rank within each ensemble member and the median rank of all ensemble members are summarized in Table 3.1. Starting with the most important predictor both the overall rank and the importance scores drop sharply. Of the five preselected important loss influencing variables shown in Table 3.1, we find three variables for *rloss* and four variables for *dam* to improve the predictive performance in the probabilistic loss models. Variable importance values for all 44 variables and differences between the machine learning algorithms are shown in SI (Results section).

3.4.2 Predictive Performance of Probabilistic Models

The prediction performance of the six probabilistic models (univariable and multivariable models for *Gaussian*, *RandomForest*, and *Beta*) for the cross-validated predictions are summarized in Table 3.2. Looking solely on the error of the point estimate of the predictions (median of the predictive distribution), we find only a minor nonsignificant reduction in root-mean-square error for the three models for both the univariable and multivariable versions. However, for the 90% HDI of each predictive distribution, the parametric *Beta* and *Gaussian* models are significantly more reliable with an average HR of 97% and 95% for the univariable and multivariable *Beta* models and 91% for both *Gaussian* models compared to 67% and 49% for the *RandomForest* counterparts. However, when we control the HR of the predictive distributions for dispersion and distance to missed observations using the interval score, the high HR scores of the *Gaussian* models can be attributed to consistently wider 90% HDI's (see Figure 3.1b) compared to the other two models. The difference in shape and width of the predictive distributions of the different models is illustrated in Figure 3.1a, for the example of a loss estimate for a single building with an observed *rloss* of 0.016. While the *RandomForest* models tend to give very sharp predictive distributions with shapes close to a normal distribution, the predictive distributions of the *Gaussian* and *Beta* models both have longer tails. The almost lognormal shape of the *Gaussian* models is caused by the back transformation of the logit-transformed predictive distribution. Although the sharp predictive distributions of the *RandomForest* models lead to considerably narrower prediction intervals it significantly increases the risk of the 90% HDI not covering the actual observed loss (see Table 3.2). With its flexibility in shape and clearly defined interval of the response distribution, we find the *Beta* models to provide the best trade-off between reliability and dispersion. Compared to the widely used reference function (univariable *Gaussian*), the univariable and multivariable have between 47% and 50% narrower HDI's with HRs above 90%. Comparing the difference between the univariable and multivariable models, we find an increase in the variability in shape and width of the predictive distributions for all multivariable models. Although this increase in variability only show a minor, nonsignificant improvement in accuracy, reliability, and dispersion (see Table 3.2), we find that multivariable models perform significantly better compared to models using the water depth as only predictor when individual predictions are aggregated (see Figure 3.3c).

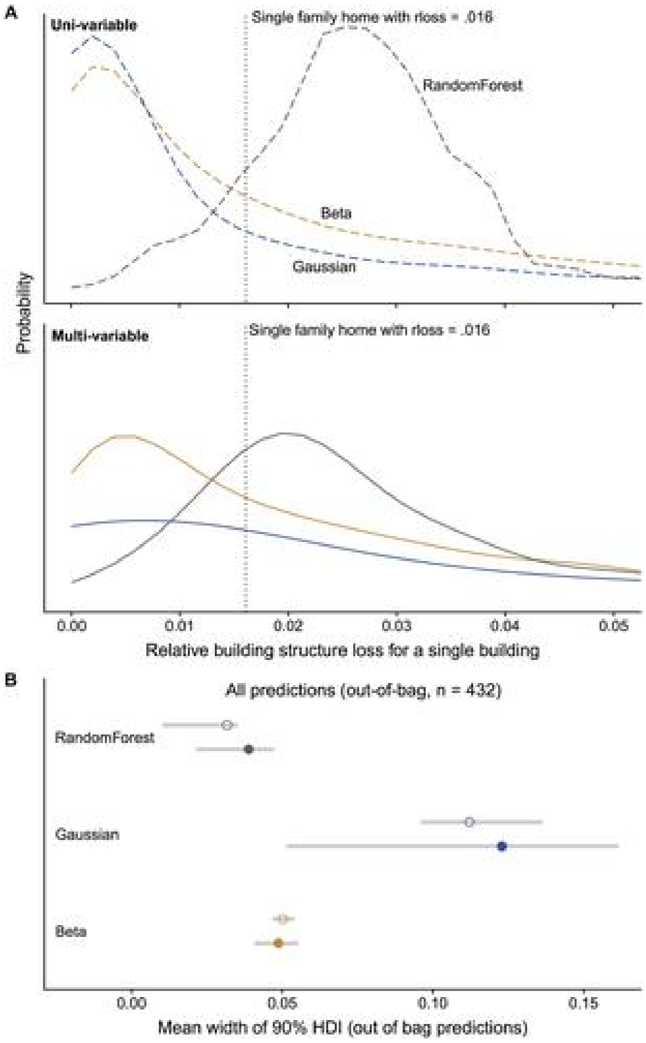


Figure 3.1: Probabilistic predictive distributions of different univariable and multivariable models (RandomForest, Gaussian, and Beta) for cross-validated observations. The predictive distributions for *Gaussian* and *Beta* models are based on 2000 MCMC samples from the respective posterior predictive distributions. The predictive distributions from *RandomForest* model are based on the predictions of 2,000 individual trees used for training the forest. (a) The different predictive distributions for a single household (single-family home) with a recorded relative loss of 0.016 (dotted vertical line). The upper plot of (a) shows the predictive distributions for three univariable models using the water level as only predictor (dashed lines). The lower plot of (b) shows the same three model types, but with five additional predictors (solid lines). (b) The widths of the 90% HDI for the predictive distributions of all cross-validated observations ($n = 431$) are summarized. The points show the medians for the univariable (hollow) and multivariable (solid) models for the three different model types. The gray boxes show the 25th to 75th percentile ranges for each model. HDI = highest density interval.

Table 3.2: Performance of Loss Model Predictions for Out of Sample Observations (Median)

Model type	Variables	RMSE	MBE	Hitrate (90% PI)	Interval Score (90% PI)
Gaussian	univariable	0.028 (0.018)	0.015 (0.008)	0.91 (0.01)	0.26 (0.01)
	multivariable	0.027 (0.017)	0.013 (0.007)	0.91 (0.02)	0.25 (0.02)
Random Forest	univariable	0.028 (0.017)	0[a] (0.009)	0.49[a, b](0.07)	0.17[a] (0.11)
	multivariable	0.025 (0.016)	0.005 (0.008)	0.67[a, b](0.08)	0.11a (0.08)
Beta	univariable	0.027 (0.017)	0.010 (0.008)	0.97 (0.06)	0.09[a] (0.08)
	multivariable	0.025 (0.017)	0.009 (0.008)	0.95 (0.07)	0.08[a](0.08)

Note. Standard deviation in brackets. RMSE = root-mean-square error; MBE = mean bias error.

[a] Significantly different from Gaussian model for the 0.05 significance level (univariable and multivariable models, respectively).

[b] Significantly different from univariable models for the 0.05 significance level for each model type.

3.4.3 Effect of Zero-Loss Cases on the Damage Estimates

The often low water levels of pluvial flooding compared to river or coastal flooding increases the chances that direct building loss can be completely avoided, although water entered the building. Analyzing different zero-loss proportions, we find that not explicitly accounting for these cases can considerably affect model predictions in terms of reliability and dispersion of the predictive distribution. For the *Gaussian* models, none, and for multivariable *RandomForest* model, 28 of the 38 zero-loss observations in the data set were inside the respective 90% HDI. For increasing the zero-loss proportions we observe a significant increase in the reliability of the *RandomForest* model and a significant increase in the width of the 90% HDI of the loss prediction for the *Gaussian* model (Figure 3.2). The increase in reliability of the *RandomForest* model reflects the capability of the model to learn implicitly to account for zero-loss cases, when the learning sample becomes large enough. Without the possibility to consider zero-loss cases, a higher proportion of zero-loss observation simply adds additional variability, which the *Gaussian* models cannot explain. Bias caused by varying zero-loss proportions is found to be reduced to a minimum by explicitly accounting for zero-loss observation in the (zero-inflated) *Beta* models (see *Beta* model in Figure 3.2). Findings for the univariable models are, for the sake of readability, shown in SI (Results section).

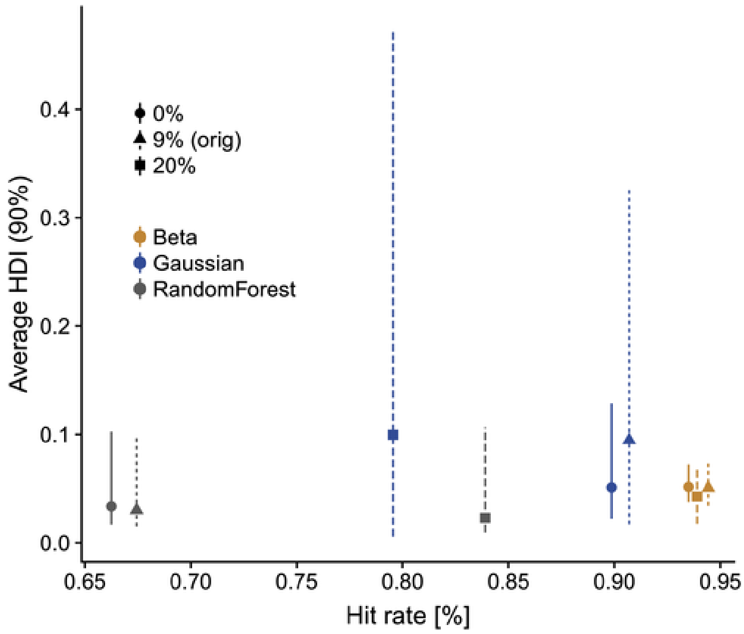


Figure 3.2: Trade-off between reduction in uncertainty and reliability for cross-validated predictions for different multivariable loss models and different proportions of zero-loss observations in the data set. Results for univariable models are shown in SI (Results section). Uncertainty is represented as mean width of the 90% HDI for all observations. Reliability is represented as proportion of the out-of-sample observation, which are inside the respective 90 % HDI. Error bars represent the 90% interval for the HDI width of all out-of-bag predictions. HDI = highest density interval.

3.4.4 Hurricane Harvey Building Loss for Harris County, TX

Modeled direct losses to the building structure caused by pluvial flooding during Hurricane Harvey in Harris County, TX, are summarized in Figure 3.3. Our main finding is that the width of the 90% HDI of the predictive distribution for individual buildings can be reduced by 21% or U.S.$3,685 on average when using the multivariable *Beta* model instead of the univariable *Gaussian* model representing the current standard in empirical flood loss estimation. Panel (b) shows the mean relative reduction in the width of the 90% HDI between the two models for individual buildings on the zip code level. For individual buildings we find spatial variations for the average building structure loss ranging from U.S.$544 to U.S.$10,134 with the majority of areas being in the range of U.S.$2,000 to U.S.$5,000. The highest average building structure loss with values above U.S.$7,500 are found west and southwest of Downtown Houston (panel a).

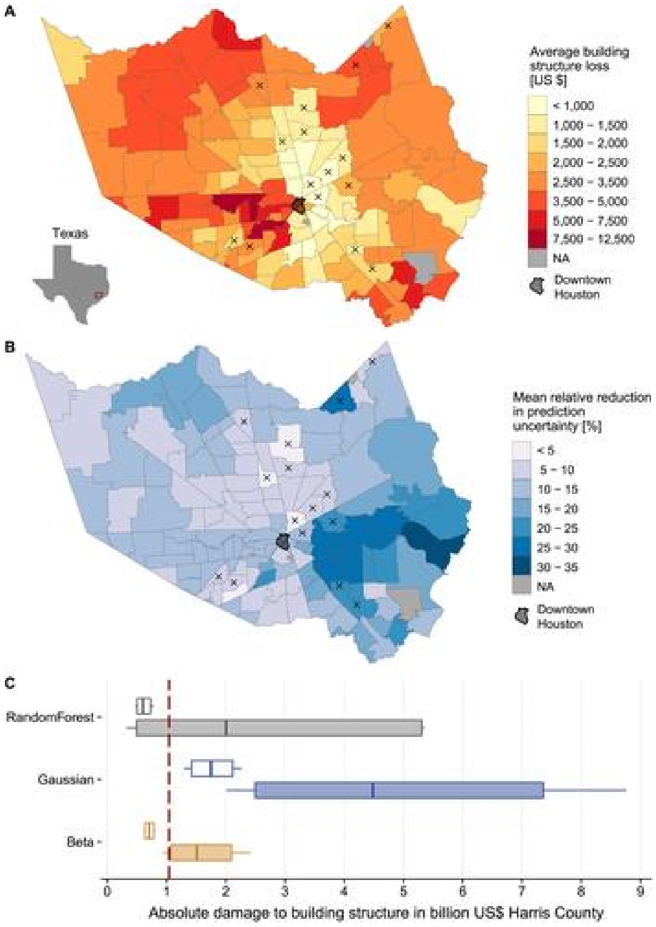


Figure 3.3: Modeled direct building structure losses for Harris County, TX, caused by pluvial flooding during Hurricane Harvey. (a) The modeled average building structure loss per building aggregated on the zip code level using the multivariable *Beta* model. (b) The average relative reduction in uncertainty (expressed through the width of the 90% HDI) per building between the univariable *Gaussian* model (reference function) and the multivariable *Beta* model in percent aggregated on the zip code level. Crosses in (a) and (b) indicate zip code areas where the reported average building loss is outside the 90% HDI of the modeled average building loss. (c) Box plots of the aggregated predictive distributions of the absolute direct building structure damage for the entire county for three different model types (*RandomForest*, *Gaussian*, and *Beta*) in their univariable (hollow) and multivariable (solid) versions. Bars indicate the median absolute loss, boxes the 90% HDI, and whiskers the 98% HDI of the absolute direct building loss for Harris County. The red dashed line represents the official reported absolute building structure loss based on data from the Federal Emergency Management Agency Housing Assistance Program. HDI = highest density interval.

For the aggregated predictive distribution of the absolute loss to the building structure of over 304,000 affected residential buildings (single-family and multifamily homes) in Harris County, the corresponding samples of the individual predictive distributions of each building are summed up. This leads to an effect, known as the central limit theorem, where the Beta-distributed predictive distributions for individual buildings coming from the *Beta* model tend to form a normal distribution when enough individual predictive distributions are summed. In combination with a higher variability, introduced by the additional variables, the considerably higher reliability and lower dispersion of the multi-variable *Beta* model compared to the univariable *Gaussian* model on the building-level vanishes when the predictions are aggregated over a large amount of individual buildings (panel c).

This effect is also described by Sieg (2019) and provides further evidence why univariable stage damage functions based on *Gaussian* response distributions yield sufficiently accurate loss predictions on larger scales while the same model produces highly uncertain loss estimates on the building level. For results aggregated to the county level, we find univariable and multivariable *Gaussian* models to overestimate the absolute building structure losses by U.S.$0.7 and U.S.$3.4 billion, respectively. This can be partly attributed to the underestimation of zero-loss cases described in the previous chapter, which leads to higher intercepts in the model. For the multivariable model this effect is considerably stronger as the model is fit as a linear instead of a square root function (see section 3.3.3). Of the six models tested, none of the univariable models, and only the aggregated predictive distributions of the multivariable *RandomForest* and *Beta* models are covering the reported loss from FEMA's Housing Assistance Program (U.S.$1.04 billion). Here the multivariable *Beta* performs significantly better with a total reduction in width of the 90% HDI of U.S.$3.8 billion (or 78%) compared to the multivariable *RandomForest* model, providing the best trade-off between dispersion and reliability.

3.5 Discussion and Conclusions

Despite causing severe losses in cities around the globe, pluvial flooding is still widely neglected when estimating the current and future flood risk in urban areas. This results in a widespread underestimation of flood risk especially in urban areas where fluvial or coastal floods are not the dominant sources of flooding (Rosenzweig et al., 2018). One key limitation in reliably quantifying pluvial flood risk is the local extend of pluvial floods, requiring loss estimates on spatial scales where damaging processes are still hardly understood and the associated uncertainties are often unknown. We present the first consistent quantification of uncertainties in pluvial flood loss models for private buildings in the shape of predictive distributions using a fully probabilistic modeling approach. We train and validate different univariable and multivariable probabilistic loss models with a local training data set and use these models for a probabilistic estimate of building structure losses of over 304,000 individual buildings in Harris County during Hurricane Harvey. Our analysis reveals significant differences in the dispersion and reliability of the continuous predictive distributions between different models depending on (i) the use of additional predictors, (ii) the choice of response distribution, (iii) the ability of the model to account for zero-loss cases, and (iv) the spatial scale of the analysis. We find that the assumption of a normal or lognormal distribution of uncertainties in loss estimates, which most loss models implicitly use today, results in unnecessarily wide prediction intervals. In the case of property level predictive distributions, we find that the width of the 90% HDI exceeds the median of the prediction by factor 30 on average. Our results suggest that the width of the 90% HDI for pluvial flood loss estimates on the property level can be significantly reduced by 47% when using a zero-inflated beta distribution instead of normal response distributions without sacrificing the reliability (Table 3.2). While not evident on the property level, we find that using water depth as only predictor results in an underestimate of the prediction intervals leading to unreliable loss estimates when spatially aggregating loss predictions (Figure 3.3c). Here, we find additional predictors to improve the pluvial flood loss predictions in two ways: (i) by increasing the variability of individual predictive distributions leading to a more realistic representation of uncertainties when aggregating estimates and (ii) by improving the detection of cases where water entered the building but did not cause any monetary damage to the structure (Figure 3.2). For the latter our analysis indicates the ability of households to prevent direct damage to their homes should be included in loss models.

The analysis of important loss influencing variables has further shown that the probability of a household to not have any monetary loss to the building structure is—other than for the degree of loss—strongly influenced by household characteristics such as the number of people living in a household and their prior knowledge about the pluvial flood hazard. This highlights the need to account for differences in the ability of households to reduce or avoid damage to their homes in loss models for pluvial floods.

For loss estimates in Harris County, the use of additional predictors in zero-inflated beta models considerably increases the reliability while at the same time significantly reduces the dispersion of the predictive distribution given validation data. For direct building losses aggregated on the county level this reduction accounts for U.S.$3.8 billion or 78% compared to loss models based on normal response distributions. These findings are relevant for a larger discussion on using probabilistic loss estimates for decision making in FRM. This includes the potential of probabilistic approaches to improve the spatial transferability of loss models. We further demonstrate the potential to significantly improve the dispersion and reliability of pluvial flood loss estimates using probabilistic models, which goes beyond previous studies considering only point estimates (Van Ootegem et al., 2015; Zhou et al., 2012). Although these results are limited to a quantification of uncertainties of loss predictions, the results can easily be extended for robust decision making on adaptation strategies based on exceeding probabilities, which can be directly derived from predictive distributions. While our results suggest that models that use a zero-inflated beta response distribution provide predictive distributions with a significantly lower dispersion and higher reliability, a general paradigmatic change toward probabilistic models would greatly aid a better understanding of uncertainties in loss models (Todini, 2018). Same is true for multivariable models, where emerging cloud-based reporting systems and open data portals now allow the use of high-dimensional data sets in flood loss modeling.

3.6 Supporting Information (SI)

3.6.1 SI Data

3.6.1.1 Survey data

The data set contains information collected by computer aided telephone interviews (CATI) of private households affected by pluvial floods in 2005, 2010 and 2014 in five German cities (Table 3.3). Altogether 783 completed interviews are available from these surveys. On the basis of information from fire brigades or flood reports and press releases, lists of inundated streets were compiled for each flood event. These lists served as a basis to select telephone numbers of all potentially affected households from the public telephone directory. Computer-aided telephone interviews were undertaken by a market research institute with the help of the VOXCO software package (www.voxco.com) about 15 to 19 months after the events (Table 3.3). At the beginning of the interview, it was asked to interview the person in the household with the best knowledge about the flood event. The questionnaires used for the surveys were based on a questionnaire developed by Kreibich, Thieken, Petrow, Müller, and Merz (2005) and A. H. Thieken, Müller, Kreibich, and Merz (2005) for river floods, but was adapted for the special characteristics of pluvial flooding. The interviews lasted 25 to 30 min on average and contained approximately 110 questions on the following topics: flood impact, warning, emergency measures, evacuation, clean-up, characteristics of and damage to household contents and buildings, recovery of the affected household, precautionary measures, flood experience, and socio-economic characteristics of the household. Building loss includes all costs associated with repairing the damage to the building structure, such as plastering, replacing broken windows and repairing the heating system. The questionnaire contained detailed questions addressing not only total loss but also the affected stories, many information on the building itself necessary to estimate the building value. This generated the most accurate information possible about the flood loss. Post-processing was performed, like correcting or removing implausible inputs, for example, by comparing reported water levels inside and outside the house and by comparing reported floor areas with building footprint. More details on the surveys and dataset are provided by Rözer et al., (2016) and Spekkers, Rözer, Thieken, ten Veldhuis, and Kreibich (2017).

Table 3.3: Overview survey data

Characteristics	**Surveys**		
Survey period	Nov 2006	Feb/Mar 2012	Oct/Nov 2015
Affected cities	Lohmar, Hersbruck	Osnabrück	Münster, Greven
Event	Jun-05	Aug-10	Jul 2014
Number of households interviewed	173	100	150
References	URBAS (2008), Rözer et al., (2016)	Rözer et al., (2016)	Spekkers et al., (2017)

3.6.1.2 Relative loss

The relative loss (rloss) describes the proportion of the direct monetary damage to the structure of a building in relation to its total value. It is bounded between the interval [0,1], where 0 is equivalent to no monetary damage at all and 1 to a total loss of the building. Modeling the proportion of the direct monetary damage instead of the total values has two main advantages when modeling losses: (i) the loss estimates become independent from the actual building values, which is expected to lead to more stable relationships between rloss and the predictors. (ii) The relative loss is dimensionless, which means it creates comparable results over space and time without the bias of inflation or varying building costs in different regions (as discussed for floods in Merz, Kreibich, Schwarze, and Thieken (2010) and in the general context of natural hazards by Neumayer and Barthel (2011)). For pluvial floods the majority of the values for rloss are typically in the lower range (< .4) including cases where water entered the building in such low quantities that it did not cause any direct damage to the building structure (we refer to these cases as zero-loss observations). The bounded outcomes as well as the concentration of values at or close to 0 makes the modeling of rloss challenging in the context of (ordinary least squares) regression, which does not account for bounded intervals and therefore may lead to biased estimates. To overcome these limitations the response variable has to be either transformed to map the outcomes to the [0,1] interval or using a regressor where the response variable is assumed to be beta-distributed on the (0,1) interval (Schmid et al., 2013). For the variable selection using machine learning as well as for the stage-damage and multivariate ensemble model we use the logit-transformation to transform rloss:

$$logit(rloss) = log(\frac{rloss}{1 - rloss}) \qquad \textit{Equation 3.6}$$

This avoids nonsensical predicted values for rloss below 0 or above 1. To deal with observations where rloss = 0 that would create a transformed value of $-\infty$ we set the values for rloss = 0 to the smallest non-zero value in the dataset as suggested by Warton and Hui (2011). This provides more flexibility in the selection of different learning algorithms as machine learning for beta distributed response variables is not yet well established. For the probabilistic multi-variate damage model, where the focus of the model is on prediction, we model rloss as zero-inflated beta distribution. The quantile-quantile (Q-Q) plots in Figure 3.4 show the logit-transformation as well as the beta-distribution compared to the untransformed empiric distribution of rloss.

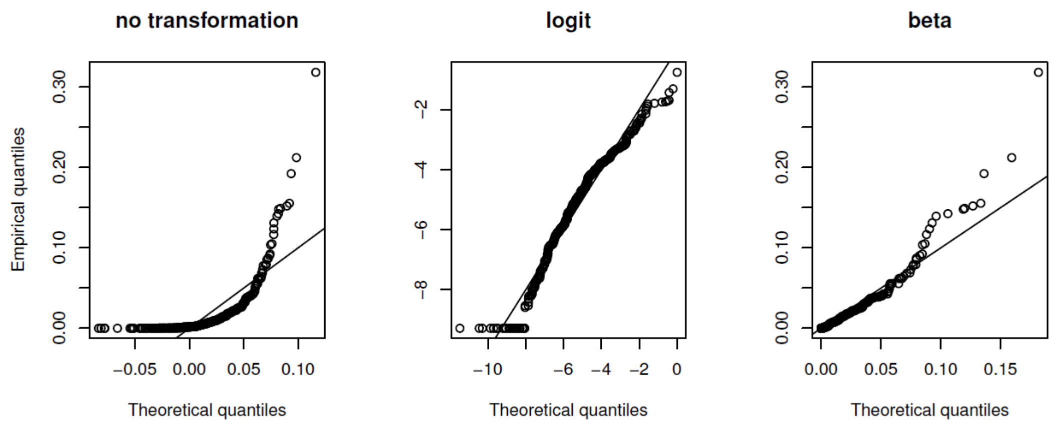


Figure 3.4: Quantile-quantile plots for different transformations/distributions for rloss

3.6.1.3 Harris County data

Based on a high-resolution pluvial flood map (spatial resolution approx. 30m) containing modeled inundation depth, we construct a multi-variable data set to be used with the proposed uni- and multi-variable probabilistic flood models. The pluvial flood map is provided by JBA Risk Management based on model runs covering the period from August 25 2017 to August 28 2017 and represents the maximum water depth per cell in this period. The entire map covers the Gulf coast from Corpus Christi, TX to Lake Charles, LA to Huntsville, TX in the north, but only the inundated areas in Harris County, TX are used for this study.

Table 3.4: Overview of all candidate variables

Group	Var.	Description	Scale	Range	Missing [%]
Damage	rloss	Relative building structure damage; Normalized with building value	c	0: damage 1: total damage; actual range: 0 - 0.4	-
	dam	Binary building structure damage	b	yes/no	-
	af1*	Basement affected	b	yes/no	< 1
	af2	Ground flood/first floor affected	b	yes/no	< 1
	af3*	Higher floors affected	b	yes/no	< 1
Hazard	wd	Water level relative to the ground level	c	-247 cm below ground to 453 cm above ground	5.6
	d	Flood duration	c	1 - 840 h	4.2
	con	Contamination with chemicals, sewage or oil	b	yes/no	4.6
	v	Flow velocity indicator	o	0: still to 6: high velocity	4.6
	rfi1h	Maximum amount of rainfall in 1 hour over the whole storm event	c	15.6 - 141.8 mm	-
Warning	ws1*	Warning source: Severe weather warning	b	yes/no	1.6
	ws2	Warning source: Friends, neighbors, family	b	yes/no	1.6
	ws3*	Warning source: National news	b	yes/no	1.6

	ws4	Warning source: Own observation	b	yes/no	1.6
	wt	Early warning lead time	c	0 - 72 h	2.3
Emergency response	em1	Saving documents and valuables	b	yes/no	< 1
	em2	Put movable content upstairs	b	yes/no	< 1
	em3*	Safeguard oil tanks	b	yes/no	< 1
	em4	Pump out water	b	yes/no	< 1
	em5*	Safeguard domestics animals and pets	b	yes/no	< 1
	em6	Protect building against inflowing water	b	yes/no	< 1
	em7	Redirect water on the property	b	yes/no	< 1
Precaution	pre1	Inform about flood hazard/protection	b	yes/no	< 1
	pre2	Participate in flood protection network	b	yes/no	2.5
	pre3	Flood insurance	b	yes/no	2.3
	pre4	Inferior use of exposed floors	b	yes/no	1.6
	pre5	Avoid expensive permanent interior	b	yes/no	2.3
	pre6*	Relocate heating / electricity to higher floors	b	yes/no	3.5
	pre7	Reduce contamination risk (protect oil tank, store chemicals in safe place)	b	yes/no	1.1
	pre8	Improve flood safety of the building	b	yes/no	2.5
	pre9	Install backflow protection device	b	yes/no	1.9
Experience	fe	Flood experience indicator based on no. of previous floods, previous damage and time since last flood	o	0: no experience to 9: recent experience with loss > 1000 Eur	2.1

	npf	Number of previous floods	c	0 - 5	< 1
Building characteristics	bt1	Building type: Multi-family home	b	yes/no	< 1
	bt2	Building type: Semi-detached house	b	yes/no	< 1
	bt3	Building type: Rowhouse	b	yes/no	< 1
	by1*	Building year: <1924	b	yes/no	< 1
	by2	Building year: 1924 - 1948	b	yes/no	< 1
	by3	Building year: 1949 - 1964	b	yes/no	< 1
	by4	Building year: 1964 - 1990	b	yes/no	< 1
	ht	Oil heating Y/N	b	yes/no	< 1
	bu	Building has basement	b	yes/no	< 1
	bq	Building quality	c	1: very good to 9: very bad	< 1
	bv	Building value	c	88440 to < 19682400	
	nfb	Number of apartments per building	c	1 - 45	1.1
	fsb	Floor space building	c	55-4900 sq.m	2.1
	bm1*	Building material: Timber frame	b	yes/no	1.2
	bm2	Building material: Steel-enforced concrete	b	yes/no	1.2
	bm3	Building material: Masonry	b	yes/no	1.2
	bm4*	Building material: Wood	b	yes/no	1.2
	tpi500	Topography index. Relative height of building location compared to surroundings. Radius 500m	c	-17.74 building below to 10.91 above surrounding areas	-
	age	Age of respondent	c	20 - 90 years	3.2

Socio-economic	hs	Number of people living in household	c	1-8 persons	1.4
	chi	Number of children <14 years in household	c	0 - 3	3.2
	eld	Number of adults >65 years in household	c	0 - 4	2.8
	own1*	Ownership status: Tenant	b	yes/no	< 1
	own2*	Ownership status: Apartment owner	b	yes/no	< 1
	own3	Ownership status: Home owner	b	yes/no	< 1

*Variables have zero or near-zero variance and are not used in the model

As the inundation depth is the result of modeling work and could only partly validated using observations, the provided inundation depths are inherently uncertain. More detailed information on the pluvial flood map are available in the meta data of the flood map and from JBA Risk Management (JBA Risk Management, 2017). Using the footprint of the maximum extent of flooding from the pluvial flood inundation map, the following additional variables are derived from publicly available data sets: estimate of the inundation duration (d), information on contamination (con), average household size (hs), knowledge of the hazard (pre1), type of building (bt) as well as the value of the building. For an estimate of the inundation duration (d), we use the revised estimated dry times provided by the Pacific Northwest National Laboratory (PNNL). The dry times are estimates based on model simulations of a 2-day hindcast and 5-day quantitative precipitation forecast and do not reflect the operational control of dams. The dry times reflect the estimated number of days the water is expected to take from its peak state to a dry state not including base flow conditions (PNNL, 2017). Point information on the contamination is derived from incidents report data base of the National Response Center (NRC) of the United States and filtered for reports in relation to Hurricane Harvey for a report period between August 27 2017 and September 9 2017(NRC, 2018). Reports not related to water pollution were excluded from the data set. In addition, a data base compiled by the Sierra Club containing a collection of national and state level reports from known incidents during Hurricane Harvey were used to validate and/or complement the NRC data (Sierra Club, 2017). In total 98 records are available. We use 2Dkernel density estimation to create a probability map reflecting the probability that an area was contaminated based on the proximity to locations were contamination was reported. The point information of contamination is only interpolated for locations that were within or close to a flooded area (< 30m) and also only within the flooded areas based on the assumption that contaminants (oil, gas, sewage etc.) are only transported through flood water. This approach does not consider flow fields of the surface water or sewage system and is only an estimate of potentially contaminated areas. To estimate the household size of a building, we use information about the average household size separated by tenants and house owners on the block level, based on the 2016 American Community Survey (ACS) (U.S. Census Bureau, 2016). The knowledge of the household about flood risk is based on the flood zone information provided by FEMA. The assumptions are made, that households lying within an area with a 1-percent annual chance of getting flooded (Zone A) are aware of the flood risk. This assumption is based on the requirement that property owners have to buy flood insurance in these areas when making, increasing, renewing, or extending a loan (FEMA, 2018b). Information on the type (bt) and value (bv) of the affected buildings can be directly derived from the property data base of the Harris County Appraisal District (HCAD, 2018). For this study we only use private single- and multi-family homes. All commercial or public buildings are not considered and excluded from the data set. For the building value we use the development replacement cost new less depreciation (RCNLD) to quantify the current replacement value of each building on the property. Based on the pluvial flood inundation map we link the flooded areas and other hazard characteristics with the exposed building. This results in a data set with a total of 304,441 affected buildings in Harris County including information on the estimated inundation in centimeter, the flood duration in hours, the probability of the building being contaminated by oil, gas, chemicals or sewage as well as several information on the household size and knowledge about the flood hazard. The modeled relative building losses are multiplied by the RCNLD to obtain loss values in US$. The loss values are validated for the zip code and county level based on reported loss values from FEMA Housing Assistance Program (FEMA, 2018a). For validation purposes only, the building structure damage of home owners are considered. As only households whose losses are not covered by insurance are eligible to receive funds from FEMA's Housing Assistance Program the validation data might underestimate the total loss when excluding insured losses. However, the underestimation is expected to be minor as FEMA estimated that only 15% of all homes (20% of flooded homes according to estimates by the Consumer Federation of America) in Harris County had flood insurance prior to Hurricane Harvey (Kunreuther, 2018).

3.6.2 SI Materials and Methods

3.6.2.1 Determining important predictors

For building an effective predictive model, the selection of input variables is a crucial step. However, when the number of variables is large, detailed exploratory analysis of all possible predictors is inefficient and often not feasible (Kuhn and Johnson, 2013). Since input data for loss estimations are scarce and often difficult to obtain, one would strive for a parsimonious loss estimation model, that optimizes the trade-off between number of predictors and predictive accuracy (Grömping, 2009). In this study, we rank all 44 candidate variables (total number of variables is 56, but only 44 considered for variable selection due to near-zero variance of 12 variables) based on the intrinsic variable importance measures of four different predictive models. The respective models are used in a regression context to find the strongest predictors for the level of relative loss (rloss) and in a classification context for the presence or absence of loss (dam). The variables that show a strong relationship with rloss and dam respectively are selected to be used as input for the probabilistic loss estimation model.

The supervised predictor selection routine is shown in Figure 3.5. We use the same routine with the same predictors and the same type of models independently to identify the predictors with a strong relationship to rloss and dam respectively. To increase the robustness of the variable ranking and compensate for recurrence issues frequently appearing in machine learning, the predictor selection is based on the variable importance measures of four different models (Dasgupta et al., 2011). The four models were selected out of a large pool of models provided in the caret package (Kuhn, 2008) based on the following criteria: (a) provide inherent variable importance measure, (b) can be used for regression and classification (c) combination of models that is able to detect linear- and non-linear relationships. For the variable importance of non-linear predictors, we use two different non-parametric tree-based ensemble models: conditional inference forests (Strobl et al., 2007) based on an ensemble of independent randomized regression-/classification trees, similar to the random forest model originally proposed by Breiman (2001); and gradient boosting machines based on a stage-wise additive model with correlated trees (Friedman, 2001). To detect predictors with possible linear relationships between the two dependent variables, we use two different penalized linear regression models: the least absolute shrinkage and selection operator (LASSO) with L1 regularization (Tibshirani, 1996) and Ridge regression with L2 regularization (Hoerl and Kennard, 1970). For the classification routine, the proportion of cases with no loss is increased from 9% to 50% through random sampling with replacement to compensate for class imbalance. Each model is tuned individually using 10-fold cross-validation with 10 repeats. That means the dataset is split and resampled to result in 100 individual training and validation datasets.

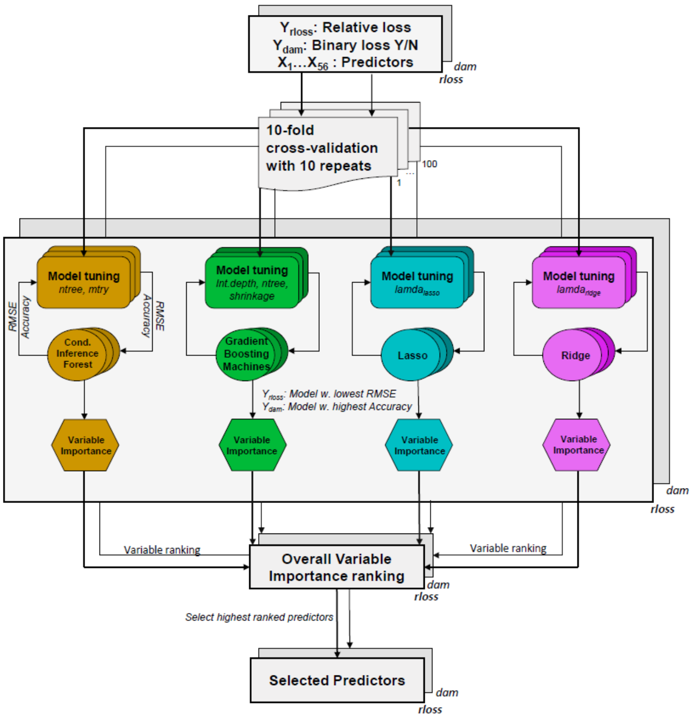


Figure 3.5: Flowchart of the machine learning routine for the variable importance measures of the level of loss (rloss) and the the classification of loss/no loss (dam). For the variable importance each of the four models are tuned for the lowest RMSE (resp. highest accuracy).

The variable importance measures are normalized and each variable is ranked in each of the four models based on their variable importance score. The most important loss influencing variables are selected based on their overall rank(median). The cross-validation routine is repeated for each tuning configuration and the optimal tuning configuration for each model is selected based on the lowest root mean square error (RMSE) for rloss and the highest mean accuracy for dam. The model configuration with the lowest RMSE (the highest mean accuracy respectively) are considered as the optimal models and for these models the variable importance of each predictor is determined. The variable importance measures are scaled from 0 (removed from the model) to 100 (most important variable) using unity-based normalization to make the scores comparable between the different models. Based on its variable importance score, each variable is ranked from 1 (most important) to 44 (least important) for each variable. The overall rank of each variable is then determined based on the median rank of all four algorithms. The top five variables with the highest overall rank for rloss and dam are then considered for the actual probabilistic loss model.

3.6.2.2 Bayesian zero-inflated beta regression

Section 3.3.3 of the main text describes the Bayesian zero-inflated Beta regression model used to develop probabilistic prediction of flood loss. Here we provide details on computation and the priors used. In Bayesian models involving empirical observations, obtaining analytical solutions for predictions are almost impossible. Hence, we estimate an approximated posterior distribution (Kruschke and Vanpaemel, 2015). Markov Chain Monte Carlo (MCMC) samplers create tens of thousands of parameter replications based on the data generation process to represent the posterior distributions. The probabilistic multi-variate flood loss model is implemented in the stan modeling language (Carpenter et al., 2017) using the brms package version 3.3.2 (Bürkner, 2017) in R using the No-U-Turn Sampler (NUTS) by Hoffman and Gelman (2014). The MCMC sampler is run with two chains, with 2000 iterations each and a burn-in period of 1000, resulting in a total number of 2000 MCMC samples for each posterior distribution. Model convergence is assessed using the Gelman-Rubin $\hat{R}$ statistic (Gelman and Rubin, 1992), which compares between-chain and within-chain variances to assess MCMC convergence. We obtained $\hat{R} < 1.1$, suggesting good convergence. We also compute the effective sample size of posterior draws, which accounts for the autocorrelation to measure the equivalent number of independent samples (Kass et al., 1998). We confirmed the ratio of effective sample size to nominal sample size to be between 0.999 and 1.001. Bayesian modeling also requires the application of a prior distribution on the parameters applied; Bayes' rule gives

$$p(\theta|y) = \frac{p(\theta, y)}{p(y)} = \frac{p(\theta)p(y|\theta)}{p(y)p(\theta)p(y|\theta)} \propto p(\theta)p(y|\theta) \qquad \textit{Equation} - 3.7$$

and the prior is just the value assigned to p(μ). Priors are often classified as uninformative, weakly informative, or informative, although these terms are not clearly defined.

We applied weakly informative priors, which we define following Gelman et al., (2017) and Simpson et al., (2017) as priors explicitly designed to encode information that applies to a general class of problems without taking full advantage of problem-specific knowledge. In other words, weakly informative priors provide coverage over all parameters which might be plausible. We do not modify the weakly informative priors provided by default in the brms package but note that results are not sensitive to alternative specifications of weakly informative priors. These default priors assume that the group-level priors (i.e. γ and β) follow a normal distribution with mean zero and unknown covariance matrix, and assigns an improper flat prior to scale parameters (i.e.φ). For further discussion of the priors applied to the unknown covariance matrix we refer the reader to Bürkner (2017).

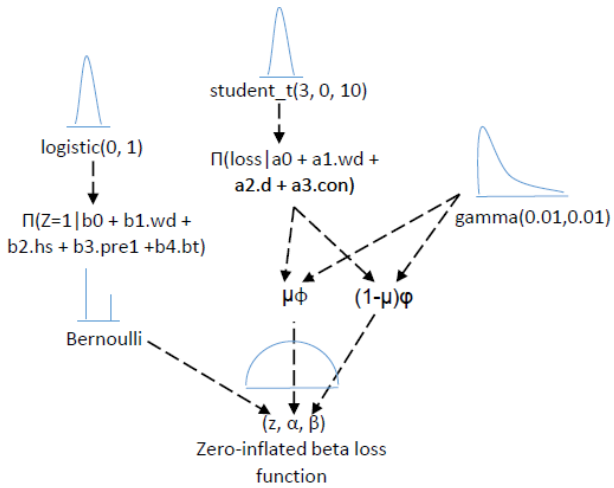


Figure 3.6: Visualization of the zero-inflated Beta model including priors for the Bernoulli and Beta parts.

3.6.2.3 Models for comparison

We compare the predictive performance of out-of-sample predictions of the uni-and multivariable Beta model with two different model types from the literature: first a stage-damage model based on a root function using the water depth as only predictor as representation of a concept that is widely used in academia and among practitioners to describe the flood vulnerability of a building (Penning-Rowsell et al., 2005; Merz et al., 2013; Scawthorn et al., 2006; Thieken et al., 2008). Root functions are used as reference damage function in (among others) Merz et al., (2013), Schröter et al., (2014) and Wagenaar et al., (2017). Second, a nonlinear, non-parametric tree-based model based on the Random Forest algorithm by Breiman (2001), representing the current state of the art for loss models (Schröter et al., 2014; Wagenaar et al., 2017). The root function follows:

$$rloss = c_1 + c_2\sqrt{wd} \qquad \textit{Equation - 3.8}$$

and is fit to the survey data set using Bayesian inference in Rstan. Each prediction consists of 2000 Markov-Chain-Monte-Carlo (MCMC) samples from the posterior predictive distribution based on the No-U-Turn Sampler (NUTS) implemented in Stan (Hoffman and Gelman, 2014). Priors for c1 and c2 are set weakly informative (normal distribution with μ= 0 and σ = 10) to avoid any bias in the prediction. Before fitting the model rloss is logit-transformed to map the bounded outcomes of rloss to the whole real line (-Inf, Inf) and to satisfy the assumptions of OLS-regression (i.e. normally distributed residuals).

The Random Forest model uses the original Random Forest algorithm by Breiman (2001) as implemented in the randomForest R package (Liaw and Wiener, 2002). The Random Forest model is learned with 2000 independent trees (ntree=2000). For each loss prediction of an individual household a predictive distribution is generated by using the mean of the respective terminal node for each of the 2000 trees. The number of trees is set to 2000 to make sure that the predictive distribution of the RandomForest model is generated based on the same number of samples as for the posterior predictive distributions of the Beta and Gaussian models. All multi-variable models are fit using the same 7 variables (see Table 3.1 in the main text, three variables for rloss, four variables for dam) used in the probabilistic multi-variable Beta model. The uni-variable Beta and Gaussian models are fit as squareroot function using the water level as only predictor. The Gaussian models are fit using logit-transformed values of rloss following the reference function used in Schröter et al., (2014). For the uni-variable Random Forest model different split points in the water level are selected based on bootstrap samples of the original dataset; the multi-variable model is fit by randomly selecting two out of six variables at each split (mtry = 2).

3.6.2.4 Model comparison and scoring methods

The predictive performance of the three previously described models are compared using 10-fold cross-validation, where for each iteration 90% of the data are used for fitting/training the models and 10% are used for prediction. For all three models the composition of the folds is the same and each observation in the dataset is used at least once for training and once for prediction. The same dataset described in Section 3.6.1 - Survey data is used for all three models. The predictive performance is evaluated in terms of accuracy of the point estimate based on the median of the predictive distribution, using the root mean squared error (RMSE) and mean bias error (MBE); the reliability of the 90% highest density interval (HDI) of the predictive distributions is evaluated using the hit rate (HR) and the dispersion of the interval using the interval score (IS) and mean width of the 90% HDI. Accuracy of point estimate (median of the predictive distribution) Q_{50i}):

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Q_{50i} - O_i)} \qquad \textit{Equation -3.9}$$

$$MBE = \frac{1}{n}\sum_{i=1}^{n}(Q_{50i} - O_i) \qquad \textit{Equation –3.10}$$

$$HR = \frac{1}{n}\sum_{i=1}^{n} h_i;\ h_i = 1\ if\ rloss_i \in HDI_{90i};\ 0\ , otherwise \qquad \textit{Equation – 3.11}$$

$$IS = HDI_{90i} + \frac{1}{n}\sum_{i=1}^{n}\frac{2}{\beta}\left(\min(HDI_{90i}) - \widetilde{rloss_i}\right)|\{O_i < \min(HDI_{90i})\} + \frac{2}{\beta}\,(O_i - \max(HDI_{90i})|\{O_i > max(HDI_{90i})\} \qquad \textit{Equation – 3.12}$$

with HDI_{90ilow} and HDI_{90iup} marking the upper and lower bounds of HDI_{90i}.

3.6.3 SI Results

3.6.3.1 Model validation for Hurricane Harvey

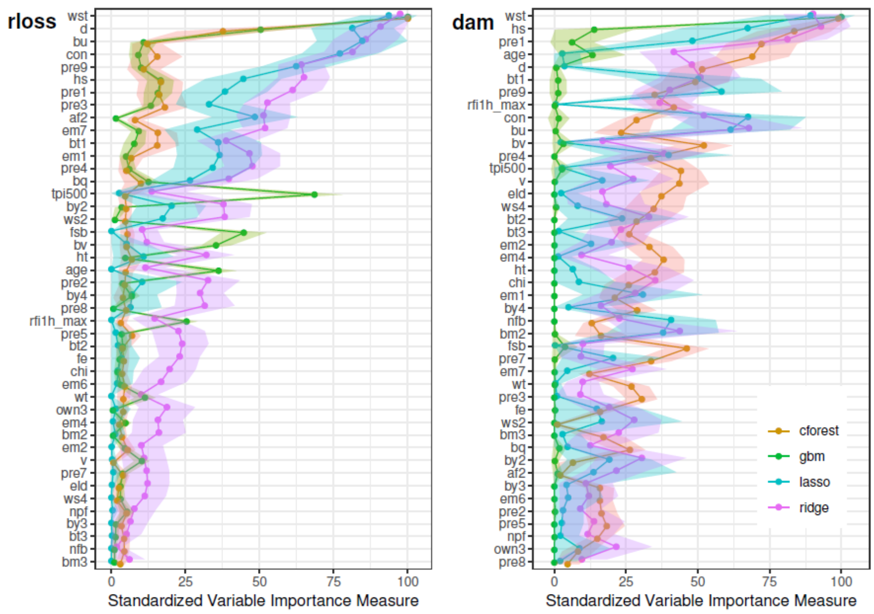


Figure 3.7: Variable importance of the 44 candidate variables using conditional inference forests (cRF), gradient boosting machines (GBM), lasso (LASSO) and ridge (Ridge) regression for the level of loss (rloss) and the classification between loss/no loss (dam). The variable importance values were rescaled for the interval (0,100) using unity normalization. The most important variable for each model was set to 100. A value of 0 means, that the variable was removed from the model (feature selection). The uncertainty bands represent the error of one standard deviation considering all 100 re-sampling rounds.

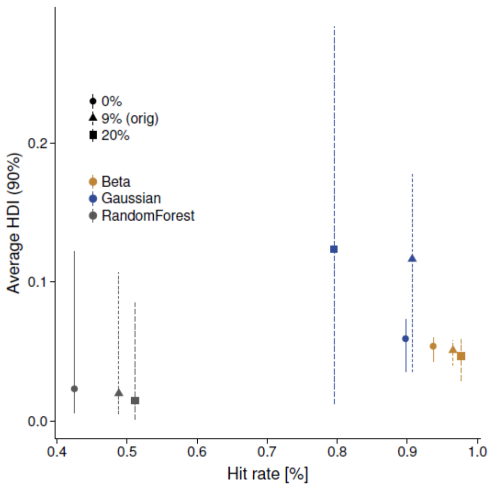


Figure 3.8: Trade-off between reduction in uncertainty and reliability for cross-validated predictions for different uni-variable loss models and different proportions of zero-loss observations in the dataset. Uncertainty is represented as mean width of the 90% HDI for all observations. Reliability is represented as proportion of the out-of-sample observation, which are inside the respective 90 % HDI. Error bars represent the 90% interval for the HDI width of all out-of-bag predictions.

Comparing the aggregated modeled predictive distributions with FEMA's Housing Assistance Program, we find that in 103 out of 136 modeled zip code areas, the reported average loss lies within the 90% HDI of the modeled average loss (76%). For the modeled total loss this is true for 112 out of 136 modeled zip code areas (82%). For the total absolute loss for the entire county only the 90% HDI of multi-variable RandomForest model covers the reported total absolute loss by FEMA. For the multi-variable Beta model, the 98% HDI cover the reported loss with a considerably sharper prediction making it the model with the highest IS.

Acknowledgments: The data collection campaigns after the flood events in Germany were supported by the project "EVUS Real-Time Prediction of Pluvial Floods and Induced Water Contamination in Urban Areas" (BMBF, 03G0846B), the University of Potsdam, Deutsche Rückversicherung AG, the German Research Centre for Geosciences GFZ and URBAS - Urban Flash Floods project (BMBF; 0330701C). We would also like to acknowledge JBA Risk Management, who supported our work by providing the pluvial flood inundation map for Hurricane Harvey. The pluvial flood inundation map from JBA Risk Management is available via the OASIS Hub (https: //oasishub.co/dataset/surface-water-flooding-footprint-hurricane-harvey-august-2017-jba). The data sets of the flood events in Germany from 2005 and 2010 are available via the German flood damage data base HOWAS21 (http://howas21.gfz-potsdam.de/howas21/). The data set from 2014 will be made available via the HOWAS21 data base in June 2023. All other data sets used for the application in Harris County, TX are openly available and cited in the text and SI. Detailed information on all data sets used for this study and how to access them are available in the Supporting Information (Section 3.6.1).

4 Bayesian Data-Driven Approach Enhances Synthetic Flood Loss Models

Summary

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.

Under review: Sairam, N., Schröter, K., Carisi, F., Wagenaar, D., Domeneghetti, A., Molinari, D., Brill, F., Priest, S., Viavattene, C., Merz, B., & Kreibich, H. Bayesian Data-Driven Approach Enhances Synthetic Flood Loss Models. Environmental Modelling and Software.

4.1 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; IPCC, 2012; Merz et al., 2012; Domeneghetti et al., 2015). Decisions concerning FRM focusing on new flood defense schemes and resilience initiatives are generally based on risk assessment encompassing of future hazard scenarios and the resulting damage. 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).

Flood loss models are an essential component of the risk chain as they quantify flood risk in terms of economic losses (Bubeck & Kreibich, 2011; Merz et al., 2010b). Flood loss models are generally developed using two approaches: 1. Synthetic or Engineering functions, 2. Empirical modelling. Synthetic models use expert opinions or engineering solutions that result in a set of What-If scenarios to estimate flood losses. They are not based on statistical analysis of observed data (Penning-Rowsell and Chatterton, 1977). Though one of the major advantages of synthetic loss models is their non-dependency on empirical data, the development of detailed damage scenarios covering all damage possibilities and building characteristics requires high effort (Smith, 1994). Since these models are synthesized based on a variety of data sources, such as expert knowledge and technical papers, the models are more generalized and lead to higher levels of standardization compared to empirical models (Smith, 1994; Merz et al., 2010b; Amadio et al., 2019). For practical applications, the outputs from the synthetic models are required to capture the observed damage processes. However, except in very few models such as the INSYDE (Dottori et al., 2019), the empirical loss values do not constitute the model development.

Empirical models are developed based on real damage information observed from past events and hence, require large amounts of high-quality detailed data on flood damage and the damage-influencing factors, such as water depth (Merz et al., 2010b; Smith, 1994). These models aim to represent the relationship between flood damage and its influencing factors using patterns that occurred in the past events. The empirical models may be based on data from a single event (localized model) or cumulative data from multiple events (generalized model). Flood loss models purely based on localized empirical datasets are unable to reliably predict building damage for other events (Wagenaar et al., 2018). In contrast, generalized models (e.g. Bayesian Network, multi-level parameterization) based on data from multiple events cover a wider range of damage processes and perform better for new events (Wagenaar et al., 2018; Sairam et al., 2019). As empirical models are based on real damage data, it is expected that they capture the observed damage processes and are less prone to surprises (Merz et al., 2015). However, an important disadvantage is their requirement for detailed damage surveys. These are often expensive and time consuming. Survey campaigns that are conducted after extreme events may result in a large sample of respondents that reported damage. However, in the case of surveys conducted after small localized events, the resulting datasets are often insufficient to model different damage processes. Owing to lack of detailed object-level damage data, only a few studies have validated the flood loss models against observed loss estimates (Gerl et al., 2016; Amadio et al., 2019). In the case of empirical models, it is still possible to use a part of the empirical data for validation during model development. Since synthetic models are generally developed when empirical data is unavailable, both calibration and validation of synthetic models remain a challenge.

More than 95% of the state-of-the-art flood loss models (both synthetic and empirical) are deterministic (Gerl et al., 2016). Deterministic models result in one damage estimate based on the influencing factors. However, in reality, there exists variability in damage predictions given by the loss model based on the influencing factors. This may be due to the inherent stochastic nature of damage processes and other reasons such as uncertainty in empirical data, model structure and missing influencing factors in the model (Schröter et al., 2014; Winter et al., 2018). Decision makers and administrators are required to consider thoroughly the reliability of the flood loss models, in order to base FRM decisions and investments on the loss predictions. Hence, flood loss models should provide loss predictions along with an estimate of their uncertainty and reliability. A probabilistic flood loss model estimates the probability of occurrence of all possible loss scenarios for each object and results in a distribution of predicted losses. Probabilistic models potentially account for all sources of uncertainty in model parameters, structure and variability in the modelled processes based on observed data and assumptions concerning damage processes. Hence, there is an increasing interest in developing probabilistic approaches for flood loss modelling (Schröter et al., 2014; Wagenaar et al., 2018; Rözer et al., 2019; Lüdtke et al., 2019). In the presence of large detailed empirical datasets, advanced approaches for the development of probabilistic loss models are given by Wagenaar et al., (2018) and Rözer et al., (2019). These probabilistic approaches are applicable for empirical models. Since the synthetic models are not fitted to observed losses during development, they are commonly not calibrated. Hence, it is impossible to estimate the reliability of the synthetic model without validating the model against empirical loss data (Zischg et al., 2018).

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

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

4.2 Methods and Data

4.2.1 Setting up the framework for BDDS model:

The BDDS model describes the relationship between empirical losses and their corresponding deterministic loss predictions from synthetic models by means of a full Bayesian approach. The parameters of the BDDS model are indicators pertaining to the deviation between the synthetic model predictions and empirical observations. Also, the full joint posterior probability distribution of the BDDS model parameters can be obtained along with the predictive distribution of flood losses given the synthetic model and empirical losses from events that occurred in the region. From the credibility intervals of the predictive distributions, it is possible to estimate the uncertainty in the flood loss predictions.

The BDDS model is based on the premise that the empirical losses and synthetic loss predictions may be seen as components of a statistical model, in which the synthetic loss predictions are considered as exogenous variables (one that is determined outside the model, and imposed on the model) that are used to determine the observed losses. The BDDS model estimates losses using a linear function with empirical loss as the dependent variable regressed against the synthetic loss prediction. We assume that the BDDS model is identifiable for households within a region: i.e., the damage processes that occur in households belonging to one region are the same. Hence, the BDDS model assumes a single set of parameters for each region. In order to make the loss predictions comparable across the different case studies, we use relative loss to buildings, $rloss$, which is the ratio of absolute building loss to its total reconstruction value in the respective currencies, at the time of the event (Elmer et al., 2010). The rloss values are between 0 and 1, where 0 indicates no damage and 1 indicates complete damage, requiring reconstruction of the building. The BDDS model is given in Equation 4.1.

$$\widetilde{rloss}|rloss\_syn \sim Beta(\alpha, \beta) \qquad \textit{Equation – 4.1}$$
$$\alpha = \mu \times \varphi$$
$$\beta = (1-\mu) \times \varphi$$
$$\mu = inv\ logit\ (\lambda \times rloss\_syn + \varepsilon)$$

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

4.2.2 BDDS model construction

In reality, we are particularly interested in the capability of the BDDS model to estimate expected flood losses to buildings after an event (post-event scenarios) or predict expected losses for a potential future event. Hence, the BDDS model is constructed considering two scenarios:

1. **Post-event:** Comparison of a BDDS model developed using empirical data from one event against synthetic loss predictions, for the same event using 10-fold Cross Validation (local 10-fold CV). The empirical dataset from the event is split into 10 parts, a BDDS model is trained with 9 parts of the dataset and validated on the left-out data (10^{th} part). This is repeated 10 times, i.e., until all of the dataset is validated.

2. **Future event:** Comparison of a BDDS model developed using empirical data from one or more events against synthetic loss predictions, for a future event that occurs in the same region (Temporal one-step ahead Cross Validation; see Figure 4.1). A BDDS model (BDDS e_1) is developed using synthetic model and empirical flood loss data from the first event (e_1). This model provides calibrated probabilistic loss predictions for the future event, e_2. After the occurrence of the event e_2, a BDDS model (BDDS e_1, e_2) is developed using the same synthetic model and empirical loss data from both events e_1 and e_2. This model results in calibrated probabilistic loss predictions for the event e_3, which may potentially happen in the future.

$$p\left(\widetilde{rloss}_{b'e}\middle|\widetilde{rloss}_{be}\right) = \int_{\Theta} p(\widetilde{rloss}_{b'e}|\theta)p(\theta|\widetilde{rloss}_{be})d\theta \qquad \textit{Equation} - \textit{4.2}$$

$$p\left(\widetilde{rloss}_{b'e'}\middle|\widetilde{rloss}_{be}\right) = \int_{\Theta} p(\widetilde{rloss}_{b'e'}|\theta)p(\theta|\widetilde{rloss}_{be})d\theta \qquad \textit{Equation} - \textit{4.3}$$

The BDDS model definition for the two scenarios of CV are given in equations 4.2 and 4.3, respectively. We are particularly interested in the posterior predictive distribution of the target variable $\widetilde{rloss}$ of residential buildings b' that are not included in training the BDDS model conditioned on the observed losses from the empirical dataset, $\widetilde{rloss}_{be}$ from buildings b and events e. For the post-event damage prediction, the posterior prediction consists of residential buildings that are from the same event e as the empirical data used in the BDDS model training/calibration (Equation 4.2). For the future event damage prediction, the posterior prediction of $\widetilde{rloss}$ are estimated for residential buildings from a future event e' that was not used in the BDDS model training/calibration. θ contains the beta model parameters (φ, λ and ε) as shown in Equation 4.1. Hence, after specifying a prior for θ, one finds the posterior distribution $p(\theta|\widetilde{rloss_{be}})$.

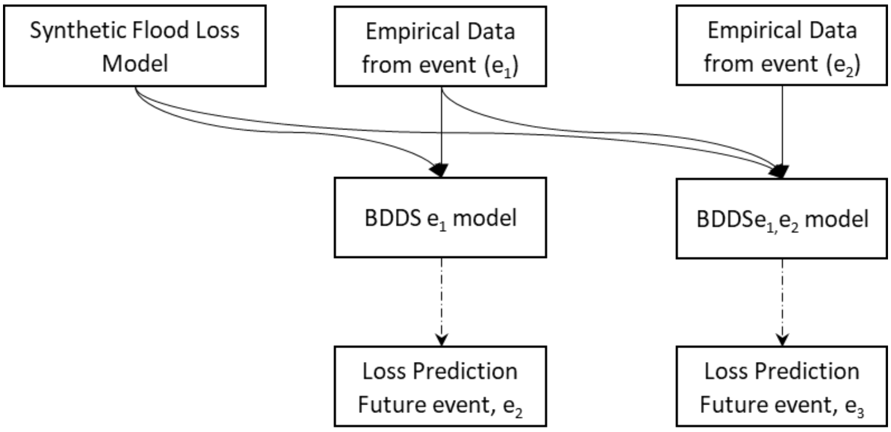


Figure 4.1: Framework for Temporal one-step ahead CV using a synthetic flood loss model and continuous collection of empirical flood loss data. The components involved in the development of BDDS model are shown with solid lines and the predictions are shown as dot-dash lines.

4.2.3 Metrics for assessing model performances

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

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|\widetilde{rloss}_i - rloss_i\right| \qquad \textit{Equation – 4.4}$$

$$MBE = \frac{1}{n}\sum_{i=1}^{n}\widetilde{rloss}_i - rloss_i \qquad \textit{Equation – 4.5}$$

$$HR = \frac{1}{n}\sum_{i=1}^{n}h_i;\ h_i = 1\ if\ rloss_i \in HDI_{90i};\ 0\ , otherwise \qquad \textit{Equation – 4.6}$$

$$IS = HDI_{90i} + \frac{1}{n}\sum_{i=1}^{n}\frac{2}{\beta}\left(\min(HDI_{90i}) - \widetilde{rloss}_i\right)\Big|\left\{\widetilde{rloss}_i < \min(HDI_{90i})\right\} + \frac{2}{\beta}\left(\widetilde{rloss}_i - \max(HDI_{90i})\right|\left\{\widetilde{rloss}_i > max(HDI_{90i})\right\} \qquad \textit{Equation – 4.7}$$

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

4.2.4 Case studies: Synthetic models, event description and empirical data

4.2.4.1 Cumbria, United Kingdom

Synthetic model: Multi Coloured Manual (MCM)

The Multi-Coloured Manual (MCM) (Penning-Rowsell et al., 2013) was initiated in 1977 and incrementally improved thereafter and was developed for the purpose of benefit appraisal for flood investment. It aims to represent national economic losses in sterling. Adopting a deterministic approach, the MCM provides a range of synthetically-generated absolute depth-damage functions for residential and non-residential properties of different types which have been developed to provide national consistent values. The damage functions are generated for individual inventory items and building contents per social grade based on the best ownership and economic values available from market-based surveys and synthetically generated susceptibility curves. For residential properties, unique damage functions are provided according to the type and duration of flooding, warning lead time, building type, year of construction and social class; and estimates of damage are provided for the building fabric and contents and the costs of drying and cleaning. Weighted average damage function curves are then obtained for the different properties considering the national distribution of properties in flood prone areas. For comparability, we utilize MCM loss data to only the residential building fabric and divide by reconstruction cost to obtain an estimate of relative loss. Since empirical data concerning social class was not available, an initial MCM assessment for building fabric losses was performed utilizing different damage functions based on type of flooding, water depth, duration of inundation, warning lead time, building type and year of building construction.

Event description and empirical data: Cumbria 2015

The December 2015 flood event in Cumbria (Storm Desmond) was characterized by exceptionally high rainfall, temperature and soil moisture. This is the biggest recorded flooding in Cumbria in almost all the river basins. In comparison, the meteorological winter of 2015/2016 was the wettest on record across all of the UK. The December 2015 event with a return period of 800 to 1,000 years in some parts of Cumbria broke numerous climate records resulting in extreme flooding and strong winds. This event is estimated to have caused impacts between £520 and £662 Million (Szönyi et al., 2016). In most parts of Cumbria, the flooding occurred due to overtopping of the structural protection measures such as dikes and flood walls. In Cockermouth and Keswick, the improved flood protection reduced the impacts of the 2015 event. Further information on the event can be found in Szönyi et al., (2016) and Cumbria County Council (2018). The households reported up to 3 meters of inundation depth and the duration of inundation was between a few hours to almost 48 hours in many regions. After the 2015 event, computer-aided telephone surveys were undertaken targeting the households that suffered damage during the 2015 flooding. A list of affected streets was obtained using the flood outlines published by the Environment Agency DEFRA (Environment Agency DEFRA, 2019) and the telephone numbers of households in these streets were obtained from public telephone directory. The survey locations were mainly spread over northern UK, mainly focused on the Cumbria region covering, Appleby, Keswick, Kendal, Carlisle and Cockermouth. The survey consisted of questions concerning the hazard (water depth, duration, velocity, contamination etc.), exposure (rebuilding cost and content value), vulnerability (building type, construction year, private precautionary measures, emergency measures, warning information etc.) and incurred damage to building structure and contents. The reconstruction costs for the houses were obtained from the Association of British Insurers (ABI, 2003). The households that provided water depth and building loss information from the Cumbria region were selected for this analysis. This resulted in a dataset with 33 residential buildings. All of these households provided information pertaining to the initial appraisal of the MCM. The summary statistics of the responses from the households are provided in Table 4.1.

4.2.4.2 Meuse, Netherlands

Synthetic model: SSM

SSM is a flood loss model developed for the Dutch national government (De Bruijn et al., 2014). It is the standard model applied in all Dutch FRM studies for the national government. It is an update of an earlier model called Standard Damage and Fatality assessment model (HIS-SSM) (Kok et al., 2005). The damage function applied in this paper, for residential structural damage was first proposed in Duiser (1982). This damage function is based on a combination of information synthesized from empirical observations concerning flood damage from three events: the coastal floods in Zeeland in 1953, the Wieringermeer flood of 1945 from a large lake and a flood in Tuindorp-Oostzaan in 1960 (canal dike breach), interviews from experts and damage functions from Penning-Rowsell et al., (1977).

Event description and empirical data: Meuse 1993

This dataset is based on the 1993 flood of the Meuse River in the Dutch province of Limburg. It has been described in WL Delft (1994), Wind et al., (1999) and Wagenaar et al., (2017). The 1993 Meuse discharge was 3,120 m3/s, the highest recorded up to that point. 8% of the province was flooded causing about 180 Million Euro damage (price level 2016) (Wagenaar et al., 2017). Unlike most of the rest of Dutch rivers, in 1993 the Meuse River didn’t have dikes yet in Limburg. The data was collected to compensate affected households. Every flooded building was visited, resulting in a complete data set of 5,780 records. The data collection was carried out by insurance experts who visited the affected buildings weeks after the flood, often before restoration activities were completed. The experts also recorded the water depth in the buildings but this wasn’t their primary objective and was sometimes difficult to assess because the flood had happened weeks prior. In Wagenaar et al., (2018) the recorded flood losses have been transferred to relative losses. The summary statistics of the survey responses are given in Table 4.1.

4.2.4.3 Adda, Caldogno and Secchia, Northern Italy

Synthetic model: INSYDE (Dottori et al, 2016)

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

Event descriptions and empirical data: Adda 2002, Caldogno 2010, Secchia 2014

In this case study three flood events in Northern Italy are considered. The first one happened in November 2002 in the town of Lodi. The flood resulted from a most critical combination of events for the lower part of the Adda river, namely the simultaneous increase of the discharges from the Como lake and of the Brembo river, that is the largest tributary of the Adda upstream of Lodi. Between the 25th and 26th of November, the Adda reached the hydrometric height of 3.43 m above the reference level (68.28 m a.s.l.), corresponding to a discharge between 1,800 and 2,000 m^3/s. The return period has been estimated as 100-200 years. Large portions of the town were flooded with water levels above 2 m in some neighbourhoods. The second flood event happened in the Veneto region, where from the 31st of October to the 2nd of November 2010, persistent rainfall affected the pre-Alpine and foothill areas, with peaks of more than 500 mm in some locations (ARPAV, 2010). Consequently, about 140 km^2 of land was inundated, involving 130 municipalities, some of which were particularly negatively affected. The situation of Bacchiglione River and its tributaries was especially critical, where hydrometric levels overcame historical records (water velocities in the river higher than 330m^3/s were registered; see Belcaro et al., 2011), causing the opening of a breach on the right levee of the river on the morning of the 1st of November. The countryside and the settlements of Caldogno, Cresole and Rettorgole were flooded with an average water depth of 0.5 m (ARPAV, 2010) for about 48 hours. The total damage, including residential properties, economic activities, agriculture and public infrastructures, was estimated to be about EUR 26 million, of which EUR 7.5 million relate to the residential sector (Scorzini and Frank, 2017). Finally, the last event happened in January 2014 in the central area of the Emilia–Romagna region (Modena province), where in the early morning of the 19th of January the water started to overtop the right levee of the Secchia River, flooding the countryside. The breach was not caused by an extreme river discharge (the return period of the event was estimated around 5 years), but by the collapse of the river embankment, weakened by animal burrows (D'Alpaos et al., 2014). Seven municipalities were affected with an inundated area of around 52 km^2 with the small towns of Bastiglia and Bomporto suffering the largest impacts remaining flooded for more than 48 h. The total volume of overflowing water was estimated about $36x10^6$ m^3, with an average water depth of 1 m (D'Alpaos et al., 2014). The economic cost inflicted on residential properties, according to damage declaration, amounted to EUR 36 million. After the three floods, public funding was made available by the national Civil Protection Authority. In order to be reimbursed, with similar procedures for all inundation events, citizens were requested to fill in pre-filled claim forms; the latter were then mostly collected by the affected municipalities and, in a small part, by the Regional Authorities. In total, our dataset includes 1,158 buildings in the flooded areas (Amadio et al., 2019). They include information on the owner, the address of the flooded building, its typology (e.g. apartment, single house), the number of affected floors, a description of the physical damage and its translation into monetary terms (distinguishing for the different rooms among damage to walls, windows and doors, floor and content). More information about the individual flood events, their hydrodynamic simulations and the data collection campaigns were published in Scorzini et al., (2018), Molinari et al., (submitted), Scorzini and Frank (2017), Carisi et al (2018), Amadio et al., (2019). The summary of empirical data from this case study is provided in Table 4.1.

4.2.4.4 Danube, Germany

Synthetic model: Rhine Atlas Model (RAM) (ICPR, 2001)

The Rhine Atlas Model (RAM) was developed in 2001 in order to determine the regions with high flood risk in the Rhine catchment based on the 1995 floods and develop FRM strategies (ICPR, 2001). Since, the RAM is intended for the Rhine catchment, an inherent transfer scenario exists when the RAM is generalized to the other catchments within Germany. However, given that a number of studies consider RAM as a standard synthetic flood loss model (Jongman et al., 2012, de Moel. 2011), we use the model as the standard synthetic flood loss model for Germany. The RAM is mostly based on expert judgment as well as some information based on the HOWAS empirical flood damage data (Buck & Merkel. 1999). It is a stage-damage function using water depth as the only predictor. The RAM loss prediction is based on the resolution of land-use classes similar to that of the CORINE land use data (Jongman et al., 2012). We apply the stage-damage function corresponding to losses to building structure in the residential land-use class to estimate flood loss for each residential building.

Event descriptions and empirical data: Danube 2002-2013

In this case study, three flood events that occurred between 2002 and 2013 in the Danube catchment is considered. Among the events, the 2013 flood was quite extreme with return period up to greater than 1000 years in some parts of the catchment. These were summer floods caused due to heavy rainfall resulting in surface water flooding and flash floods (Vogel et al., 2018). The 2013 floods were characterized by high antecedent soil moisture combined with heavy precipitation resulting in large spatial extent of flood peaks with high magnitudes resulting in the most severe flooding in Germany over the past 6 decades (Merz et al., 2014a; Schröter et al., 2015). Another distinguishing feature is the occurrence of dike breaches during the Danube 2013 event. Many properties were affected after dike breaches (e.g. at Deggendorf). After these events, computer-aided cross-sectional telephone surveys of private households that had suffered from losses were undertaken using a standardized questionnaire. A list of affected streets was obtained using the flood masks derived from satellite data, (DLR, https://www.zki.dlr.de/), and the telephone numbers of households in these streets were obtained from public telephone directory. The survey campaigns always focused on a single event. Depth of water within the house is determined using the reported water level in the highest affected storey by applying corrections based on the presence of a basement and height of the ground floor. Building reconstruction costs are adjusted for inflation to values as of 2013 using the building price index (DESTATIS, 2013). We consider all datasets which refer to households with basement (for unbiased measurements of water depth) and for which information on water depth and relative building loss were provided. Hence, the empirical data used in this study consists of 408 buildings from three events in the Danube catchment, that have a considerable number of completed surveys (sample size>25). The summary of empirical data from this case study is provided in Table 4.1.

4.2.4.5 Elbe, Germany

Synthetic model: Rhine Atlas Model (RAM) (ICPR, 2001)

The Rhine Atlas Model (RAM), described in section 4.2.4.4 is implemented for estimating losses in the Elbe catchment.

Event descriptions and empirical data: Elbe 2002-2013

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

Empirical damage data was collected from the affected households in the Elbe catchment during the same survey campaigns, explained in the previous section. The study uses four events comprising of a total of 1,110 households, that provided information on water depth and relative building loss and have a considerable number of completed surveys (sample size>25). The summary of empirical data from this case study is provided in Table 4.1. More information about the individual flood events in the Elbe and Danube, the surveys and their results were published in Thieken et al., (2007), Kreibich et al., (2011, 2017(b)), Kienzler et al., (2015) and Vogel et al., (2018).

Table 4.1: Sample size, the summary (average) of water depth (wd) in meters, exposed building value (bv) in EUR, absolute and relative losses to residential buildings (bloss in EUR, rloss) for the five case studies

Case study	Event	Sample size	wd	bv	bloss	rloss
Cumbria, United Kingdom (UK)	Cumbria 2015	33	0.6	287,000	24,000	0.07
Meuse, Netherlands (NL)	Meuse 1993	5780	0.4	197,356	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	354,785	6,258	0.015
	Danube 2005	104	2.0	406,012	7,874	0.015
	Danube 2013	79	3.0	571,536	45,000	0.060

Elbe, Germany (DE)	Elbe 2002	518	3.5	302,005	43,805	0.096
	Elbe 2006	42	2.9	307,800	6,962	0.018
	Elbe 2011	58	2.7	475,456	9,140	0.015
	Elbe 2013	492	2.7	427,680	23,250	0.051
Total		8489				

4.3 Results and discussion

Comparison of predictions from synthetic loss models and BDDS model

The performance of the BDDS model is compared with the synthetic models from the respective regions. Since the development of BDDS models requires empirical data, new BDDS models are trained for each of the local 10-fold CV as well as temporal one-step-ahead CV. During both validation scenarios, there are no variations in definition and parameterization of the synthetic models. Point estimates are assessed via MAE and MBE and prediction uncertainty and reliability via IS and HR (section 4.2.3). Reliability and uncertainty of loss predictions are provided by all BDDS models, representing an enhancement over the deterministic synthetic models (4 out of 5 models). The model validation is performed by bootstrap sampling of the synthetic and BDDS model predictions with 1,000 iterations with replacement, while preserving the sample size of the empirical data during each iteration.

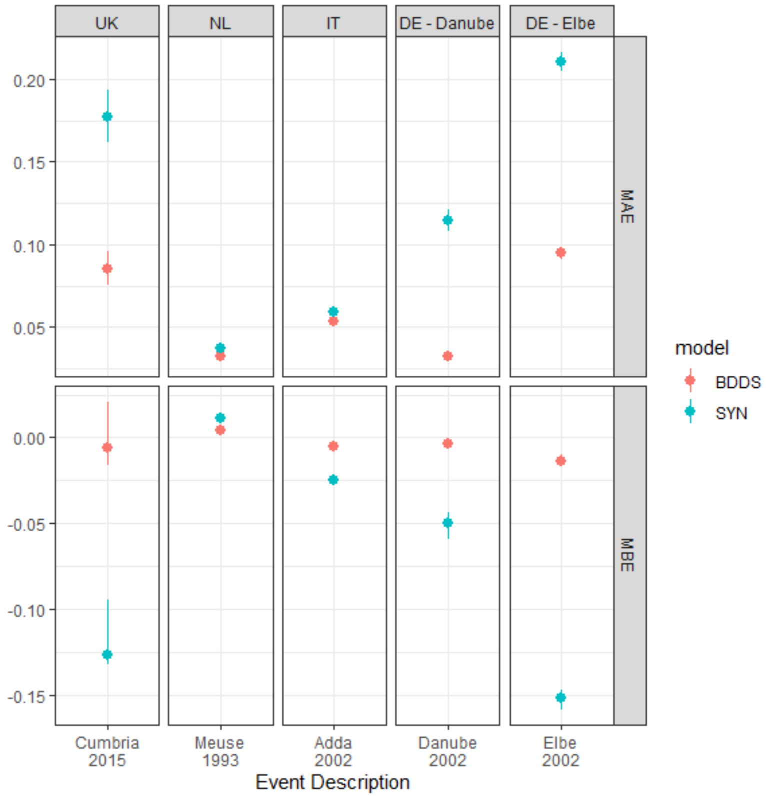
4.3.1 Local 10-fold CV

We perform a local 10-fold CV in order to validate the BDDS model predictions against the synthetic model predictions for the post-event scenario. The case studies with no empirical data from the region prior to the event are used for local 10-fold CV. This scenario (Equation 4.6) is applicable for the Cumbria 2015, Meuse 1993, Adda 2002, Danube 2002 and Elbe 2002 flood events. These events are either the only available empirical data from the respective regions or the first event of the continuous empirical data collection campaigns. All synthetic models, except SSM, result in a negative MBE which indicates that on average, all these synthetic models over-estimate the building losses (see Figure 4.2a).

The prediction performance of the BDDS model with one event is compared against the performance of the synthetic models from the corresponding countries (Figure 4.2a). The BDDS model performs better than the synthetic model in terms of point estimates. As described in Equation 4.6, during the local 10-fold CV, the model is iteratively validated on residential buildings that are not used in the model development. Thus, the local 10-fold CV evaluates out-of-sample model performance of the BDDS model. The BDDS model with RAM and empirical data from the Elbe 2002 event results in the highest improvement in predictive performance in terms of MAE and MBE. Small improvement in predictive performance is exhibited by the BDDS models - SSM and empirical data from Meuse 1993 event and INSYDE with empirical data from the Adda 2002 event. However, among the tested synthetic models, the INSYDE and SSM models result in the smallest errors in the 10-fold CV. Among the two catchments in Germany, the RAM results in larger errors predicting losses for the Elbe 2002 event compared to the Danube 2002 event. The BDDS model consistently improves the predictions for the 2002 event in both catchments.

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

Among the tested synthetic models, the SSM and INSYDE models result in the least errors (see, Figure 4.2a). 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 4.1). This may lead to smaller errors in terms of MAE and MBE (Wagenaar et al., 2018).



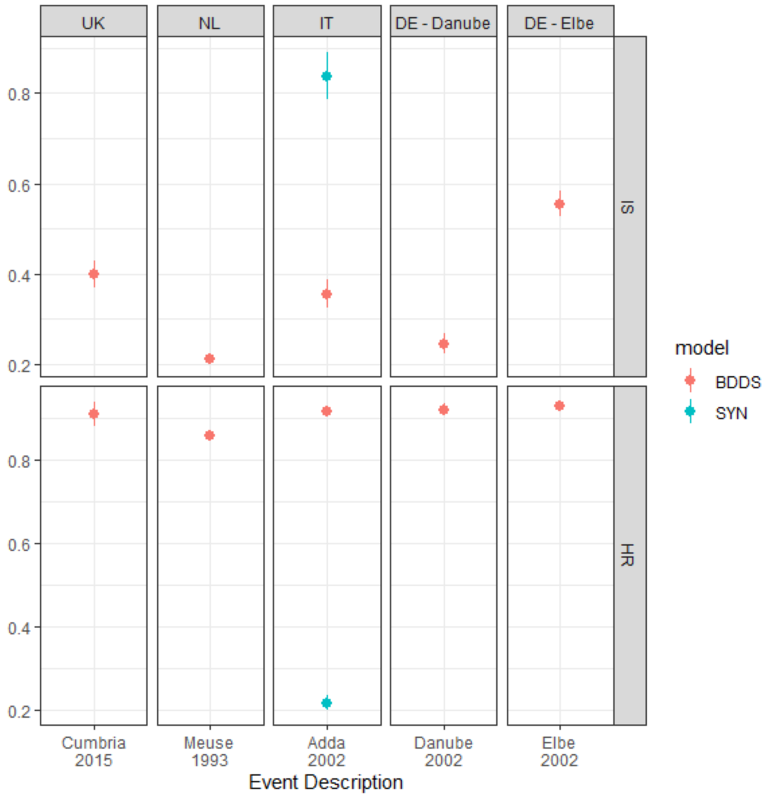


Figure 4.2: Model performances for local 10-fold CV using events and their corresponding synthetic loss models (shown in brackets) — Cumbria 2015 (MCM), Meuse 1993 (SSM), Adda 2002 (INSYDE),

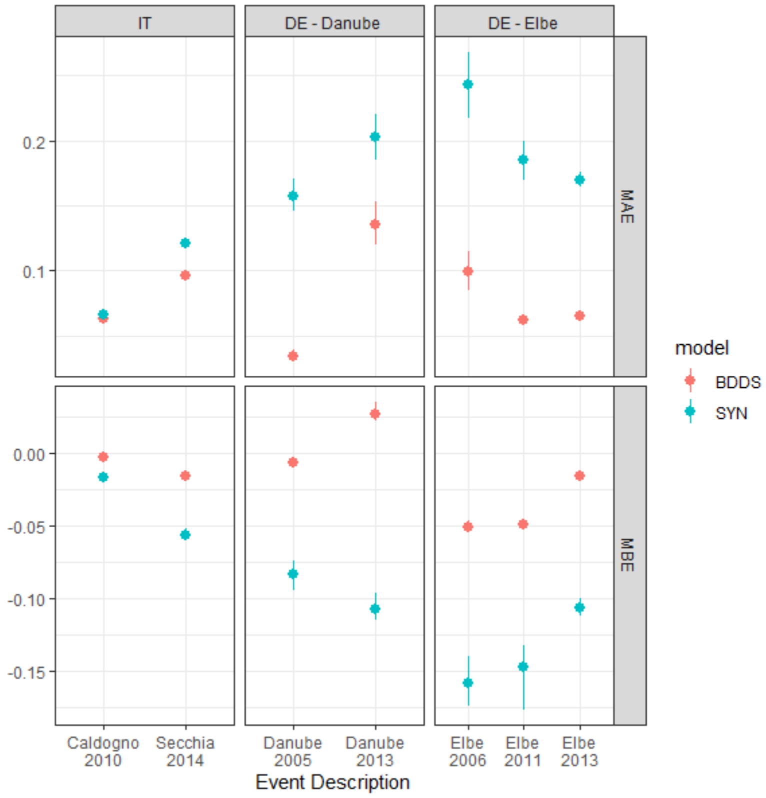
Danube 2002 (RAM) and Elbe 2002 (RAM). (a) MAE and MBE of flood loss predictions using synthetic models and BDDS models (b) IS and HR of loss predictions using BDDS models.

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 2002, the hit rate of the BDDS model is high and comparable with the performance of other BDDS models. However, the high IS indicates that the loss distributions are not sharp. This high uncertainty may be attributed to variability in damage processes that are not adequately captured by the variables in the RAM (i.e. water depth only). This quantification of uncertainty and reliability from BDDS model is an enhancement over the established synthetic models, which is crucial for risk-based decision making (Polasky et al., 2011).

4.3.2 Temporal One-step ahead CV

In regions where, continuous empirical flood damage data is available, the predictions from synthetic models and BDDS models are compared using temporal one-step ahead CV. The losses suffered by residential buildings due to an event in the future is predicted from a BDDS model developed using the synthetic model and all available empirical data from the past events (Figure 4.1 and Equation 4.7). From our case studies, empirical damage data from northern Italy and Germany can be used to implement temporal one-step ahead CV.

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



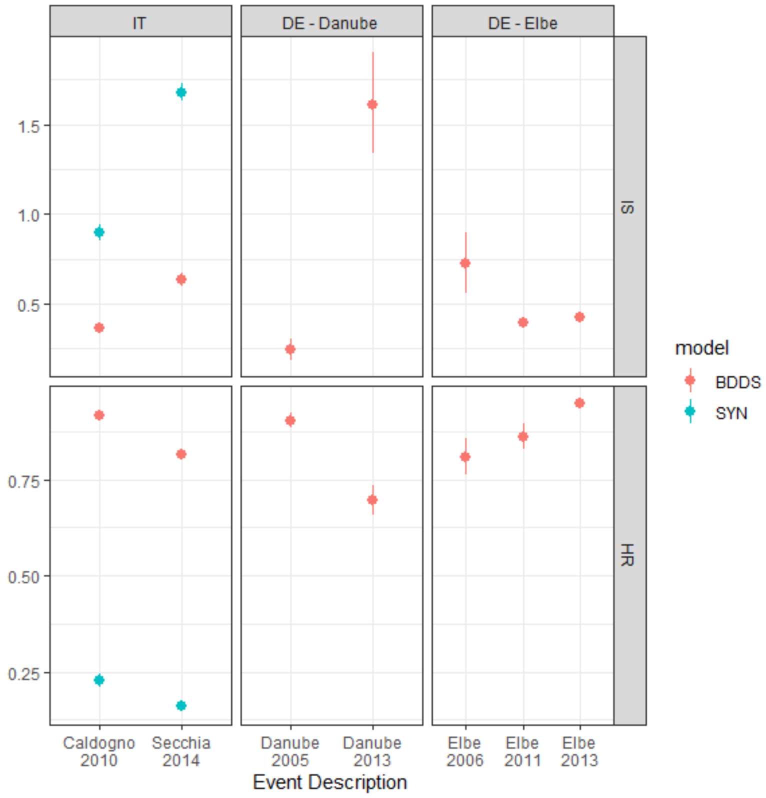


Figure 4.3: Model performances for temporal one-step ahead CV of events using empirical data from past events and their corresponding synthetic loss models (shown in brackets) -– Caldogno 2010 (Adda 2002; INSYDE), Secchia 2014 (Adda 2002, Caldogno 2010; INSYDE), Danube 2005 (Danube 2002;

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.

The results of the temporal one-step ahead CV are provided in Figure 4.3a. For all the case studies, the errors (MAE and MBE) from the BDDS model temporal one-step ahead prediction are smaller than the errors from the corresponding synthetic models. The results show that compared to the INSYDE model, the performance of the INSYDE model continuously integrated with empirical data from more events is higher. For the Elbe catchment, the BDDS model's improvement in predictive performance is observed for all future event predictions when integrated with a continuous collection of empirical data. These results suggest that, in these two regions, parameterizing the BDDS model with empirical data from events in the recent past improves the damage prediction for following events

In the Danube catchment in Germany, the BDDS model outperforms the RAM for temporal one-step ahead predictions. However, the BDDS model shows a lower performance when data from an additional event is integrated. We also notice a change from negative to positive bias. This suggests that in the case of Danube 2013 event, the BDDS model developed by integrating RAM with empirical data from 2002 and 2005 events under-estimates the losses. The uncertainty and reliability estimates, i.e. IS and HR, from BDDS model one-step ahead temporal predictions are shown in Figure 4.3b. The two BDDS models developed for the case study in Northern Italy result in better HR and IS estimates compared with the INSYDE model. The BDDS model shows best reliability and least uncertainty for the Elbe 2013 event with a HR close to 100% and a relatively small IS, suggesting small uncertainty. On the other hand, loss predictions for the 2013 event in the Danube catchment from the BDDS model performs the worst with the least HR of 70% and a high IS, suggesting low reliability and large uncertainty.

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 4.3b). This effect is also in agreement with the inferences from MBE for Danube 2013 estimated from the same model (Figure 4.3a). 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 flood experience and were not prepared for flooding. Hence, high intensity flooding combined with low preparedness resulted in large damage (e.g. oil contamination from heating systems). Such effects are not sufficiently captured either by the uni-variable RAM or the empirical data from past events. A potential approach to capture the difference in damage processes between events is to introduce a multi-level model that allows both shared and separate parameters representing the similarities and differences between the damage processes exhibited by the different events (Sairam et al., 2019).

In order to interpret the importance of local empirical data, we discuss the performances of the BDDS model that is built with empirical data from the same event (local 10-fold CV) and past events (temporal one-step ahead CV). Local empirical data from the same event improves the overall reliability of the BDDS model and also results in low uncertainty, i.e. reduces IS and increases HR (Figures 4.2b and 4.3b). Hence, the use of empirical data from the same event is useful for post-event risk analysis and damage estimation. For risk-based decision making for future scenarios, we need accurate and reliable models, which can only be validated using empirical data from past events. Therefore, the IS and HR estimates obtained from the temporal one-step ahead loss predictions are more relevant. These metrics can be considered by decision makers and flood risk managers as the estimates of uncertainty and reliability of the damage model for future flood risk portfolios. In general, the BDDS model enhances synthetic models using local empirical data.

4.4 Conclusions

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

Our validation results show that empirical loss data from past events are valuable for enhancing the synthetic models to predict damage more accurately. Hence, for improving estimates of future risk, empirical data collection campaigns after flood events are crucial. However, the loss predictions from the post-event scenario show higher reliability compared to the future risk predictions. This suggests that flood damage processes show variability across events and dynamic damage models are required to capture this variability. An important feature of the presented approach is the uncertainty quantification of the damage estimate, since this provides valuable information for improved decision making. Thus, the Bayesian Data-Driven approach is valuable for flood risk managers.

Acknowledgements: This research has received funding from the European Union's Horizon 2020 research and innovation program under Grant Agreement 676027 MSCA ETN System-Risk.

Data and Software Availability: The Multi Coloured Manual (MCM) database handbook is published by the Flood Hazard Research Centre (FHRC) at MiddleSex Univerisity, London, UK. The functions are proprietary and not publicly accessible. The SSM model is available from deBruijn et al., (2014). INSYDE functions are available for download as R open source code, currently hosted on GitHub (https://github.com/ruipcfig/insyde/). The Rhine Atlas Model (RAM) is available from ICPR (2001). The data implemented in the Cumbria 2015 case study is currently not publicly accessible. The dataset may be obtained upon request. The data used in the Meuse 1993 case study is available from Wagenaar et al., 2017. The dataset used in the Northern Italy case study are not publicly accessible. The first reason behind this is that some data come from private sources (i.e., businesses, utilities companies) that agreed on sharing their data only for research objectives, including sensitive information. The dataset may be obtained upon request. For the Danube and Elbe case studies, 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. The models presented in this paper are implemented in the stan modeling language (Carpenter et al., 2017) using the brms package version 3.3.2 (Bürkner, 2017) in R (R Core Team, 2019).

5 Hierarchical Bayesian Approach for Modelling Spatiotemporal Variability in Flood Damage Processes

Summary

Flood damage processes are complex and vary between events and regions. State-of-the-art flood loss models are often developed on 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 (multi-level) 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 is available, we use variables pertaining to specific region- and event-characteristics representing commonly available expert knowledge as group-level predictors within the HBM.

Published as: 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 8233-8237. https://doi.org/10.1029/2019WR025068

5.1 Introduction

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

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 spatio-temporal 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 et al., 2017b).

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, sealing oil tanks, etc (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 spatio-temporal variability in the damage processes.

HBMs can be theoretically conceptualized and implemented to account for causal effects in processes (Gelman, 2006; Kruschke and Vanpaemel, 2015; Levy et al., 2012; Feller & Gelman et al., 2015). 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 spatio-temporal trends in precipitation extremes; Ahn et al., (2016) developed a HBM to forecast seasonal stream flows. Das et al., (2018) showed the potential of using a hierarchical modelling approach for modelling irrigation withdrawals over the US, especially for data-sparse years. However, as per our knowledge, there are no studies which implemented a HBM for flood loss estimation.

The localized model considers that each region and event has distinct damage processes which 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 (multi-level) 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 spatio-temporal transferability of flood loss models, the damage processes pertaining to different events and regions are modelled separately, while also accounting for similar processes across regions and events. Bayesian probabilistic modelling 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, i.e. 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 spatio-temporal 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 spatio-temporal 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 5.2.1. The functional form and different model structures of HBM, localized and generalized models are discussed in section 5.2.2. Methods and metrics to assess model performance are discussed in section 5.2.3. The best performing HBM structure is chosen in section 3.1. The HBM, localized and generalized model parameters are explained in section 5.3.2. The development of a HBM with group-level predictors is described in Section 5.3.3. The predictive performance of the models and inferences are explained in Section 5.3.4.

5.2 Data and methods

5.2.1 Data

Object (household)-level empirical flood loss data is 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 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 storey 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, 2017(b)), Kienzler et al., (2015) and Vogel et al., (2018).

Table 5.1: 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 sq.m 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), multi-family houses (bt2) and semi-detached houses (bt3) for each of the spatio-temporal groups.

Catchment	Event	Sample size	wd	bv*	bloss*	rloss	pre	fe	d	bt			ba	rp
										bt1	bt2	bt3		
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	>1000
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		1663												

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

For our study we selected from these surveys, all datasets which 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 spatio-temporal 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 datasets (>25) are considered in this study, resulting in total 1663 datasets. Information regarding each of the spatio-temporal groups is reported in Table 5.1. From table 5.1, 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 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 meters, 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.

5.2.2 Modelling flood damage processes

5.2.2.1 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 ($\sqrt{wd}$), since this functional form has been proven to be suitable (Merz et al., 2013; Schröter et al., 2014; Wagenaar et al., 2017; Rözer et al., 2019). Values of $rloss$ lie between 0 and 1. In contrast to deterministic models that assumes 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 un-bounded distributions may result in implausible values. Therefore, rloss is modelled 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 5.1). μ is the mean $rloss$ which is a function of $\sqrt{wd}$ 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 5.1, weakly informative generic priors are provided. For example, the water depth is constrained to be positively correlated with rloss.

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

Equation - 5.1

5.2.2.2 Modelling Approaches

<u>HBM</u>

A HBM is a multi-level 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 5.1, the damage processes can be modelled to vary either randomly between region- and event-groups (varying intercept model) or conditioned on $\sqrt{wd}$ (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 is 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 are 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 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 over-arching 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 5.2: model structure - M5). Similarly, when the slope and intercept are made to vary between region- and event-groups, there are over-arching distributions at two levels, governing their variability across the region- and event-groups and also across regions (see Table 2: model structure – M8).

Table 5.2: Specification of the eight HBM structures (M1 – M8) tested in this study. In the Model Structure Specification, $\sqrt{wd_i}$ is the square root of water depth at i^th household, θ and ε are the coefficient of water depth and intercept, respectively. The shared parameters that are common to all region- and –event groups are represented without subscripts, i.e. 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 Structures	Description	Model Structure Specification
M1	Varying intercept between spatial groups (regions)	$\log(\mu_i) = \theta \times \sqrt{wd_i} + \varepsilon_r$ $\varepsilon_r \sim normal(\mu'_r, \sigma'_r)$
M2	Varying intercept between spatiotemporal groups (regions-events)	$\log(\mu_i) = \theta \times \sqrt{wd_i} + \varepsilon_{re}$ $\varepsilon_{re} \sim normal(\mu'_{re}, \sigma'_{re})$ $\mu'_{re} \sim normal(\mu'_r, \sigma'_r)$
M3	Varying slope between spatial groups (regions)	$\log(\mu_i) = \theta_r \times \sqrt{wd_i} + \varepsilon$ $\theta_r \sim normal(\mu_r, \sigma_r)$
M4	Varying slope between spatiotemporal groups (regions-events)	$\log(\mu_i) = \theta_{re} \times \sqrt{wd_i} + \varepsilon$ $\theta_{re} \sim normal(\mu_{re}, \sigma_{re})$ $\mu_{re} \sim normal(\mu_r, \sigma_r)$
M5	Varying slope and intercept between spatial groups (regions)	$\log(\mu_i) = \theta_r \times \sqrt{wd_i} + \varepsilon_r$ $\theta_r \sim normal(\mu_r, \sigma_r)$ $\varepsilon_r \sim normal(\mu'_r, \sigma'_r)$
M6	Varying slope between spatio-temporal groups (regions-events) and varying intercept between spatial groups (regions)	$\log(\mu_i) = \theta_{re} \times \sqrt{wd_i} + \varepsilon_r$ $\theta_{re} \sim normal(\mu_{re}, \sigma_{re})$ $\mu_{re} \sim normal(\mu_r, \sigma_r)$ $\varepsilon_r \sim normal(\mu'_r, \sigma'_r)$

M7	Varying slope between spatial groups (regions) and varying intercept between spatio-temporal groups (regions-events)	$\log(\mu_i) = \theta_r \times \sqrt{wd}_i + \varepsilon_{re}$ $\theta_r \sim normal(\mu_r, \sigma_r)$ $\varepsilon_{re} \sim normal(\mu'_{re}, \sigma'_{re})$ $\mu'_{re} \sim normal(\mu'_r, \sigma'_r)$
M8	Varying slope and intercept between spatio-temporal groups (regions-events)	$\log(\mu_i) = \theta_{re} \times \sqrt{wd}_i + \varepsilon_{re}$ $\theta_{re} \sim normal(\mu_{re}, \sigma_{re})$ $\mu_{re} \sim normal(\mu_r, \sigma_r)$ $\varepsilon_{re} \sim normal(\mu'_{re}, \sigma'_{re})$ $\mu'_{re} \sim normal(\mu'_r, \sigma'_r)$

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 are always conditioned on the region in which the households are located (see Table 5.2: model structures M2, M4, M6, M8). Among the tested model structures, the one with the best prediction capability is chosen and compared against the generalized and localized models which 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). $\theta \sim normal(0,1)$, $\varepsilon \sim cauchy(0,10)$ and the coefficient of $\sqrt{wd}$, 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. $\mu_r \sim normal(0,1)$, $\sigma_r \sim cauchy(0,10)$, $\mu'_r \sim normal(0,1)$, $\sigma'_r \sim cauchy(0,10)$. The parameters for standard deviation, σ_r and σ'_r are constrained to be non-negative.

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 (Merz et al., 2013, Kreibich et al., 2017a). 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 5.1) generalizes the damage processes conditioned on $\sqrt{wd}$ and results in a single slope estimate (θ), intercept (ε) and precision (φ), as shown in Figure 5.1. Weakly informative priors are provided for θ and ε; θ and φ are constrained to be non-negative.

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 5.1 results in slope (θ_{re}), intercept (ε_{re}) and the precision parameter (φ_{re}), as shown in Figure 5.1. θ_{re} and φ_{re} are constrained to be non-negative. 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 modelling approach may result in unreliable, noisy estimates.

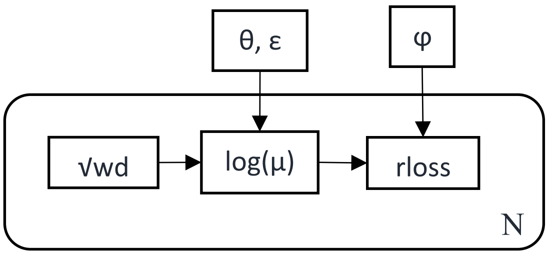
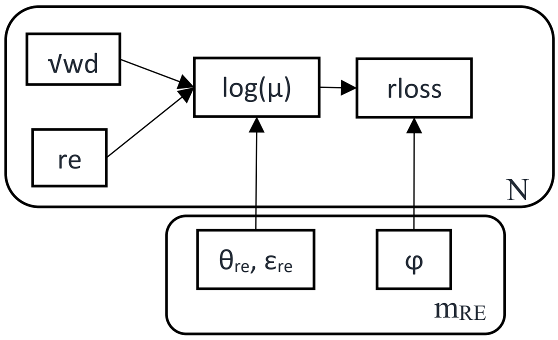
Model	**Graphical structure**	**Model specification**
Generalized Model		$\log(\mu_i) = \theta \times \sqrt{wd_i} + \varepsilon$ $\theta \sim normal(0,1);\ \theta \geq 0$ $\varepsilon \sim normal(0,1)$ $\alpha_i = \mu_i \times \varphi$ $\beta_i = (1 - \mu_i) \times \varphi$
Localized model		$\log(\mu_i) = \theta_{re} \times \sqrt{wd_i} + \varepsilon_{re}$ $\theta_{re} \geq 0$ $\alpha_i = \mu_i \times \varphi_{re}$ $\beta_i = (1 - \mu_i) \times \varphi_{re}$

Figure 5.1: 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 (1663); 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.

5.2.2.3 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, i.e. out-of-sample and out-of-group validations, are performed using three performance metrics – Expected Log-pointwise Predictive Density ($elpd$), Mean Absolute Error (MAE), Mean Bias Estimate (MBE). $elpd$ (Equation 5.2) is a measure of the predictive accuracy of the model for data points considered (Vehtari et al., 2016).

$$elpd = \sum_{i=1}^{n} \int p_t(\tilde{y}_i) \log p(\tilde{y}_i|y)\, d\tilde{y}_i \qquad \textit{Equation – 5.2}$$

Where $p_t(\tilde{y}_i)$ is the true density of observed rloss for i^{th} household; $p(\tilde{y}_i|y)$ 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 to 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).

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|\widetilde{rloss}_i - rloss_i\right| \qquad \textit{Equation – 5.3}$$

$$MBE = \frac{1}{n}\sum_{i=1}^{n}\widetilde{rloss}_i - rloss_i \qquad \textit{Equation – 5.4}$$

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 5.3 and 5.4, n refers to the total number of households in the validation dataset; $rloss_i$ 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.

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 $\hat{k}$ 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 (S.E) whose corresponding p-value is < 0.0001. MAE for out-of-sample validation are determined using a 10-fold cross-validation performed by iteratively removing ten 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.

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 (LOGO-CV). 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 are 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, 200 from each region- and event-group) with replacement is performed while estimating $elpd$ for out-of-group validation.

5.3 Results and Discussion

5.3.1 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 S.E) 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. Amongst 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 and 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 spatio-temporal groups. Therefore, we choose 'M4 - Varying slope between spatio-temporal groups', as the appropriate model structure. Its graphical structure and model specifications are shown in Figure 5.2. Thus, this HBM structure is proposed for flood loss estimation and in the following tested against the generalized and localized models.

Table 5.3: Out-of-sample predictive performances of potential HBM structures. LOO-CV with PSIS approximation is used to estimate and compare the out-of-sample expected log-pointwise. The Standard Errors - S.E of the comparisons are shown in brackets

Model Comparison	**Out-of-sample LOO-CV with PSIS approximation**	
	Elpd difference (S.E)	**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

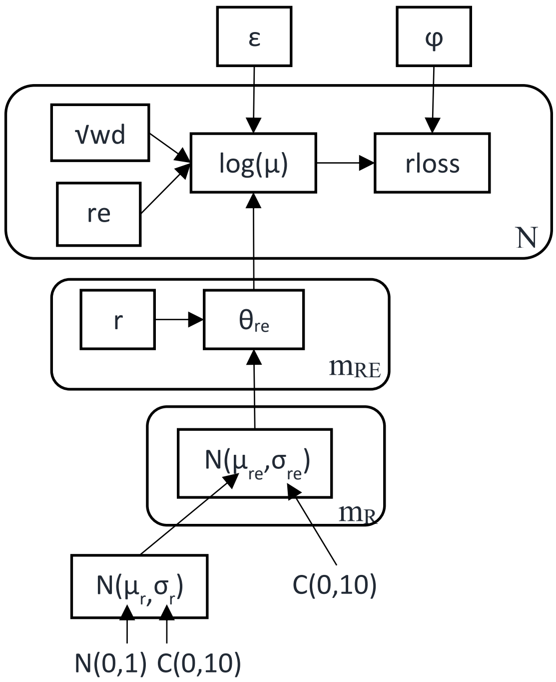
Model	Graphical structure	Model specification
HBM - M4		$\log(\mu_i) = \theta_{re} \times \sqrt{wd_i} + \varepsilon$ $\theta_{re} \sim normal(\mu_{re}, \sigma_{re})$ $\mu_{re} \sim normal(\mu_r, \sigma_r)$ $\sigma_{re} \sim cauchy(0,10)$ $\mu_r \sim normal(0,1)$ $\sigma_r \sim cauchy(0,10)$ $\alpha_i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

Figure 5.2: 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 (1663); m_{RE} refers to the number of region- and event-groups (9); m_R refers to the number of region-groups (4), see Table 5.1. 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

According to the chosen structure, HBM - M4 (Figure 5.2), $rloss$ is modelled 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 $rloss$,
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 $\sqrt{wd}$. 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 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 5.1: median water depth for all the events were 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 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 which 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 socio-economic 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.

5.3.2 Model parameters

The HBM - M4 has a single shared intercept (ε = -3.75), but separate slopes for each region- and event-group with overarching distributions as shown in Figure 5.3a. 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 over-arching distributions are provided in the SI, section 5.5.1. 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 from each other (Figure 5.3a). 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 ($\theta_{Elbe2002}$ = 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 ($\theta_{Elbe2013}$ = 2.79) indicating the improved resistance of households to flood damage compared to the 2002 event. The means of the distributions governing the variability of the slopes (θ_{re}) within each region, μ_{re} (Figure 5.3a) do not show much variation between the Elbe ($\mu_{Elbe(02,06,11,13)}$ = 2.23) and Danube ($\mu_{Danube(02,05,13)}$ = 2.26) regions. This suggests that, the variability in damage processes across different events within the same region are 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 5.3b. From the distributions of slope and intercept for every region- and event-group provided in SI, section 5.5.1, 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 dataset. Consistent inferences regarding damage processes cannot be made from these noisy parameter estimates. The generalized model results in a single slope parameter (coefficient of $\sqrt{wd}$, θ = 2.78) and intercept (ε = -3.77) for all region- and event-groups (Figure 5.3c). The damage processes represented by the parameters of the generalized model are more inclined towards extreme events (such as Elbe 2002, 2013 and Danube 2013) and may not accurately capture the damage processes of small events (such as Elbe 2006, Rhine 2011). The distributions of the slope and intercept from the generalized model are provided in SI, section 5.5.1 (see, Figures 5.5 and 5.8)

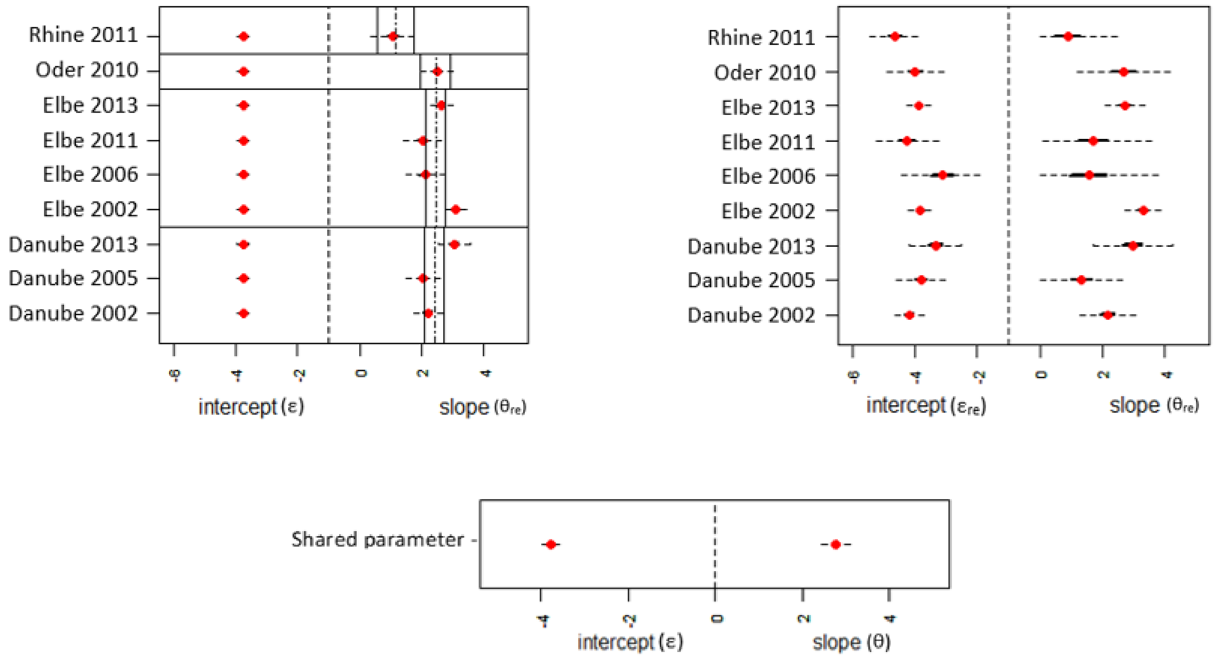


Figure 5.3: Intercept (ε) and slope (θ_{re}) parameters estimated from the HBM - M4 with μ_{re} for each region is represented as dot-dash vertical lines with solid vertical lines showing the 95% confidence interval; Intercept (ε_{re}) and slope (θ_{re}) parameters estimated from the localized model structure; Intercept (ε) and slope (θ) parameters estimated from the generalized model structure (clockwise).

5.3.3 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 is unavailable, we include group-level predictors which 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.

Identifying group-level predictors that improves flood loss predictions during spatio-temporal 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 step-wise linear regression (Venables & Ripley, 2002) with 1000 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 AIC (Akaike Information Criterion) and BIC (Bayesian Information Criterion). Both AIC and BIC score the model based on goodness-of-fit and also penalizes the model for over-fitting 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 5.1, prevalence of flood experience (fe_{re}), duration of inundation (d_{re}) and return period of the event (rp_{re}) are crucial in explaining the spatio-temporal variability in damage processes (Table 5.4). 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.

Table 5.4: Results of step-wise regression predicting varying slopes of HBM - 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

We hypothesize that, region- and event-group-level predictors such as the percentage of households that have prior flood experience, and 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 5.4. 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 is included in SI section: 5.5.1 (see, Figure 5.9).

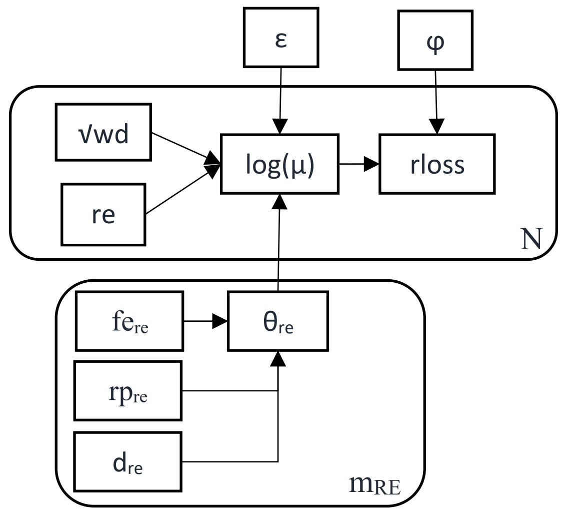
<table>
<tr><th>Model</th><th>Graphical structure</th><th>Model specification</th></tr>
<tr><td>HBM - M4 with group-level predictors</td><td>
</td><td>$$\log(\mu_i) = \theta_{re} \times \sqrt{wd_i} + \varepsilon$$
$$\theta_{re} \sim normal(fe_{re} * A + d_{re} * B + rp_{re} * C + D, \sigma_{re})$$
$$\alpha_i = \mu_i \times \varphi$$
$$\beta_i = (1 - \mu_i) \times \varphi$$
$$rloss_i \sim beta(\alpha_i, \beta_i)$$</td></tr>
</table>

Figure 5.4: 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 (1663); m_{RE} refers to the number of region- and event-groups (9), see table 5.1. In the model specification, subscript i refers to i^{th} 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.

5.3.4 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 Table 5.5a and 5.5b, respectively. The elpd comparison for the models is provided aggregated for all the region- and event-groups in Table 5.5c. 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, 5.5 and 5.8) of Elbe 2002 are very close to the slope and intercept of the generalized model. Since, a large sample of households in the dataset 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 5.1). 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 5.5 and 5.8) 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 modelling of the regional variability in damage processes. Though the HBM - M4 results in an overall best fit (refer to section 5.3.1), generalizing the damage processes (varying slope and constant intercept – M4) between Rhine 2011 and other events leads to over-estimation 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 5.5b) 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 S.E) in their out-of-sample prediction capabilities (LOO-CV). For all LOO-CV model comparisons (Table 5b), the PSIS $\hat{k}$ values were less than 0.7 indicating a reliable estimation of $elpd$ (Vehtari et al., 2017).

The ability of the models to perform in spatio-temporal 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 (Leave-one-event-out cross-validation) 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 dataset, they are not used in Leave-one-event-out cross-validation. All the nine region- and event-groups are used in predicting held-out households from the regions (Leave-one-region-out cross-validation). The HBM - M4 performs best during spatio-temporal transfer compared to the generalized and localized models in terms of point estimate errors MAE and MBE (Table 5.5a and 5.5b). This result agrees with the conclusions from previous studies (Cammerer et al., 2013; Schröter et al., 2014; Wagenaar et al., 2016; Jongman et al., 2012; Vogel et al., 2018) 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 5.5c). Hence, the predictive performance of the HBM - M4 is significantly higher than that of the generalized and localized models during spatio-temporal transfer. The performance of the HBM - M4 with group-level predictors is assessed for held-out-households using Leave-one-event-out cross-validation. The HBM - M4 with group-level predictors performs better than HBM - M4 in terms of point estimate errors and elpd estimates (Tables 5.5a, 5.5b and 5.5c). 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 spatio-temporal transfer.

Table 5.5: 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 2006	Danube 2013	Elbe 2002	Elbe 2005	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 2006	Danube 2013	Elbe 2002	Elbe 2005	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 (S.E)	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)*	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

* Insignificant difference (elpd difference from LOO-CV with PSIS approximation < 4 S.E)

5.4 Conclusions

A HBM is developed for capturing spatio-temporal variability in flood damage processes. Parameterization of the widely used depth-damage functions, i.e. 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 is 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 which 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 modelling, particularly in spatio-temporal 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.

5.5 Supporting Information (SI)

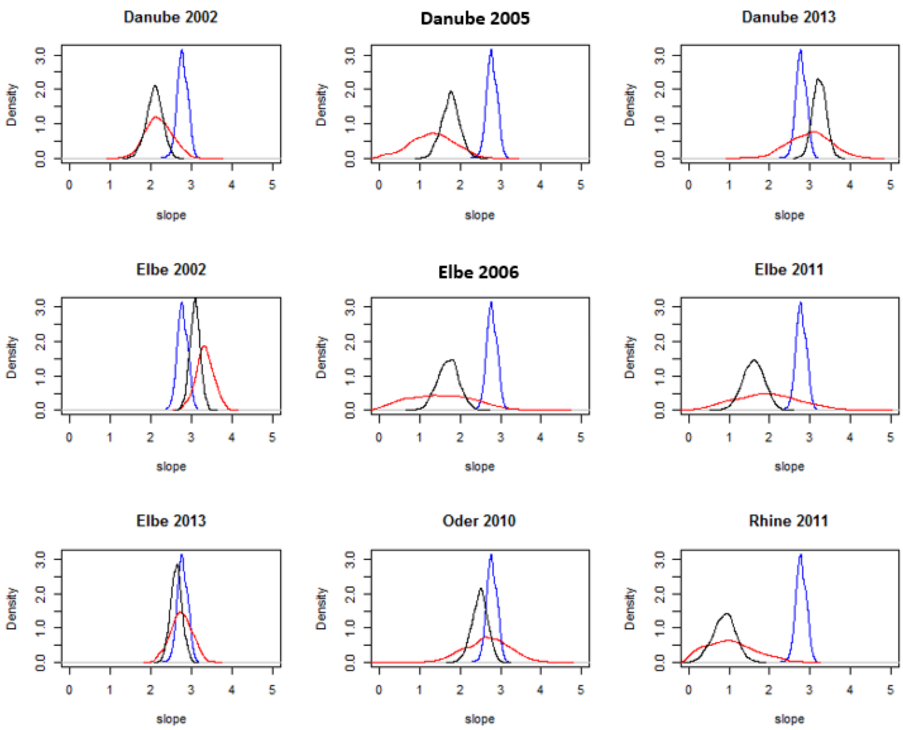


Figure 5.5: Slope ($\boldsymbol{\theta}$); The posterior of slope parameter ($\boldsymbol{\theta}$) from the generalizd model is shown in blue; the slopes ($\boldsymbol{\theta}_{re}$) for each region- and event-groups (re) from the ungeneralized and HBM – M4 models are shown in red and black respectively

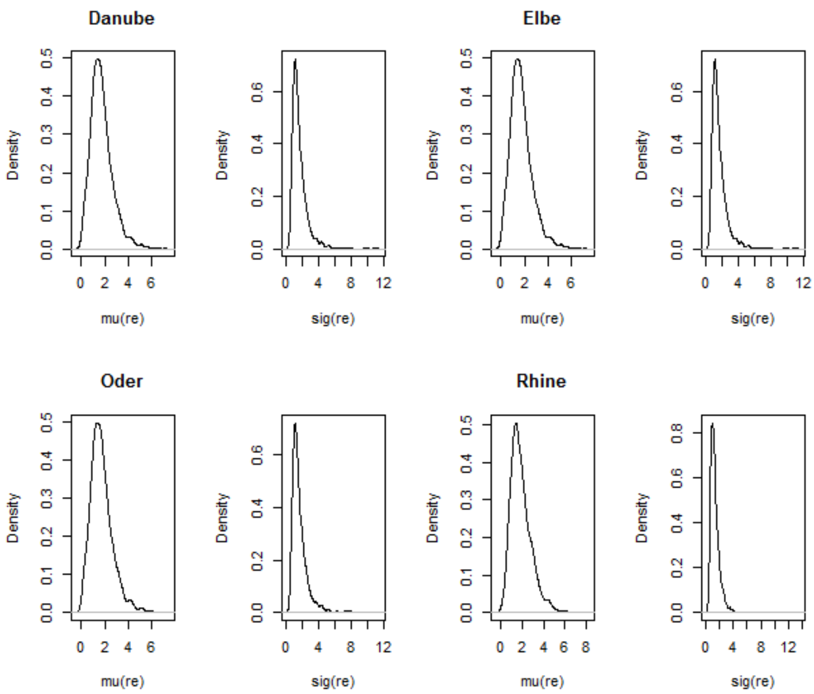


Figure 5.6: Region-level parameters (μ_{re}, σ_{re}) governing variability of slope in HBM – M4

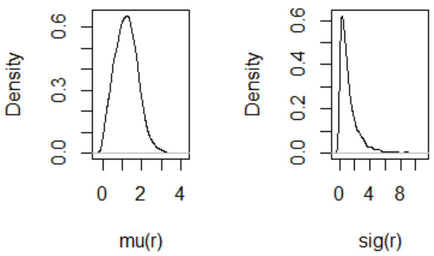


Figure 5.7: Parameters (μ_r, σ_r) governing variability of region-level parameters (μ_{re}, σ_{re}) in HBM - M4

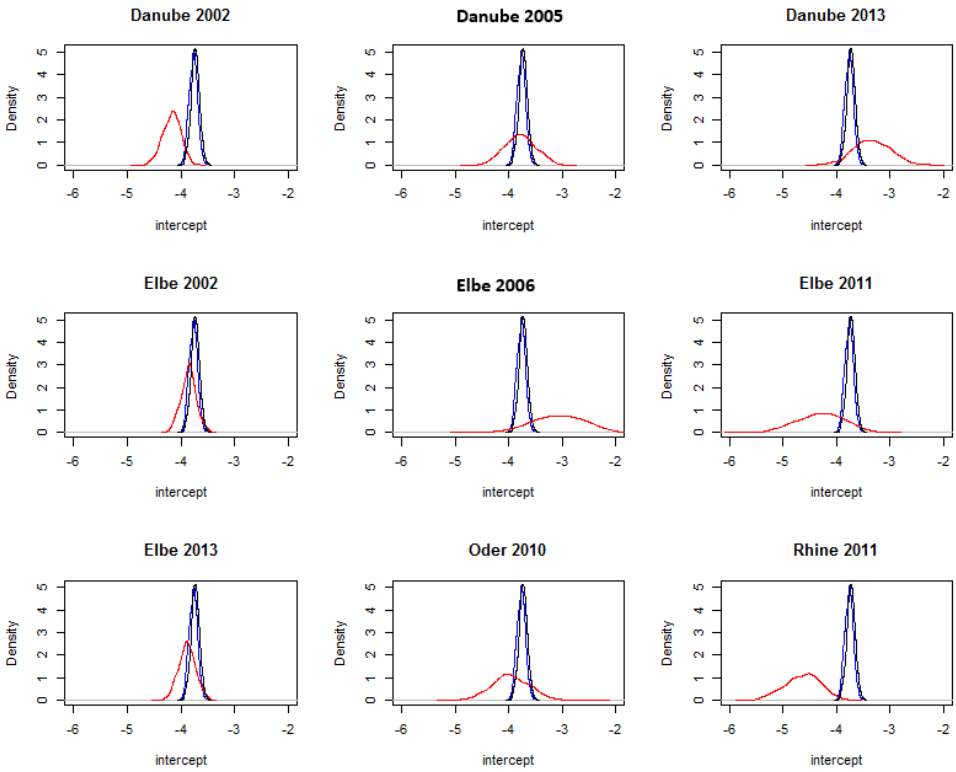


Figure 5.8: Intercept (ε) from the generalized and HBM - M4 models are shown in blue and black respectively. The intercepts (ε_{re}) corresponding to each region- and event-groups from the localized model are shown in red.

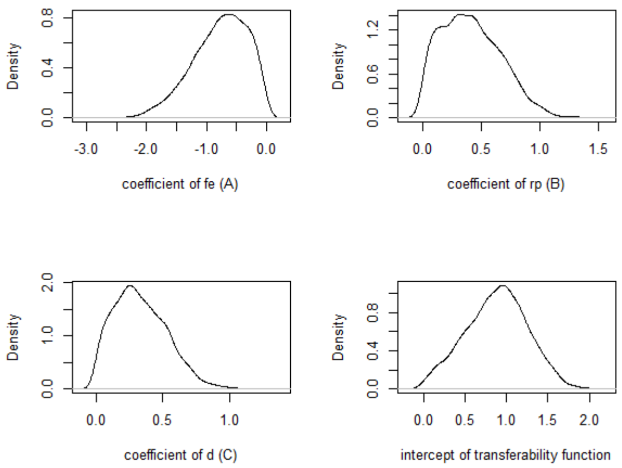


Figure 5.9: Parameters from HBM – M4 with group-level predictors (flood experience - fe, return period - rp and duration of inundation - d)

Acknowledgements: This research has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 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), no. 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.

6 Discussion, Conclusions and Outlook

6.1 Discussion

The overarching objective of this thesis was to develop Bayesian modelling approaches that improve the representation of flood damage processes. Specifically, changes in vulnerability via private precaution was quantified and appropriate flood loss models capturing these changes were identified; challenges in the representation of flood damage processes with respect to reliability of loss predictions and spatiotemporal transferability of flood loss models were addressed. The research questions from section 1.4 are addressed in this section. Inferences regarding the damage processes, obtained during the model development along with the potential of these models regarding implementation, applicability and limitations are also discussed. Following the discussions, key points from the thesis are highlighted.

1. *What is the role of private precaution in reducing the vulnerability of households?*

In contemporary FRM, the implementation of private precaution is given a lot of importance. However, quantitative knowledge concerning the effectiveness of private precaution is rarely available. Therefore, the loss reduction due to implementation of private precaution is not generally considered in risk assessment for planning of large-scale adaptation measures or determining insurance premiums. Recent studies have focused on the implementation of Agent Based Models (ABM) and use of Protection Motivation Theory (PMT) to model potential adaptation scenarios for Natural hazards (Haer et al., 2016; Bubeck et al., 2018; Yang et al., 2018). These approaches quantify the propensity of implementation of private precaution based on several identified drivers and possible scenarios. In order to assess dynamics of flood risk, the propensity of private precaution should be considered in a flood loss model along with its feedbacks and effectiveness. Therefore, an important step in moving towards modelling changes in flood vulnerability for risk assessment is the robust estimation of effectiveness of private precaution in reducing flood losses.

In chapter 2, prevalent private precautionary measures in Germany including waterproof sealing, flood adapted use and flood adapted interior fitting were considered. The Average Treatment Effect (ATE) of private precaution was estimated to be between 11 and 15 thousand Euros per household. This is approximately equal to 27% of average building losses suffered by all the flooded households in Germany during the six events between 2002 and 2013. The

ATE along with the propensity of households to implement private precaution is valuable for risk-based decision making. Also, the ATE estimates are useful for performing cost-benefit analysis along with the expected hazard scenarios. Quantifying the economic benefits of adaptation may motivate the uptake of appropriate private precautionary measures and enhance risk-based adaptation.

The empirical damage data used in this analysis included only households that experienced damage due to flooding. Hence, an important shortcoming is the unavailability of households which had no damage due to flooding. Though unlikely, considering the high inundation depths from the riverine floods, there is still a possibility that these households avoided 100% damage due to implementation of private precaution. Hence, the analysis might have resulted in an under-estimation of the average effect of private precaution. In addition, more subjective or qualitative aspects of damage processes are not considered in this analysis. For example, one of the important factors that negatively influences implementation of private precaution is ignorance towards residual risk (Barendrecht et al., 2017). Households just behind a newly improved dike ring or flood wall tend to feel very safe and indifferent towards implementing private precaution. Hence, in addition to flood experience and coping abilities, risk communication and risk perception before the implementation of private precaution are important confounders. Empirical data from a cross-sectional survey do not show a clear temporal precedence from responses concerning risk communication and risk perception to the occurrence of the flood event. Though this limitation is prone to introduce a bias in the estimated effectiveness of private precaution, the estimate is evaluated to have low sensitivity to missing confounders based on Rosenbaum's sensitivity bounds. The sensitivity test uses Hodges-Lehmann point estimate, which is an established indicator of robustness of ATE.

Additionally, it is important to acknowledge the contextual nature of the effectiveness of private precaution estimated in this study. Though the estimate is localized to the contexts of flood risk in Germany, the methodology and analysis techniques discussed in chapter 2 is applicable for estimating effectiveness of flood adaptation measures for other contexts. Hence, a potential direction of future research is to quantify the reduction in flood vulnerability due to private precaution in a spatial transferability scenario.

2. *Which of the state-of-the-art flood loss models account for the influence of this vulnerability reduction?*

Flood damage processes are represented using flood loss models. Flood loss models quantify risk in monetary values that are useful for making adaptation decisions, urban planning and developing risk profiles for insurance (Merz et al., 2010b). To model dynamics of flood risk, it is important that flood loss models capture changes in vulnerability. The implementation of private precautionary measures is an important driver of flood vulnerability.

In a few state-of-the-art flood loss models, the effectiveness of private precaution is considered to absolutely prevent damage up to certain water depth. For example, In Multi-Coloured Manual (MCM), the implementation of private precaution avoids all damage for water depths less than 0.6 m (Penning-Rowsell and Chatterton, 1977). This scenario is valid only when flood barriers are implemented. On the other hand, the effect of other private precautionary measures such as adapted building use and structural protection are influenced by different hazard attributes such as velocity and duration of inundation; building characteristics such as presence of basement, heating arrangements etc.

Hence, flood loss models were tested for their ability to capture the effect of commonly implemented private precautionary measures such as water proof sealing, flood adapted use and flood adapted interior fitting, while predicting losses. Among the tested models, the Bayesian Network based graphical model BN–FLEMOps (Wagenaar et al., 2018) and the rule-based FLEMOps model (Thieken et al., 2005; Elmer et al., 2010) combine expert knowledge with data-driven approaches. Regression Trees (Merz et al., 2013), Bagging Decision Trees (Kreibich et al., 2017a) and expert knowledge based FLEMOps and Bayesian Network models (Wagenaar et al., 2018) were pure data-driven models.

When the flood loss models were forced with an intervention concerning implementation of private precaution, the effect of private precaution on incurred loss was reflected in the loss predictions from the BN-FLEMOps and FLEMOps models. These models are capable of capturing counterfactual scenarios for risk assessments. On the other hand, the data-driven models were unable to capture the differences due to implementation of precautionary measures as their characteristics such as explanatory variables, relationships and parameterization were only based on their prediction capability. For example, the tree-based model shown in figure 2.1a considers the influence of private precaution only when water depth is lesser than -0.8 m and greater than -1.45 m, i.e. basement flooding only. For buildings experiencing these water depths, except for structural measures like sealing the basement, the role of other adaptive measures is irrelevant. Thus, this model shows low skill in capturing effect of private precaution. One plausible reason for low importance given to private precaution among the data-driven models is due to the inclusion of flood experience, which is a strong confounder of private precaution.

Since the data-driven models were aimed at maximizing the predictive performances rather than representing the damage processes, the influence of flood experience superseded the influence of private precaution. However, in reality, flood experience influences the implementation of private precaution, which reduces flood loss. The structure of the BN-FLEMOps and FLEMOps models align with this process. Hence, they are successful in capturing changes in vulnerability via private precaution. Thus, inclusion of expert knowledge representing influencing factors and causal inferences regarding damage processes are crucial to construct flood loss models that capture changes in vulnerability. From the validation results (see, Table 2.8), we see that the data-driven models perform the best in terms of prediction accuracy. Therefore, further research is required to account for the dynamic aspects of risk such as changes in vulnerability without compromising on prediction accuracy.

3. *What factors affect the uncertainty and reliability of flood loss models?*

Due to the stochastic nature of damage processes, there is high uncertainty in the representation of damage processes. In addition, uncertainty in empirical data and model structure and parameters contribute to large errors in the model predictions. Probabilistic flood loss models provide a distribution of loss predictions from which uncertainty and reliability of the predictions can be quantified. Uncertainty is measured by the width of the highest-density interval of the predictive distribution. Reliability is measured as ability of the predictive distribution to cover the actual observed loss. Based on the validation results of the probabilistic models developed in chapters 3 and 4, the important factors affecting the uncertainty and reliability of flood loss models are determined.

Chapter 3 introduced a fully probabilistic multi-variable empirical model for flood loss estimation. This resulted in consistent quantification of uncertainty via distribution of loss predictions. The choice of response distribution in a probabilistic flood loss model was found to be crucial for determining prediction uncertainties. For example, log-normal distribution of uncertainties (Merz et al., 2004) overestimate the uncertainties in the upper tail of the distribution, leading to a wide 90% HDI. Tree-based models (Merz et al, 2013; Kreibich et al., 2017a) use non-parametric distributions approximating relative loss predictions. Though these models have led to a reduction in prediction errors, the short tails of non-parametric distributions severely reduce the reliability of loss estimates (Sieg. 2019). The Bayesian Network model (BN-FLEMOps see figure 2.2) considers relative loss (degree of damage) to be a discrete quantity that follows a multinomial distribution. However, in the case of flood damage processes, the response variable, relative loss is continuous and bounded between 0 and 1. Hence, the beta distribution (Egorova et al., 2008) was identified as an appropriate response distribution for relative losses.

The beta distribution is bounded between 0 and 1, thus eliminating implausible values for loss prediction. The possibility to capture zero-loss cases using a mixed response distribution (zero-Inflated beta consisting of binomial distribution for chance of loss and beta distribution for degree of loss) is consistent with the definition of risk comprising of chance of loss and degree of loss. The presented Bayesian Beta framework is flexible and quantifies reliability of loss predictions independently of the model type. This framework can be implemented to estimate the reliability of existing flood loss models irrespective of their approach or complexity. The validation results show that this choice of response distribution significantly reduced the uncertainty and increased the reliability of the model predictions.

The Bayesian Data-Driven Synthetic (BDDS) model from chapter 4 uses the Bayesian Beta framework introduced in chapter 3 to quantify the reliability of established synthetic flood loss models using available empirical data. Based on the BDDS model performances, the uncertainty and reliability quantification were found to be dependent on the quality of empirical data used for calibration. In the presence of informative empirical data with good coverage, the BDDS model enhanced the synthetic model predictions and quantified model uncertainty.

Hence, uncertainty and reliability of models are influenced by the choice of the response distribution, influencing factors, model parameterization and representativeness of the empirical data used for model development/calibration with respect to the target region and event.

4. Can reliability of synthetic loss models be quantified?

Synthetic models are based on what-if analysis. Though established synthetic models are standardized, they are generally deterministic and not validated against observations. These models often have high uncertainties that remain obscure.

In order to calibrate the established synthetic models, a Bayesian Data-Driven approach was implemented using empirical loss data in chapter 4. The resulting Bayesian Data-Driven Synthetic (BDDS) model improved the prediction accuracy of the synthetic model and also provided a distribution of loss predictions. From the validation results based on the model performances for twelve events in Western Europe, the BDDS model reduced the point estimate errors and quantified the uncertainty and reliability of synthetic model predictions for post-event and future event scenarios.

The applicability of the BDDS model is tested for capturing damage processes in the four countries in Western Europe. Though the changes in vulnerability across different events in these regions are significant, there is no temporal changes in predominant building types and other exposed assets in these regions. Since, urbanization and exposure changes are very important drivers of flood damage in the developing countries, one direction for future research could be to test the approach on case studies from low and middle-income countries.

The BDDS model's capability to predict damage for future events is dependent on the quality and representative nature of the empirical data with respect to the future event. Obtaining high quality empirical damage data is challenging as it is time-consuming and expensive. In some regions where there is no record of past flooding or the last flood event was long ago, it is impossible to collect object-level empirical data. Also, if there are major changes in building construction types or implementation of adaptation, the empirical data from the past is no longer representative of the future scenario. Hence, one potential direction for future research is to develop approaches that replace the need for local empirical data with expert knowledge concerning changes in building stock, socio-economic characteristics and potential adaptation levels. In this respect, a suitable model structure and parameterization to use expert knowledge concerning region and event characteristics is presented in Chapter 5.

5. *How to improve spatiotemporal transfer of flood loss models?*

Applicability of flood loss models depend on their ability to represent damage processes across regions and events. However, flood loss models are often localized to the contexts in which they are trained and as such do not generalize well. Hence, flood loss models trained from specific events do not perform well when transferred in space and/or time. Approaches such as updating the model with local data have improved spatiotemporal transfer of flood loss models. For example, FLEMOps model trained with heterogenous data from several events (generalized model) performed well during spatiotemporal transfer. Also, updating the model with local empirical data improved the model performance (Wagenaar et al., 2018). Nevertheless, this was applicable only when detailed object level empirical data was available (Lüdtke et al., 2019).

In chapter 3, of this thesis, the Bayesian Beta model trained with data from Germany was applied to a recent pluvial flood event in Houston, TX (USA). Despite the regional differences between the training and application case studies, the validation results show that the model performs well for both the zip code and county levels. Though it is not possible to infer the transfer capability of the model using one case study, it supports the assumption that a successful model transfer to other regions is possible. One reason for the transfer capability of the Bayesian Beta model is that it is a generalized model, developed using heterogenous training data set containing different flood events with different event magnitudes.

The Bayesian Data-Driven Synthetic (BDDS) models developed for future event scenario in Chapter 4 also show that generalized models calibrated using heterogenous empirical data from several flood events perform well during transfer scenarios (see, figure 4.3). The improvement in performance is determined by the representative nature of the empirical data with respect to the future event. Hence, the generalized models do not perform well when there are changes in the drivers of flood damage and damage processes between the past and future events.

In order to overcome the limitations of the generalized models in spatiotemporal transfer of flood damage processes, chapter 5 of this thesis proposed a Hierarchical Bayesian Model (HBM). The HBM is a multi-level regression model with beta distribution as the response distribution for relative loss. It consists of group-level parameters and shared parameters for representing differences and similarities in food damage processes for different regions and events. The group-level parameters may be estimated either using empirical data (Figure 5.2) from several events and regions or using expert knowledge concerning their damage processes (Figure 5.4). The advantage of the HBM is that the Bayesian framework allows for a consistent quantification of uncertainties and the multi-level parameterization allows for transferability, making the presented model highly scalable. The HBM is validated using six events from four catchments within Germany. While further validation and case studies are necessary, the high prediction capabilities of the HBM during spatiotemporal transfer could be used in future studies to complement the lack of local empirical data availability.

Based on the inferences from chapters 3, 4 and 5, modelling flood damage processes using large samples of empirical data, preferably from various events and using multi-level parameterization based on the availability of empirical data and expert knowledge were found to improve the spatiotemporal transfer capabilities of the flood loss models.

Key Points

1. Private precaution significantly reduces flood vulnerability of households. In Germany, private precaution reduces damage to each household building by 11 - 15 thousand Euros.
2. Flood loss models developed using both expert knowledge and data-driven approaches (e.g. FLEMOps and BN-FLEMOps) are capable of capturing changes in flood vulnerability.
3. Probabilistic models are recommended for flood loss estimation as damage processes are stochastic and significant uncertainty exists in data and model parameters.
4. The reliability of loss prediction was found to be strongly influenced by the choice of response distribution for flood loss (Beta distribution improves model reliability compared to Gaussian or non-parametric distributions).
5. Combining synthetic loss models with local empirical data via the BDDS model enhances prediction accuracy and quantifies reliability for post-event and future event scenarios.
6. In the absence of local empirical data, the Hierarchical Bayesian Model (HBM) uses empirical data from other regions and events resulting in improved loss predictions than established stage-damage functions.
7. Using region and event-specific characteristics (e.g. flood experience, flood duration and event return period) improves the HBM performance during spatiotemporal transfer.

6.2 Conclusions

This thesis has presented Bayesian approaches that improved the representation of flood damage processes. The conclusions from the thesis are organized into three topics:

quantification of changes in flood vulnerability, reliability of flood loss models and spatiotemporal transferability of flood loss models.

6.2.1 Quantification of changes in flood vulnerability

The implementation of private precautionary measures is proven to reduce flood vulnerability of households. Using a large empirical dataset, the average treatment effect of private precautionary measures is estimated to be between 11 and 15 thousand Euros on average for each household. This is a valuable estimate for performing cost-benefit analysis which is crucial for decision making concerning implementation of private precaution. Among the state-of-the-art flood loss models, the FLEMOps and BN-FLEMOps are capable of capturing the changes in vulnerability due to implementation of private precaution. Among the two, the BN-FLEMOps shows higher loss prediction capabilities, making it an appropriate model for flood loss predictions which also captured the dynamics of flood vulnerability.

6.2.2 Reliability of flood loss models

Probabilistic models were developed based on empirical data using the Beta distribution as the response distribution. These models reduced the uncertainty in the model predictions and improved the model reliability. A significant share of flood loss models are synthetic models, developed using what-if scenarios based on expert knowledge. These models show high uncertainty which are often obscure. The developed Bayesian Data-Driven Synthetic (BDDS) models enhanced the established synthetic models using empirical data from several events. These models calibrated the synthetic models, improved their prediction accuracy as well as quantified the model reliability. Combining the vast compendium of knowledge from synthetic models and available empirical data within a Bayesian framework led to enhanced model predictions that are sharp (low uncertainty), calibrated and reliable.

6.2.3 Spatiotemporal transferability of flood loss models

Based on the availability of local empirical data and expert knowledge in the target region, the default approach for flood loss modelling may be chosen from the four Bayesian models presented in this thesis – BN-FLEMOps, Bayesian Beta, BDDS, HBM models. In the presence of detailed local empirical data, the BN-FLEMOps model is suitable for predicting flood loss. In the absence of detailed empirical data (multi-variable), the uni-variable Bayesian Beta model using only water depth as the damage predictor is suitable for predicting flood losses. In the presence of established synthetic models and some local empirical data, the BDDS model results in calibrated loss predictions with quantification of model uncertainty and reliability.

Owing to the challenges in collecting quality empirical data, the Hierarchical Bayesian Model (HBM) was developed. The HBM replaced the requirement of local empirical data with multi-level parameters and expert knowledge in the form of aggregated variables and was capable of capturing spatiotemporal variability in flood damage processes better than established stage-damage functions.

In conclusion, this thesis recommends the implementation of Bayesian modelling approaches for flood loss modelling, preferably, using heterogenous empirical data from several events. Though the choice of modelling approach, parameters and underlying assumptions has to be determined on case by case basis, probabilistic models such as BN-FLEMOps, Bayesian Beta, BDDS and HBM are recommended depending on the availability of empirical data and expert knowledge concerning flood damage processes.

6.3 Outlook

The Bayesian approaches presented in this thesis improved the representation of flood damage processes. In particular, the developed models are a step-forward in improving flood loss model applicability in the presence of little or no local empirical data. The scope for further research in continuation to this thesis may focus on improving the applicability and prediction capability of flood loss models.

Within the scope of this thesis, the flood loss models were validated using damage data from several events that occurred in Germany and some events from Italy, Netherlands, UK and the USA. Hence, one direction for future research that contributes to improving model applicability is to combine the developed models with flood inundation maps and exposure estimates from the respective regions to provide approaches for flood loss predictions along with consistent estimates of uncertainty for various countries.

Owing to the flexibility of these models with respect to empirical data requirements, another direction for future research is to integrate the developed approaches into a large-scale (European) probabilistic flood loss model that uses the vast compendium of available expert knowledge in the form of synthetic models and aggregated variables along with available empirical data.

Though the developed HBM and BDDS model perform better than established models during spatiotemporal transfer, it is important to acknowledge that the validation is still localized to the tested case studies. Hence, the models are not tested for their abilities to capture damage processes that occur in other countries during flood events characterized by different drivers of hazard, exposure and vulnerability. Hence, a future direction towards improving applicability is to develop flood loss models based on these approaches for case studies from other countries, especially, for developing countries with growing urbanization and high socio-economic variability among people.

Improvements in prediction capability of the flood loss models are commonly achieved by using more explanatory variables and complex parameterization, that may lead to overfitting. On the other hand, users of the flood loss models are more interested in parsimonious models with least input requirements. Hence, all models in this thesis were focused on limiting the input requirements and penalizing overfitting. For example, in chapter 5, HBM with expert knowledge performed similar to the model with local empirical data. Nevertheless, it cannot be ignored that in this study, the expert knowledge was derived from comprehensive household surveys which are not commonly available from other countries. Hence, future research towards model improvement could focus on extending the hierarchical Bayesian approach to integrate relevant hazard, exposure and vulnerability information mined from secondary data sources such as crowd-sourced maps (e.g. OSM), remote sensing imageries, social media information.

Bibliography

ABI (2003). Assessment of the cost and effect on future claims of installing flood damage resistant measures. *Association of British Insurers*, London.

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

Aldrich, D. P. (2012). Social, not physical, infrastructure: the critical role of civil society after the 1923 Tokyo earthquake. *Disasters*, *36*(3), 398-419.

Allaire, M. C. (2016). Disaster loss and social media: Can online information increase flood resilience? *Water Resources Research*, *52*(9), 7408-7423.

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

Apel, H., Aronica, G.T., Kreibich, H., & Thieken, A.H. (2009). Flood risk analysis-how detailed do we need to be? *Natural Hazards*, *49*(1)*,* 79-98.

ARPAV: Scheda Evento "Idro" 31 Ottobre–5 Novembre 2010, available at: http://www.regione.veneto.it/ (last access: 21 March 2019), 2010.

Arnell, N. W., & Gosling, S. N. (2016). The impacts of climate change on river flood risk at the global scale. *Climatic Change*, *134*(3), 387-401.

Atreya, A., Czajkowski, J., Botzen, W., Bustamante, G., Campbell, K., Collier, B., ... & Montgomery, M. (2017). Adoption of flood preparedness actions: A household level study in rural communities in Tabasco, Mexico. *International journal of disaster risk reduction*, *24*, 428-438.

Austin, P.C., & Mamdani, M.M. (2006). A Comparison of prognosis score methods: a case-study estimating the effectiveness of post-AMI statin use. *Statistics in medicine, 25*(12)*,* 2084-2106.

Babcicky, P., & Seebauer, S. (2019). Collective efficacy and natural hazards: differing roles of social cohesion and task-specific efficacy in shaping risk and coping beliefs. *Journal of Risk Research*, 1-18.

Barendrecht, M. H., Viglione, A., & Blöschl, G. (2017). A dynamic framework for flood risk. *Water Security*, *1*, 3-11.

Barredo, J.I. (2009). Normalized flood losses in Europe: 1970-2006. *Natural Hazards and Earth System Sciences, 9*(1), 97-104.

Belcaro, P., Gasparini, D., & Baldessari, M.: 31 ottobre–2 novembre 2010: l'alluvione dei Santi, Regione Veneto, available at: http://statistica.regione.veneto.it/ (last access: 21 March 2019), 2011.

Blake, E. S., & Zelinsky, D. A. (2018). Tropical cyclone report Hurricane Harvey (Tech. Rep. No. AL092017). Miami, FL: US Department of Commerce National Oceanic and Atmospheric Administration National Hurricane Center. Retrieved from https://www.nhc.noaa.gov/data/tcr/AL092017_Harvey.pdf

Blöschl, G., Hall, J., Parajka, J., Perdigão, R. A., Merz, B., Arheimer, B., ... & Čanjevac, I. (2017). Changing climate shifts timing of European floods. *Science, 357*(6351), 588-590.

Botzen, W. W., Bouwer, L. M., Scussolini, P., Kuik, O., Haasnoot, M., Lawrence, J., & Aerts, J. C. (2019). Integrated disaster risk management and adaptation. In *Loss and damage from climate change* (pp. 287-315). Springer, Cham.

Breiman, L. (2001). Random forests. *Machine learning, 45*(1), 5–32.

Brooks, N., Adger, W. N., & Kelly, P. M. (2005). The determinants of vulnerability and adaptive capacity at the national level and the implications for adaptation. *Global environmental change*, *15*(2), 151-163.

Bruneau, M., Chang, S. E., Eguchi, R. T., Lee, G. C., O'Rourke, T. D., Reinhorn, A. M., ... & Von Winterfeldt, D. (2003). A framework to quantitatively assess and enhance the seismic resilience of communities. *Earthquake spectra*, *19*(4), 733-752.

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

Bubeck, P., Botzen, W.J.W., Kreibich, H., & Aerts, J.C.J.H. (2012). Long-term development and effectiveness of private flood mitigation measures. *Natural Hazards and Earth System Sciences*, *12*(11), 3507-3518.

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., Botzen, W.J.W., Laudan, J., Aerts, J.C.J.H. & Thieken, A.H. (2017): Insights into flood-coping appraisals of Protection Motivation Theory: Empirical evidence from Germany and France. *Risk Analysis*, online first 17 Nov 2017, DOI:10.1111/risa.12938.

Bubeck, P., & Thieken, A. H. (2018). What helps people recover from floods? Insights from a survey among flood-affected residents in Germany. *Regional environmental change*, *18*(1), 287-296.

Buck, W., & Merkel, U. (1999). *Auswertung der HOWAS-Schadendatenbank* (No. HY98/15). Institut für Wasserwirtschaft und Kulturtechnik: Universität Karlsruhe.

Bundesministerium des Innern. (2013). *Bericht zur Flutkatastrophe 2013*: Katastrophenhilfe, Entschädigung, Wiederaufbau. Tech. rept.

Burby, R.J., Bollens, S.A., Holloway, I.M., Kaiser, E.J., Mullan, D. & Sheaffer, J.R. (1988) Cities under Water: A Comparative Evaluation of Ten Cities. Efforts to Manage Flood Plain. *Land Institute of Behavioral Science, University of Colorado,* Boulder, CO

Bürkner, P.C. (2017), 'brms: An R Package for Bayesian Multilevel Models Using Stan'*, Journal of Statistical Software 80*(1), 1–28.

Caliendo, M., Kopeinig, S. (2008). Some practical guidance for the implementation of propensity score matching. *Journal of Economic Surveys, 22*(1), 31-72.

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.: Development and assessment of uni- and multi-variable flood loss models for Emilia-Romagna (Italy), *Nat. Hazards Earth Syst. Sci. Discuss*., https://doi.org/10.5194/nhess-2017-342, 2018.

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., ... & Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of statistical software*, *76*(1).

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. doi:10.3390/w9050313.

Chipman, H. A., George, E. I., & McCulloch, R. E. (2010). BART: Bayesian additive regression trees. *The Annals of Applied Statistics*, *4*(1), 266-298.

Crichton, D. (2012). Flood plain Speaking. URL: www.floodproblems.co.uk/wp-content/uploads/2012/12/Crichton_2012_flood_plain_speaking_3rd_ed.pdf (last accessed 01/03/2020)

Cumbria County Council. (2018). Flooding in Cumbria, December 2015 Impact Assessment, Carlisle, UK: Cumbria County Council – Performance and Intelligence Team. URL: https://www.cumbria.gov.uk/eLibrary/Content/Internet/536/671/4674/17217/17225/43312152830.pdf (last accessed 9 May 2019).

Cumiskey, L., Priest, S., Valchev, N., Viavattene, C., Costas, S., & Clarke, J. (2018). A framework to include the (inter) dependencies of disaster risk reduction measures in coastal risk assessment. *Coastal Engineering*, *134*, 81-92.

D'Alpaos, L., Brath, A., & Fioravante, V.: Relazione tecnicoscientifica sulle cause del collasso dell' argine del fiume Secchia avvenuto il giorno 19 gennaio 2014 presso la frazione San Matteo, available at: http://www.comune.bastiglia.mo.it/ (last access: 21 March 2019), 2014.

da Costa, R. T., Manfreda, S., Luzzi, V., Samela, C., Mazzoli, P., Castellarin, A., & Bagli, S. (2019). A web application for hydrogeomorphic flood hazard mapping. *Environmental Modelling & Software, 118*, 172-186.

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.

Dasgupta, A., Sun, Y. V., König, I. R., Bailey-Wilson, J. E., & Malley, J. D. (2011). Brief review of regression-based and machine learning methods in genetic epidemiology: the genetic analysis workshop 17 experience. *Genetic epidemiology, 35* (S1).

Dawson, RJ, Ball, T., Werritty, J., Werritty, A., Hall, JW, & Roche, N. (2011). Assessing the effectiveness of non-structural flood management measures in the Thames Estuary under conditions of socio-economic and environmental change. *Global Environmental Change*, *21*(2), 628-646.

De Bruijn, K.M. (2004) Resilience and flood risk management. *Water Policy 6*(1) 53-66.

De Bruijn, K. M., Wagenaar, D. J., Slager, K., De Bel, M., & Burzel, A.: Proposal for the new flood damage assessment method: SSM2015, Deltares, 2014

De Moel, H., van Vliet, M., & Aerts, J. C. (2014). Evaluating the effect of flood damage-reducing measures: a case study of the unembanked area of Rotterdam, the Netherlands. *Regional environmental change*, *14*(3), 895-908.

De Moel, H., Jongman, B., Kreibich, H., Merz, B., Penning-Rowsell, E. & Ward, P. J. (2015). Flood risk assessments at different spatial scales. *Mitigation and Adaptation Strategies for Global Change, 20*(6), 865-890. DOI: http://doi.org/10.1007/s11027-015-9654-z

Defra, A. (2008). Framework for pro-environmental behaviors. *Department for Environment, Food and Rural Affairs,* London.

Defra/Environment Agency. (2009). Review of the Pilot Flood Protection Grant Scheme in a Recently Flooded Area. (Report no. FD2651/TR). URL: http://evidence.environment-agency.gov.uk/FCERM/en/Default/FCRM/Project.aspx?ProjectID=73f1f745-5d08-42dc-8093-145afae754ea&PageID=25ed1548-e755-452a-9d94-f2fc7d984e56 (last accessed 7 March 2019).

Dehejia, R.H. & Wahba, S. (1999). Causal effects in nonexperimental studies: Reevaluating the evaluation of training programs. *Journal of the American Statistical Association*, *94*(448), 1053-1062.

Desai, B., Maskrey, A., Peduzzi, P., De Bono, A., & Herold, C. (2015). Making development sustainable: the future of disaster risk management, global assessment report on disaster risk reduction.

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).

Di Baldassarre, G., Viglione, A., Carr, G., Kuil, L., Salinas, J. L., & Blöschl, G. (2013). Socio-hydrology: conceptualising human-flood interactions. *Hydrology and Earth System Sciences*, *17*(8), 3295.

Diamond, A., & Sekhon, J. S. (2006). Genetic Matching for Estimating Causal Effects: A General Multivariate Matching Method for Achieving Balance in Observational Studies. *UC Berkeley: Institute of Governmental Studies*. Retrieved from https://escholarship.org/uc/item/8gx4v5qt

Dietz, H.: Wohngebäaudeversicherung. Kommentar, 2nd edition, VVW Verlag Versicherungswirtschaft GmbH, Karlsruhe, 756 pp., 1999.

DiPrete, T. A., & Gangl, M. (2004). 7. Assessing Bias in the Estimation of Causal Effects: Rosenbaum Bounds on Matching Estimators and Instrumental Variables Estimation with Imperfect Instruments. *Sociological methodology*, *34*(1), 271-310.

Dittes, B., Kaiser, M., Špačková, O., Rieger, W., Disse, M., & Straub, D. (2018). Risk-based flood protection planning under climate change and modeling uncertainty: A pre-alpine case study. *Natural Hazards and Earth System Sciences, 18*(5), 1327–1347.

Domeneghetti, A., Carisi, F., Castellarin, A., & Brath, A. (2015). Evolution of flood risk over large areas: Quantitative assessment for the Po river. *Journal of Hydrology*, *527*, 809-823.

Donat, M. G., Lowry, A. L., Alexander, L. V., O'Gorman, P. A., & Maher, N. (2016). More extreme precipitation in the world's dry and wet regions. *Nature Climate Change, 6*(5), 508–513.

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, *Nat. Hazards Earth Syst. Sci., 16*(12), 2577–2591, doi:10.5194/nhess-16-2577-2016.

Duiser, J. A. (1982). Een verkennend onderzoek naar methoden ter bepaling van inundatieschade bij dijkdoorbraak, 82-0644 TNO, Delft, the Netherlands

Dutta, D., Herath, S., & Musiake, K. (2003). A mathematical model for flood loss estimation. *Journal of hydrology*, *277*(1-2), 24-49.

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.

Environment Agency DEFRA. (2019): Recorded Flood Outlines. URL: https://environment.data.gov.uk/dataset/8c75e700-d465-11e4-8b5b-f0def148f590 (last accessed 01/10/2019)

Egorova, R., van Noortwijk, J. M. & Holterman, S. R. (2008), 'Uncertainty in flood damage estimation', *International Journal of River Basin Management 6*(2), 139–148.

Falter, D., Schröter, K., Dung, N. V., Vorogushyn, S., Kreibich, H., Hundecha, Y., Apel, H. & Merz, B. (2015). Spatially coherent flood risk assessment based on long-term continuous simulation with a coupled model chain. *Journal of Hydrology, 524*, 182–193.

Federal Emergency Management Agency (2018a). Housing assistance data. Retrieved from https://www.fema.gov/media-library/assets/documents/34758, (Accessed: 2018-09-12).

FEMA. (2018b). Mandatory Purchase of NFIP Coverage. https://www.fema.gov/ faq-details/Mandatory-Purchase-of-NFIP-Coverage. (Accessed: 2018-09- 12)

Fekete, A., Hufschmidt, G., & Kruse, S. (2014). Benefits and challenges of resilience and vulnerability for disaster risk management. *International journal of disaster risk science*, *5*(1), 3-20.

Feller, A., & Gelman, A. (2015). Hierarchical models for causal effects. *Emerging Trends in the Social and Behavioral Sciences: An interdisciplinary, searchable, and linkable resource*, 1-16.

Ferrari, S., & Cribari-Neto, F. (2004). Beta regression for modelling rates and proportions. *Journal of Applied Statistics, 31*(7), 799–815.

Few, R. (2003). Flooding, vulnerability and coping strategies: local responses to a global threat. *Progress in Development Studies*, *3*(1), 43-58.

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.

Ford, J. D., Tilleard, S. E., Berrang-Ford, L., Araos, M., Biesbroek, R., Lesnikowski, A. C., ... & Bizikova, L. (2016). Opinion: Big data has big potential for applications to climate change adaptation. *Proceedings of the National Academy of Sciences*, *113*(39), 10729-10732.

Freni, G., La Loggia, G., & Notaro, V. (2010). Uncertainty in urban flood damage assessment due to urban drainage modelling and depth-damage curve estimation. *Water Science and Technology,61*(12), 2979–2993.

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *Annals of Statistics, 29*, 1189–1232.

Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical science, 7 (4),* 457–472.

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.

Gneiting, T. & Raftery, A. E. (2007): Strictly Proper Scoring Rules, Prediction, and Estimation, *Journal of the American Statistical Association, 102*(477), 359–378, doi:10.1198/016214506000001437.

Gouldby, B., Samuels, P., Klijn, F., Messner, F., Van, Os A., Sayers, P., & Schanze, J. (2005) Language of Risk - Project definitions. *FLOODsite Report: T32-04-01*, accessed (29/06/2017).

Grigg, N.S. & Helweg, O. J. (1975). State-Of-The-Art of estimating flood damage in urban areas. *JAWRA Journal of the American Water Resources Association*, *11*(2), 379-390.

Grömping, U. (2009). Variable importance assessment in regression: linear regression versus random forest. *The American Statistician*, *63*(4)*,* 308–319.

Güneralp, B., Güneralp, İ., & Liu, Y. (2015). Changing global patterns of urban exposure to flood and drought hazards. *Global environmental change*, *31*, 217-225.

Haer, T., Botzen, W. W., & Aerts, J. C. (2016). The effectiveness of flood risk communication strategies and the influence of social networks—Insights from an agent-based model. *Environmental Science & Policy*, *60*, 44-52.

Hasanzadeh Nafari, R., Ngo, T., & Mendis, P. (2016). An assessment of the effectiveness of tree-based models for multi-variate flood damage assessment in Australia. *Water*, *8*(7), 282.

HCAD (2018). "Harris County appraisal district—Real and personal property database". Retrieved from http://pdata.hcad.org/download/ index.html, (Accessed: 2018-09-12).

Heckerman, D., Geiger, D., & Chickering (1995), D.M. Machine Learning, 20-197. https://doi.org/10.1023/A:1022623210503.

Heckerman, D. (1998). A tutorial on learning with Bayesian networks. *In Learning in graphical models*. Springer Netherlands, 301-354.

Ho, D. E., Imai, K., King, G., & Stuart, E. A. (2007). Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. *Political analysis*, *15*(3), 199-236.

Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics, 12*(1), 55–67.

Hoffman, M. D., & Gelman, A. (2014). The no-u-turn sampler: adaptively setting path lengths in hamiltonian monte carlo. *Journal of Machine Learning Research*, *15*(1)*,* 1593–1623.

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

Holub, M., & Fuchs, S. (2008). Benefits of local structural protection to mitigate torrent-related hazards. *WIT Transactions on Information and Communication Technologies*, *39*, 401-411.

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.

ICPR (2002). International Commission for the Protection of the Rhine: Non-Structural Flood Plain Management – Measures and their Effectiveness, Koblenz.

IPCC (2012). Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation. A Special Report of Working Groups I and II of the Intergovernmental Panel on Climate Change [Field, C.B., V. Barros, T.F. Stocker, D. Qin, D.J. Dokken, K.L. Ebi, M.D. Mastrandrea, K.J. Mach, G.-K. Plattner, S.K. Allen, M. Tignor, and P.M. Midgley (eds.)]. Cambridge University Press, Cambridge, UK, and New York, NY, USA, 582 pp.

JBA Risk Management (2017). Oasis hub—Pluvial flooding footprint—Hurricane Harvey—28th August 2017. Retrieved from https:// oasishub.co/dataset/surface-water-flooding-footprint-hurricane-harvey-august-2017-jba, (Accessed: 2018-09-12).

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

Jongman, B., Hochrainer-Stigler, S., Feyen, L., Aerts, J. C., Mechler, R., Botzen, W. W., ... & Ward, P. J. (2014). Increasing stress on disaster-risk finance due to large floods. *Nature Climate Change*, *4*(4), 264-268.

Jongman, B., Winsemius, H. C., Aerts, J. C., De Perez, E. C., Van Aalst, M. K., Kron, W., & Ward, P. J. (2015). Declining vulnerability to river floods and the global benefits of adaptation. *Proceedings of the National Academy of Sciences*, *112*(18), E2271-E2280.

Kass, R. E., Carlin, B. P., Gelman, A., & Neal, R. M. (1998). Markov chain monte carlo in practice: a roundtable discussion. *The American Statistician*, *52*(2), 93–100.

Keele, L. (2010). An overview of rbounds: An R package for Rosenbaum bounds sensitivity analysis with matched data. *White Paper*. Columbus, OH, 1-15.

Keller, L., Zischg, A. P., Mosimann, M., Rössler, O., Weingartner, R., & Martius, O. (2019). Large ensemble flood loss modelling and uncertainty assessment for future climate conditions for a Swiss pre-alpine catchment. *Science of the total environment*, *693*, 133400.

Kelly, P.M., & Adger, W.N. (2000). Theory and practice in assessing vulnerability to climate change and facilitating adaptation. *Climatic change*, *47* (4), 325-352.

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. *Floods across Europe. Hazard assessment, modelling and management. London. S*, 67-106.

Kok, M., Huizinga, H. J., Vrouwenvelder, A. C. W. M., & van den Braak, W. E. W. (2005). Standaardmethode 2005, Schade en Slachtoffers als gevolg van overstroming, HKV, *TNObouw, Rijkswaterstaat DWW*.

Kreibich, H., Thieken, A.H., Petrow, T., Müller, M. & Merz, B. (2005). Flood loss reduction of private households due to building precautionary measures--lessons learned from the Elbe flood in August 2002. *Natural Hazards and Earth System Science, 5*(1), 117-126.

Kreibich, H. & Thieken, A.H. (2009). Coping with floods in the city of Dresden, Germany. *Natural Hazards, 51*(3), 423.

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., Christenberger, S., & Schwarze, R. (2011). Economic motivation of households to undertake private precautionary measures against floods. *Natural Hazards and Earth System Sciences, 11*(2), 309.

Kreibich, H., Van Den Bergh, J. C., Bouwer, L. M., Bubeck, P., Ciavola, P., Green, C., ... & Thieken, A. H. (2014). Costing natural hazards. *Nature Climate Change*, *4*(5), 303.

Kreibich, H., Bubeck, P., Van Vliet, M., & De Moel, H. (2015). A review of damage-reducing measures to manage fluvial flood risks in a changing climate. *Mitigation and adaptation strategies for global change*, *20*(6), 967-989.

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., ... & Chinh, D. T. (2017). Adaptation to flood risk: Results of international paired flood event studies. *Earth's Future*, *5*(10), 953-965.

Kruschke, JK, & Vanpaemel, W. (2015). Bayesian estimation in hierarchical models. *The Oxford Handbook of computational and mathematical psychology*, 279-299.

Kuhn, M. (2008). Caret package. *Journal of statistical software, 28*(5)*,* 1–26.

Kuhn, M., & Johnson, K. (2013). *Applied predictive modeling* (26). Springer.

Kundzewicz, Z. W., Kanae, S., Seneviratne, S. I., Handmer, J., Nicholls, N., Peduzzi, P., et al., (2014). Flood risk and climate change: Global and regional perspectives. *Hydrological Sciences Journal, 59*(1), 1–28.

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

Levy, R. (2012). Probabilistic models in the study of language. *Online Draft*.

Liaw, A., & Wiener, M. (2002). Classification and regression by random forest. *R News, 2*(3), 18-22. Retrieved from https://CRAN.R-project.org/doc/ Rnews/

Lüdtke, S., Schröter, K., Steinhausen, M., Weise, L., Figueiredo, R., & Kreibich, H. (2019). A consistent approach for probabilistic residential flood loss modeling in Europe. *Water Resources Research*, *55*. https://doi.org/10.1029/2019WR026213

Mechler, R., & Bouwer, L. M. (2015). Understanding trends and projections of disaster losses and climate change: is vulnerability the missing link? *Climatic Change*, *133*(1), 23-35.

Merz, B., Kreibich, H., Thieken, A., & Schmidtke, R. (2004). Estimation uncertainty of direct monetary flood damage to buildings. *Natural Hazards and Earth System Science, 4*(1), 153–163.

Merz, B., Hall, J., Disse, M., & Schumann, A. (2010). Fluvial flood risk management in a changing world. *Natural Hazards and Earth System Sciences*, *10*(3), 509.

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.

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., Elmer, F., Kunz, M., Mühr, B., Schröter, K., & Uhlemann-Elmer, S. (2014). The extreme flood in June 2013 in Germany. *La Houille Blanche*, (1), 5-10.

Merz, B., Aerts, J., Arnbjerg-Nielsen, K., Baldi, M., Becker, A., Bichet, A., Blöschl, G., Bouwer, L. M., Brauer, A., Cioffi, F., Delgado, J. M., Gocht, M., Guzzetti, F., Harrigan, S., Hirschboeck, K., Kilsby, C., Kron, W., Kwon, H.-H., Lall, U., Merz, R., Nissen, K., Salvatti, P., Swierczynski, T., Ulbrich, U., Viglione, A., Ward, P. J., Weiler, M., Wilhelm, B., & Nied, M. (2014). Floods and climate: emerging perspectives for flood risk assessment and management, *Nat. Hazards Earth Syst. Sci., 14*, 1921–1942, https://doi.org/10.5194/nhess-14-1921-2014

Merz, B., Vorogushyn, S., Lall, U., Viglione, A. & Blöschl, G. (2015). Charting unknown waters - On the role of surprise in flood risk assessment and management. *Water Resources Research, 51*, 8, 6399-6416. https://doi.org/10.1002/2015WR017464

Metin, A. D., Nguyen, D., Schröter, K., Guse, B., Apel, H., Kreibich, H., ... & Merz, B. (2018). How do changes along the risk chain affect flood risk? *Natural Hazards and Earth System Sciences (NHESS)*, *18*, 3089-3108.

Muis, S., Güneralp, B., Jongman, B., Aerts, J. C., & Ward, P. J. (2015). Flood risk and adaptation strategies under climate change and urban expansion: A probabilistic analysis using global data. *Science of the Total Environment*, *538*, 445-457.

National Response Center (2018). United States Coast Guard—National Response Center Database. Retrieved from http://www.nrc.uscg. mil/FOIAFiles/CY17.xlsx, (Accessed: 2018-09-12).

Neumayer, E., & Barthel, F. (2011). Normalizing economic loss from natural disasters: A global analysis. *Global Environmental Change*, *21*(1), 13-24.

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.

Olsen, A. S., Zhou, Q., Linde, J. J., & Arnbjerg-Nielsen, K. (2015). Comparing methods of calculating expected annual damage in urban pluvial flood risk assessments. *Water, 7*(1), 255–270.

Ospina, R., & Ferrari, S. L. (2010). Inflated beta distributions. *Statistical Papers, 51*(1), 111.

Paprotny, D., Morales-Nápoles, O., & Jonkman, S. N. (2017). Efficient pan-European river flood hazard modelling through a combination of statistical and physical models. *Natural Hazards and Earth System Sciences*, *17*(7), 1267.

Paprotny, D., Sebastian, A., Morales-Nápoles, O., & Jonkman, S. N. (2018). Trends in flood losses in Europe over the past 150 years. *Nature communications*, *9*(1), 1985.

Parker, D. J., Green, C. H., & Thompson, P. M. (1987). *Urban Flood Protection Benetits: a Project Appraisal Guide*. Gower technical press.

Parker, D., Tapsell, S., & McCarthy, S. (2007). Enhancing the human benefits of flood warnings. *Natural Hazards*, *43*(3), 397-414.

Pearl, J. (2009). Causal inference in statistics: An overview. *Statistics Surveys, 3,* 96-146.

Penning-Rowsell, E. C. & Chatterton, J. B. (1977). The benefits of flood alleviation: A manual of assessment techniques, *Gower Technical Press*, Aldershot.

Penning-Rowsell, E., Floyd, P., Ramsbottom, D., & Surendran, S. (2005). Estimating injury and loss of life in floods: a deterministic framework. *Natural Hazards*, *36*(1), 43-64.

PNNL. (2017). *Pacific Northwest National Laboratory Hurricane Harvey Summary - Revised Dry Time.* https://apps.pnnl.gov/portal/apps/MapSeries/index.html?appid=9b53801617e9426a81a77e1774a44fad. (Accessed: 2018-09-12)

Polasky, S., Carpenter, S. R., Folke, C. & Keeler, B. (2011). Decision-making under great uncertainty: environmental management in an era of global change, *Trends in Ecology & Evolution, 26*(8), 398–404, doi:10.1016/j.tree.2011.04.007

Poser, K., Dransch, D. (2010): Volunteered geographic information for disaster management with application to rapid flood damage estimation. *Geomatica, 64*, 1, 89-98

Poussin, J. K., Botzen, W. W., & Aerts, J. C. (2015). Effectiveness of flood damage mitigation measures: Empirical evidence from French flood disasters. *Global Environmental Change, 31*, 74-84.

R Core Team (2016). R: A language and environment for statistical computing. *R Foundation for Statistical Computing*, Vienna, Austria. URL https://www.R-project.org/.

R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Richert, C., Erdlenbruch, K., & Figuières, C. (2017). The determinants of households' flood mitigation decisions in France-on the possibility of feedback effects from past investments. *Ecological Economics*, *131*, 342-352.

Rojas, R., Feyen, L., & Watkiss, P. (2013). Climate change and river floods in the European Union: Socio-economic consequences and the costs and benefits of adaptation. *Global Environmental Change*, *23*(6), 1737-1751.

Rolfsen, Michael. (2009). Das neue wasserhaushaltsgesetz. *Natur und Recht,* 765.

Rosenbaum, PR, & Ruby, DB (1983). The central role of the propensity score in observational studies for causal effects. *Biometrics*, *70*(1), 41-55.

Rosenbaum, P. R., & Rubin, D. B. (1985). Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *The American Statistician*, 39(1), 33-38.

Rosenbaum, P. R. (2002). Observational studies. In *Observational studies* (pp. 1-17). Springer, New York, NY.

Rosenzweig, B. R., McPhillips, L., Chang, H., Cheng, C., Welty, C., Matsler, M., et al., (2018). Pluvial flood risk and opportunities for resilience. Wiley Interdisciplinary Reviews: *Water, 5*, e1302.

Roy, A. D. (1951). Some thoughts on the distribution of earnings. *Oxford economic papers, 3*(2), 135-146.

Rözer, V., Kreibich, H., Schröter, K., Müller, M., Sairam, N., Doss-Gollin, J., Lall, U., Merz, B., (2019). Probabilistic models significantly reduce uncertainty in Hurricane Harvey pluvial flood loss estimates. *Earth's Future.*

Rözer, V., Müller, M., Bubeck, P., Kienzler, S., Thieken, A., Pech, I., et al., (2016). Coping with pluvial floods by private households. *Water, 8*(7), 304.

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of educational Psychology*, *66*(5), 688.

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*. https://doi.org/10.1029/2019WR025068

Sairam, N., Schröter, K., Lüdtke, S., Merz, B., and Kreibich, H (2019). Quantifying Flood Vulnerability Reduction via Private Precaution. *Earth's Future.*

Scawthorn, C., Flores, P., Blais, N., Seligson, H., Tate, E., Chang, S., et al., (2006). Hazus-MH flood loss estimation methodology. II. Damage and loss assessment. *Natural Hazards Review, 7*(2), 72–81.

Schröter, K., Kreibich, H., Vogel, K., Riggelsen, C., Scherbaum, F., & Merz, B. (2014). How useful are complex flood damage models? *Water Resources Research*, *50*(4), 3378-3395.

Schröter, K., Kunz, M., Elmer, F., Mühr, B., & Merz, B. (2015). What made the June 2013 flood in Germany an exceptional event? A hydrometeorological evaluation. *Hydrology and Earth System Sciences*, *19*(1), 309–327. http://doi.org/10.5194/hess-19-309-2015

Schröter, K., Lüdtke, S., Vogel, K., Kreibich, H., & Merz, B. (2016). Tracing the value of data for flood loss modelling. *In E3s web of conferences*, (Vol. 7, pp. 05005).

Schröter, K., Lüdtke, S., Redweik, R., Meier, J., Bochow, M., Ross, L., et al., (2018). Flood loss estimation using 3D city models and remote sensing data. *Environmental Modelling & Software, 105*, 118–131.

Scorzini A.R., Radice A., Molinari D. (2018). A New Tool to Estimate Inundation Depths by Spatial Interpolation (RAPIDE): Design, Application and Impact on Quantitative Assessment of Flood Damage. *Water, 10*, 1805; doi:10.3390/w10121805

Scutari, M. (2009). Learning Bayesian networks with the bnlearn R package. *Journal of Statistical Software, 35(3),* 1-22. URL http://www.jstatsoft.org/v35/i03/.

Sieg, T., Vogel, K., Merz, B. and Kreibich, H. (2019), 'Seamless estimation of hydrometeorological risk across spatial scales', *Earth's Future 7*(5)*,* 574–581.

Sierra Club. (2017). *Hurricane Harvey Toxic Sites v1.1*. https://www.sierraclub .org/environmental-justice/hurricane-harvey-toxic-sites. (Accessed: 2018-09-12)

Simpson, D., Rue, H., Riebler, A., Martins, T. G., & Sørbye, S. H. (2017). Penalising model component complexity: A principled, practical approach to constructing priors. *Statistical Science, 32*(1), 1–28.

Skougaard Kaspersen, P., Høegh Ravn, N, Arnbjerg-Nielsen, K., Madsen, H., & Drews, M. (2015). Influence of urban land cover changes and climate change for the exposure of european cities to flooding during high-intensity precipitation. *Proceedings of the International Association of Hydrological Sciences, 370*, 21–27.

Smith, D. I. (1994). Flood damage estimation- A review of urban stage-damage curves and loss functions. *Water S. A.*, *20*(3), 231-238.

Spekkers, M., Kok, M., Clemens, F., & Ten Veldhuis, J. (2014). Decision-tree analysis of factors influencing rainfall-related building structure and content damage. *Natural Hazards and Earth System Sciences, 14*(9), 2531–2547.

Spekkers, M., Rözer, V., Thieken, A., ten Veldhuis, M.-C., & Kreibich, H. (2017). A comparative survey of the impacts of extreme rainfall in two international case studies. *Natural Hazards and Earth System Sciences, 17*(8), 1337–1355.

Strobl, C., Boulesteix, A.-L., Zeileis, A., & Hothorn, T. (2007). Bias in random forest variable importance measures: Illustrations, sources and a solution. *BMC Bioinformatics, 8*(1), 25.

Sun, X., U. Lall, B. Merz, and N. V. Dung (2015), Hierarchical Bayesian clustering for nonstationary flood frequencyanalysis: Application to trends ofannual maximum flow in Germany,*Water Resour. Res.,51*, 6586–6601,doi:10.1002/2015WR017117.

Szönyi, M., May, P. & Lamb, R. (2016). Flooding after Storm Desmond. UK: Zurich Insurance Group. https://www.jbatrust.org/wp-content/uploads/2016/08/flooding-after-storm-desmond-PUBLISHED-24-August-2016.pdf (last accessed 9 May 2019).

Teng, J., Jakeman, A. J., Vaze, J., Croke, B. F., Dutta, D., & Kim, S. (2017). Flood inundation modelling: A review of methods, recent advances and uncertainty analysis. *Environmental modelling & software, 90*, 201-216.

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*(12).

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., 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.

Thieken, A. H., Bessel, T., Kienzler, S., Kreibich, H., Müller, M., Pisi, S., & Schröter, K. (2016). The flood of June 2013 in Germany: how much do we know about its impacts. *Nat. Hazards Earth Syst. Sci*, *16*(6), 1519-1540.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological), 58*, 267–288.

Todini, E. (2018). Paradigmatic changes required in water resources management to benefit from probabilistic forecasts. *Water Security, 3*, 9–17.

Tsamardinos, I., Aliferis, C. F., Statnikov, A. R., & Statnikov, E. (2003). Algorithms for Large Scale Markov Blanket Discovery. In *FLAIRS conference* (Vol. 2, pp. 376-380).

Turner, B. L., Kasperson, R. E., Matson, P. A., McCarthy, J. J., Corell, R. W., Christensen, L., ... & Polsky, C. (2003). A framework for vulnerability analysis in sustainability science. *Proceedings of the national academy of sciences*, *100*(14), 8074-8079.

URBAS. (2008). *Fallstudien und Untersuchungsschwerpunkte Hamburg bis Lohmar* (Tech. Rep.). Aachen, Germany: BMBF.
http://www.urbanesturzfluten.de/schlussbericht/fallstudien%20Hamburg%20bis%20Lohmar/download.

U.S. Census Bureau (2016). Occupancy characteristics 2012–2016 American Community survey 5-year estimates. Retrieved from https://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?pid=ACS_16_5YRB25010&prodType=table, (Accessed: 2018-09-12).

Van Ootegem, L., Van Herck, K., Creten, T., Verhofstadt, E., Foresti, L., Goudenhoofdt, E., et al., (2018). Exploring the potential of multivariate depth-damage and rainfall-damage models. *Journal of Flood Risk Management, 11*, S916–S929.

Van Ootegem, L., Verhofstadt, E., Van Herck, K., & Creten, T. (2015). Multivariate pluvial flood damage models. *Environmental Impact Assessment Review, 54*, 91–100.

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.

Veldhuis, J. (2011). How the choice of flood damage metrics influences urban flood risk assessment. *Journal of Flood Risk Management, 4*(4), 281–287.

Venables, W. N. and Ripley, B. D. (2002) *Modern Applied Statistics with S.* New York: Springer (4th ed).

Vincent, J. L., Baron, J. F., Reinhart, K., Gattinoni, L., Thijs, L., Webb, A., & ABC Investigators. (2002). Anemia and blood transfusion in critically ill patients. *Jama*, *288*(12), 1499-1507.

Vogel, K., Riggelsen, C., Korup, O., & Scherbaum, F. (2014). Bayesian network learning for natural hazard analyses. *Natural Hazards and Earth System Sciences*, *14*(9), 2605.

Vogel, K., Weise, L., Schröter, K., & Thieken, A. H. (2018). Identifying Driving Factors in Flood-Damaging Processes Using Graphical Models. *Water Resources Research*, *54*(11), 8864-8889.

Vorogushyn, S, Bates, PD, de Bruijn, K, et al., (2018) Evolutionary leap in large-scale flood risk assessment needed. *WIREs Water*, 5:e1266. https://doi.org/10.1002/wat2.1266

Ward, P. J., Jongman, B., Weiland, F. S., Bouwman, A., van Beek, R., Bierkens, M. F., ... & Winsemius, H. C. (2013). Assessing flood risk at the global scale: model setup, results, and sensitivity. *Environmental Research Letters*, *8*(4), 044019.

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.

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., Lüdtke, S., Schröter, K., € Bouwer, L. M., & Kreibich, H. (2018). Regional and temporal transferability of multivariable flood damage models. *Water Resources Research, 54,* 3688– 3703. https://doi.org/10.1029/ 2017WR022233

Wang, K., Wang, L., Wei, Y.-M., & Ye, M. (2013). Beijing storm of July 21, 2012: Observations and reflections. *Natural Hazards, 67*(2), 969–974.

Ward, P. J., Blauhut, V., Bloemendaal, N., Daniell, J. E., de Ruiter, M. C., Duncan, M., Emberson, R., Jenkins, S. F., Kirschbaum, D., Kunz, M., Mohr, S., Muis, S., Riddell, G., Schäfer, A., Stanley, T., Veldkamp, T. I. E., and Winsemius, H. C. (2019) Review article: Natural hazard risk assessments at the global scale, *Nat. Hazards Earth Syst. Sci. Discuss*., https://doi.org/10.5194/nhess-2019-403, in review.

Warton, D. I., & Hui, F. K. (2011). The arcsine is asinine: the analysis of proportions in ecology. *Ecology, 92* (1), 3–10

White, G. F. (1964). Choice of adjustment to floods. *CHICAGO, UNIVERSITY OF CHICAGO, 1964.* 150 P.

Wind, H. G., Nierop, T. M., Blois, C. D., & Kok, J. D. (1999). Analysis of flood damages from the 1993 and 1995 Meuse floods. *Water resources research*, *35*(11), 3459-3465.

Wing, O. E., Bates, P. D., Smith, A. M., Sampson, C. C., Johnson, K. A., Fargione, J., & Morefield, P. (2018). Estimates of present and future flood risk in the conterminous united states. *Environmental Research Letters, 13*(3), 034023.

Winsemius, H. C., Aerts, J. C., Van Beek, L. P., Bierkens, M. F., Bouwman, A., Jongman, B., ... & Ward, P. J. (2016). Global drivers of future river flood risk. *Nature Climate Change*, *6*(4), 381-385.

Winter, B., Schneeberger, K., Huttenlau, M., & Stötter, J. (2018). Sources of uncertainty in a probabilistic flood risk model. *Natural hazards*, *91*(2), 431-446.

WL Delft (1994). Onderzoek watersnood Maas, Deelrapport 1: Wateroverlast December, Deelrapport 9: Schade, WL Delft, Delft, Appendix A (in Dutch).

Wojcik, O., Holt, J., Kjerulf, A., Müller, L, Ethelberg, S., & Mølbak, K (2013). Personal protective equipment, hygiene behaviours and occupational risk of illness after July 2011 flood in Copenhagen, Denmark. *Epidemiology & Infection, 141*(8), 1756–1763.

Yang, L. E., Scheffran, J., Süsser, D., Dawson, R., & Chen, Y. D. (2018). Assessment of flood losses with household responses: Agent-based simulation in an urban catchment area. *Environmental Modeling & Assessment*, *23*(4), 369-388.

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.

Zhou, Q., Mikkelsen, P. S., Halsnæs, K., & Arnbjerg-Nielsen, K. (2012). Framework for economic pluvial flood risk assessment considering climate change effects and adaptation benefits. *Journal of Hydrology, 414*, 539–549.

Zischg, A. P., Felder, G., Mosimann, M., Röthlisberger, V., & Weingartner, R. (2018). Extending coupled hydrological-hydraulic model chains with a surrogate model for the estimation of flood losses. *Environmental modelling & software*, *108*, 174-185.