Uncertainty quantification and attribution in flood risk modelling

By

Georgios Sarailidis

Department of Civil Engineering

UNIVERSITY OF BRISTOL

A dissertation submitted to the University of Bristol in accordance with the requirements of the degree of DOCTOR OF PHILOSOPHY in the Faculty of Engineering

March 2023

Word Count: 23152

Abstract

Flood risk models are used to help manage the financial and societal risks associated with floods. They quantify risk (usually in terms of flood losses) by modelling the underlying physical hazard, the exposure and vulnerability of people, properties and/or other assets to the hazard. However, these models are subject to significant uncertainty. Knowing how much uncertainty exists in risk estimates and how it can be attributed to its various sources is essential to guide efforts for model improvement, as well as to help risk managers make better decisions. Currently, there is a lack of knowledge regarding how much uncertainty affects the inputs of flood risk models and how it can be quantified. Additionally, it remains unclear how such input uncertainties propagate and which uncertain input mostly controls the uncertainty in predicted losses. Flood risk models can now estimate risk at large scale which means they produce large and complex datasets that can be challenging to analyse. The application of Machine Learning methods (like Decision Trees) poses a set of specific challenges.

In this Thesis I intend to improve our understanding of uncertainty quantification and attribution in flood risk modelling. I performed an uncertainty and sensitivity analysis of a flood risk model over a large and heterogeneous domain, the Rhine River Basin, and in doing so developed a new Machine Learning method (interactive Decision Tree) to analyse large datasets by incorporating scientific knowledge into machine learning. I then developed an approach to quantify the uncertainty in the inputs of a flood risk model, based on a systematic literature review. Using this approach, I reported variability ranges for residential buildings value, damage ratios of the vulnerability curves, and the return period of flood events in the Rhine River Basin. I used these input uncertainty ranges to perform a global sensitivity analysis on the flood risk model over the Rhine River Basin. First, I identified the dominant input uncertainties in each spatial unit of the study domain for two key model outputs: the Average Annual Losses (AAL) and the Loss Exceedance Curves (LEC). I found that AAL is dominated by uncertainty in vulnerability, while the dominant uncertainties for the LEC change with the return period, with vulnerability becoming increasingly important with increasing return period. Finally, I attempted at linking dominant input uncertainties with the hydrological and socio-economic characteristics of the spatial units using an interactive Decision Tree. I found that topography, degree of urbanization and residential buildings' value are key characteristics for determining how dominant uncertainties change over space within the study domain.

Acknowledgements

This work was funded as part of the Water Informatics Science and Engineering Centre for Doctoral Training (WISE CDT) under a grant from the Engineering and Physical Sciences Research Council (EPSRC), grant number EP/L016214/1. This work was carried out using the computational facilities of the Advanced Computing Research Centre, University of Bristol - http://www.bristol.ac.uk/acrc/.

I am deeply thankful to my supervisors Francesca Pianosi and Thorsten Wagener, for their guidance, advice, patience, and support throughout my PhD. Thanks to them for teaching me how to explore new ideas and showing me new ways of thinking not only in research but also in life. Working with them was a delightful and fruitful life experience.

Thanks for their understanding and support during one of my most difficult times in life. For that I will always be grateful to them. Thanks to JBA Trust for supporting this studentship and to JBA Risk Management for providing me their model, data and expertise. I am thankful to my industrial supervisor Rob Lamb and collaborators Kirsty Styles and Stephen Hutchings for all their kindness, advice and support, showing interest in my ideas, and offering me the opportunity to think things from the industry perspective. Thanks to Alex, Philip, Antonia, Alison, David, Barbara for making my visits to JBA an enjoyable and fruitful experience. Special thanks to Carl Fischer for his patience and willingness in helping me solve all the technical problems I had with the model.

I am thankful to Bruno Merz, Heidi Kreibich, Sergiy Vorogushyn, Björn Guse for the interesting discussions, their advice and valuable comments on my project and for hosting me in the German Research Centre for Geosciences (GFZ) in Potsdam.

I am thankful to Roberto Quaglia for all the amazing times we spent together during my PhD. Thanks to my WISE and Woodland Road colleagues Sebastian, Lina S., Elisa C., Valentina, Elisa B., Anna, Stamatis, Doris, Simbi, Giulia, Charlie, Juliana, Dan, Mikkel with whom we had great time together. I am thankful to my friends from Greece Dimitris P., Kleopatra, Dimitris B., Katerina, Thanasis, Savvas, Vasilis, Alexia, Sofia, Giorgos who were with me all along this journey.

Most importantly, I would like to thank my parents Konstantinos and Maria and my sister Pinelopi for their unconditional love and support throughout my entire life and for teaching me to look at life with no fear and always struggle for the best.

Dedicated in the memory of my father Konstantinos Sarailidis,
who always encouraged me to chase new horizons.

“Περί ουδέν των ανθρωπίνων έργων υπάρχει βεβαιότης.”

Αριστοτέλης

“There is no certainty about any human works.”

Aristotle

Author’s Declaration

I declare that the work in this dissertation was carried out in accordance with the requirements of the University's *Regulations and Code of Practice for Research Degree Programmes* and that it has not been submitted for any other academic award. Except where indicated by specific reference in the text, the work is the candidate's own work. Work done in collaboration with, or with the assistance of, others, is indicated as such. Any views expressed in the dissertation are those of the author.

SIGNED: ..... DATE: .....

Table of Contents

List of Figures

List of Tables

Chapter 1: Introduction

1.1 Economic impact of floods

Floods are extreme hydrological and meteorological hazards and among the deadliest and costliest natural hazards. They cause over 40% of global deaths from natural disasters and over 50% of economic damages (left panel in Figure 1.1). The economic and societal impacts of recent flood events are shown on the right of Figure 1.1. In July 2021 Germany, Belgium, and the Netherlands were hit by a catastrophic flood event that led to 184 fatalities and the estimated damages for Germany alone are in the order of 30 billion euros (Apel et al., 2022). The 2022 Pakistan floods led to 1200 casualties and displaced around 33 million people (Mallapaty, 2022). Total damages are expected to exceed USD 14.9 billion, and total economic losses to reach about USD 15.2 billion (Altaf, 2022). The Thailand flood in 2011 had devastating impacts on the world economy: industrial production was reduced by 2.5%, the growth rate in 2011 was reduced from 4.1% to 2.9%, while the damages to the insurance industry were assessed at 10 billion US dollars (Haraguchi & Lall, 2015). Additionally, flood events can cause significant problems in businesses. In the insurance industry, the protection gap (the difference between the total and insured losses) is increasing. Over the last 10 years, the global losses due to floods have increased while the insured share of these losses declined from 21% to 15% (Franco et al., 2020, Zanardo & Salinas, 2022).

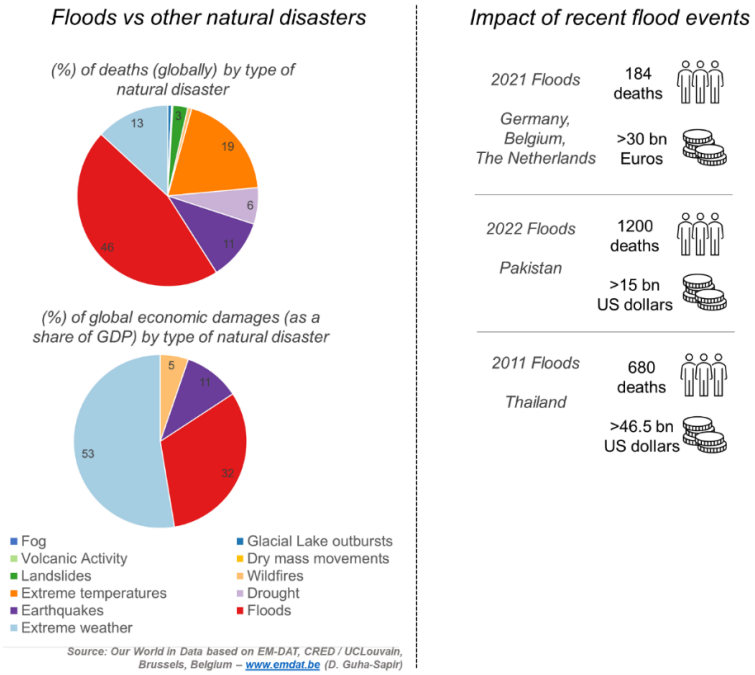


Figure 1. 1 *Statistics of flood impacts (deaths and economic losses) worldwide and for recent flood events.* ***Left:*** *Flood damages and deaths at global scale in comparison with other natural disaster (Ritchie et al., 2023).* ***Right:*** *Flood damages and deaths of recent flood events.*

Hence, national and international efforts are focused on better managing flood risk and mitigating the impacts of floods. For example, the United Nations Office for Disaster Risk Reduction has published the Hyogo Framework (United Nations, 2005) and the Sendai Framework (United Nations, 2015) that set quantitative targets for risk reduction in terms of lives, economic and environmental losses. Among the priorities is to ensure national and global cooperation on risk reduction. Another example is the European Union Flood Directive 2007/60/EC (European Commission, 2007), which requires all the EU member states to perform flood risk assessments, find areas at risk of flooding, derive flood maps and take actions to reduce flood risk (Kundzewicz et al., 2018).

1.2 Flood risk models

Flood risk assessments are performed using models that estimate flood risk as the product of the probability of occurrence of an event (hazard) with its footprint (exposure) and consequences on society and economy (vulnerability) (Duha Metin et al., 2018, Beven et al., 2018). A typical modelling cascade is shown in Figure 1.2. It starts with the estimation of peak river discharges for a set of flood events based on observed or simulated meteorological data (rainfall, river discharge etc). The peak river discharges are then fed into a flood inundation model to simulate the flood depths and extents over the region of interest. Once the flood depths and extents are available, they are intersected with exposure datasets to capture the exposed elements. Information on the damage driving characteristics of the flood event (e.g. flood depth) and on exposures (e.g. economic value of exposed assets) are then combined using vulnerability curves to assess the impact, usually expressed in terms of flood losses.

Nowadays, numerous Flood Risk Models (FRMs) are available to simulate flood risk at various scales, from catchment to regional or even global scales (Ward et al., 2015, Trigg et al., 2016, Kaczmarska et al., 2018). Traditionally, flood risk assessments were performed at catchment, regional or national scales (Hall et al., 2005, European Commission, 2007, Jonkman et al., 2008). This approach is useful in wealthy and scientifically advanced countries where models and good-quality data are available. But this is not the case everywhere and many parts of the world are still ‘unmapped’.

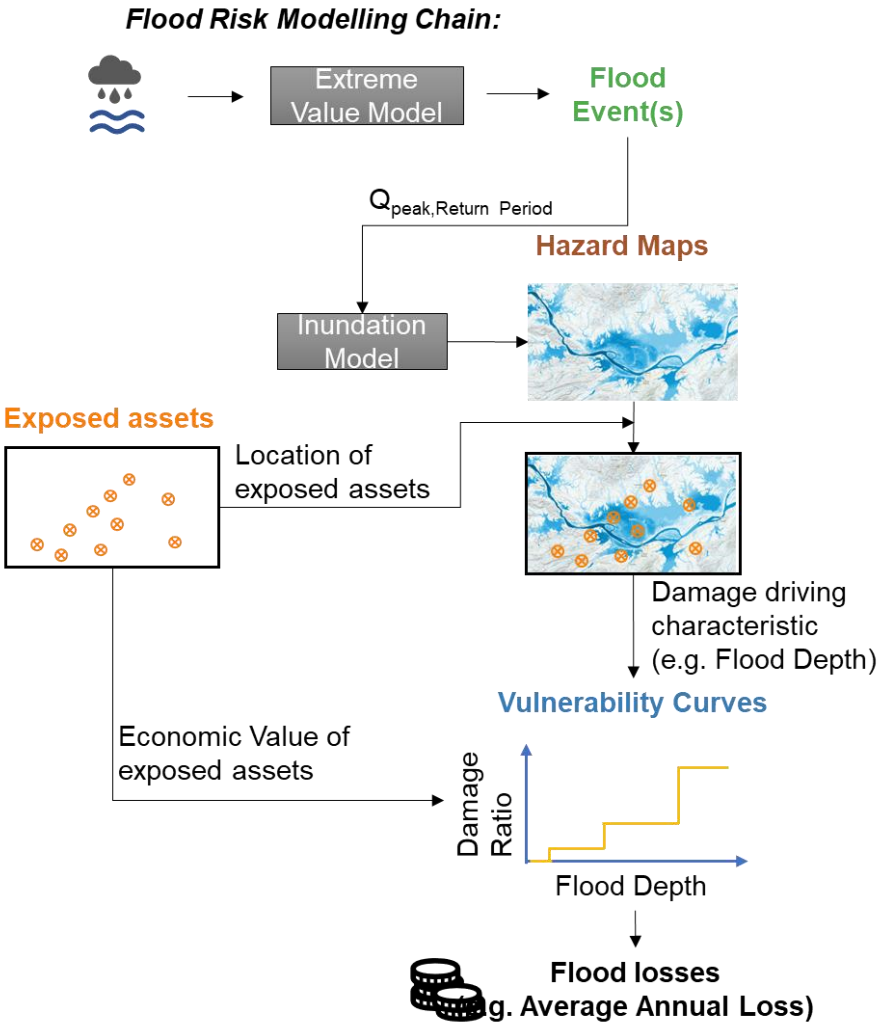


Figure 1. 2 *A typical flood risk modelling chain.*

With the increase in computational resources (Washington et al., 2009) and the growing availability of datasets at global scale (Nativi et al., 2015), large-scale FRMs have been developed to provide a consistent view of flood risk at increasingly larger scales (regional, national, continental and global), including areas where flood risk information was previously unavailable. Global investors use these risk estimates to prioritize their investments or find locations for further development, whereas international risk reduction organisations and governmental agencies use them to inform flood risk prevention and management strategies. For example, the World Bank has used large-scale FRMs to inform national-level disaster risk management plans for strengthening Nigeria's flood resilience by generating flood risk maps per state to identify hotspots of flood risk (Ward et al., 2015). The Government of Belize used the nationally consistent information of a large-scale FRM to support decision-making in spatial and infrastructure planning (Ward et al., 2015). The European Joint Research centre used a large-scale FRM to develop pan-European flood risk estimates (Hall, 2014). (Re)Insurance industry uses large-scale FRMs – or flood

"catastrophe models" in this sector's terminology – for risk pricing, accumulation control, portfolio and capital requirements management and solvency calculations (Grossi & Kunreuther, 2005; Mitchell-Wallace et al., 2017). The number of flood catastrophe models has significantly increased in the last few years (Zanardo & Salinas, 2022). Franco et al. (2020) showed that the number of catastrophe models nearly doubled from 24 to 44 between 2015 and 2020, compared to the previous 5 years.

1.3 Sources of uncertainty in flood risk models

In both research and industry, there is an increasing awareness of the importance of understanding and estimating uncertainty in flood risk assessments (Merz & Thieken, 2005, Merz & Thieken, 2009, Kaczmarska et al., 2018, Beven et al., 2018). This is due to the recognition that every model has its limitations, the difficulty in capturing the natural variability in the characteristics of flood hazard and the socio-economic processes that determine flood losses (Beven et al., 2018) and the recognition that flood risk assessments without considering uncertainty may lead to qualitatively and quantitatively disparate answers.

FRMs are subject to numerous input uncertainties which propagate through the flood risk modelling chain and make the final flood loss estimates uncertain. Sources of uncertainty in FRMs include:

- Uncertainty in processes representation. This type of uncertainty could refer to the models (e.g. statistical, physical) used to represent the physical processes, omissions (if a process is considered negligible relative to the scale) and simplifying assumptions. For example, to estimate the return period of flood events, extreme value models are fitted to observed (or simulated) data. Even with a satisfactory fitting, different probability distribution functions can lead to different estimated return periods for peak discharges (especially when extrapolating for return period beyond the data record size) (Merz & Thieken, 2005).
- The errors, gaps, and inconsistencies of the data used to build and test the models. Examples in the case of FRMs are the inaccuracies in the location, number, value and type of the exposed assets and/or the flood damage data used to build the vulnerability curves (Merz et al., 2010, Zanardo & Salinas, 2021), or the inaccuracies of the Digital Elevation Models (DEM) datasets used for flood inundation modelling (Schuman et al., 2014, Ward et al., 2015).
- Uncertainty in model parameters. An example could be the buffer area around an exposure point that represents the footprint area of the building (Winter et al., 2018).

- Set up choices that are necessary for the execution of the model on a computer (Wagener & Pianosi, 2019). Examples are the spatial resolution at which the losses will be aggregated (national, district or post-code level), or of the hazard maps to be used (Savage et al., 2016).

1.4 Uncertainty quantification and attribution

Given the complexity and non-linearity of FRMs, the impacts and relative importance of the different uncertainty sources on the final model output uncertainty is not immediately obvious. Specific Uncertainty and Sensitivity analysis techniques can be used to analyse the propagation of uncertainty through the model, quantify the resulting uncertainty in model output and attribute it to key contributing input uncertainties.

Differently from local sensitivity analysis, where uncertain inputs are varied one at the time starting from a baseline value, in this Thesis I will use a global sensitivity analysis (GSA) approach, where all input uncertainties are varied simultaneously (Pianosi et al., 2016, Wagener & Pianosi, 2019) within their variability space. GSA is a sensible choice here given that the complex nature of FRMs makes it difficult to define an appropriate baseline to which all the input uncertainties should refer to.

GSA is performed in four steps which are presented in Figure 1.3.

1. Firstly, one needs to identify the input uncertainties and characterize their variability space (e.g. assign them with probability distribution functions or variability ranges).
2. Then, randomly sample combinations of input uncertainties values from the input variability space.
3. Run the model against each input combination to propagate the input uncertainties to the model output. These output samples can be used to quantify the output uncertainty.
4. Finally, post-process the input/output samples to derive sensitivity indices. A sensitivity index typically varies between 0 and 1 and measures the amount of output uncertainty caused by input uncertainty X_i. The higher the value the more sensitive is the output to X_i.

Global sensitivity indices can be used to rank input uncertainties based on their relative contribution to the output uncertainty (Saltelli et al., 2008). This is useful in flood risk modelling to help experts in various tasks, e.g., to prioritize their efforts in uncertainty reduction (Pappenberger et al., 2008). For example, they can spend more time and

resources in better defining the vulnerability curve(s) in places where they know this is the dominant input uncertainty. Another use is for model diagnostic evaluation, where the modeller can check whether the model behaves consistently with their perceptual models, e.g., dominant input uncertainties at a particular place correspond to the physical processes that are also expected to dominate in reality, (Wagener et al., 2022).

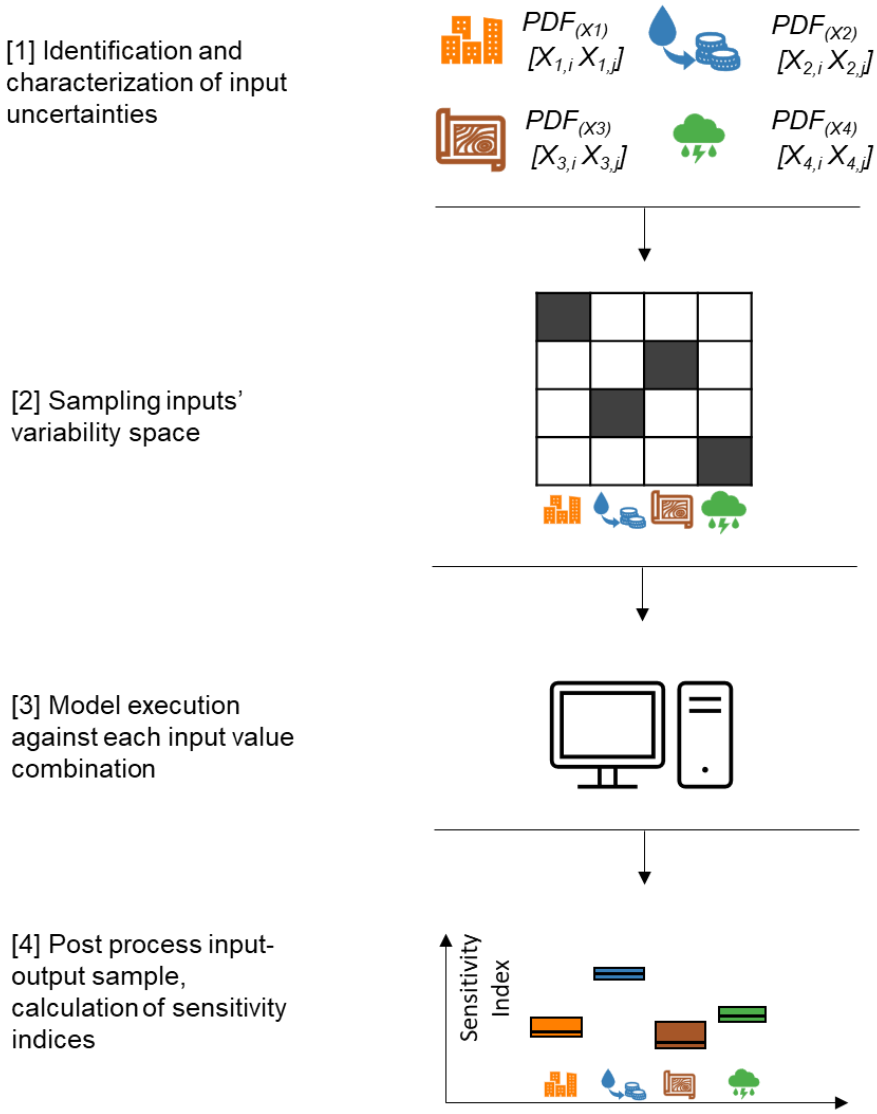


Figure 1.3 *Basic Steps of Global Sensitivity Analysis*

1.5 Challenges in uncertainty quantification and attribution of flood risk models

1.5.1 The challenge of a realistic definition of the input variability space.

The definition of the input variability space (Step 1 in Figure 1.3) is a challenging and crucial step of uncertainty and sensitivity analysis. It requires the modellers to make choices on which values are possible and which are not, and (possibly) their associated probabilities.

These (more or less) subjective choices on the characterization of the input uncertainties propagate through the model and can significantly affect the sensitivity analysis results (Savage et al., 2016, Paleari et al., 2016).

The input variability space needs to be a realistic representation of the variability existing in each input uncertainty. But, in flood risk modelling, this is challenging. Experts often deal with input uncertainties which probabilities are poorly known and constrained. For example, it is known there is considerable uncertainty in flood depths estimation, but it is unknown how much they vary with space and/or return period. Additionally, there is little (or no) availability of datasets that could be used to constrain the input uncertainties. For example, flood damage data collected from historical flood events could be used to constrain the uncertainty around vulnerability curves. While such datasets are available in data rich countries like UK, US and the Netherlands, they are lacking in most other countries (Wing et al., 2020, Bernhoven et al., 2022).

1.5.2 Conflicting evidence on the importance of input uncertainties.

Global Sensitivity Analysis has been previously applied to FRMs at city or catchment scale, providing conflicting evidence on the relative importance of input uncertainties. For example, De Moel & Aerts (2011) in a case study in Netherlands found that uncertainties in vulnerability component are dominating the uncertainty in flood loss estimates. In contrast, Apel et al. (2004) in a case study in Cologne found the extreme value modelling uncertainties to be dominant. This might be explained by the fact that they used different methods and data, or by the fact that the analysis was performed in different places. Indeed, as the hydro-climatic and socio-economic processes represented in flood risk assessments vary in space, one may expect dominant input uncertainties to also vary in space.

It is known from studies in other fields that when the same model is applied across different places, the relative importance of input uncertainties varies from place to place. For example, van Werkhoven (2008) used GSA to investigate the behaviour of a hydrological model across diverse watersheds and time periods and found that the patterns of dominant input uncertainties vary consistently with the variation in the hydroclimatic characteristics of the watersheds. Brewer et al. (2017) reached to similar conclusions regarding the sensitivity indices for a computer model simulating chemical transport in the atmosphere.

1.5.3 Suitability of Machine Learning methods to analyse geoscience datasets.

To investigate the dominant input uncertainties over a large domain and potentially link them to hydro-climatic and socio-economic characteristics of places, it is necessary to process

and analyse large and complex datasets. Machine Learning (ML) methods are promising tools that can support this effort as they are in principle capable of identifying relationships in large data sets. ML methods have been previously used to process datasets in various geosciences applications, such as seismic hazard assessment (Perol et al., 2018), volcanology (Shoji et al., 2018), hydrology and water resources (Sit et al., 2020, Zounemat-Kermani et al., 2021), and flood risk assessment (Chen et al., 2021, Paulik et al., 2022, Bentivoglio et al., 2022).

However, as ML methods originated from a different field, they may not consider the specific data characteristics and issues relevant for applications to the geosciences (Faghmous & Kumar, 2014, Bergen et al., 2019, Reichstein et al., 2019). Users of ML in geosciences have knowledge of the physical principles and laws, that could be exploited to guide the ML process. Doing this would potentially lead to ML models that are more efficient and ensure that the data analysis results are more interpretable and consistent with our knowledge. Indeed, this is an active research topic in many fields which face similar challenges to geosciences like medical sciences (Gibert et al., 2010, Holzinger 2016), visual analytics (Choo & Liu, 2018) and finances (Sinha & Zhao, 2008).

1.6 Research questions and objectives

This Thesis aims at improving our understanding of uncertainty quantification and attribution of large-scale Flood Risk Models over heterogeneous regions.

This Thesis is motivated by the following research questions:

1. How do we characterise the variability space of highly uncertain inputs of Flood Risk Models?
2. How does the relative importance of input uncertainties to Flood Risk Models change in places with different hydrological, climatic, land-cover and economic characteristics?
3. How can we incorporate scientific knowledge into machine learning methods to improve the analysis of large geophysical datasets?

In answering the above questions, I pursue the following key objectives:

1) Develop a method that combines machine learning with scientific knowledge to analyse large geophysical datasets.
2) Develop an appropriate and transferable methodological approach to quantify the variability space of input uncertainties of flood risk models (i.e. uncertainties in the

hazard, exposure and vulnerability components) when there is little knowledge or data to do so.

3) Perform a global sensitivity analysis of a flood risk model at large scale to quantify the relative importance of different input uncertainties.
4) Investigate how dominant input uncertainties vary spatially by linking them to climatic, hydrological, land cover and socio-economic characteristics.

To fulfil the above objectives, I have chosen the Rhine River basin because it presents sufficiently large heterogeneity in the hydrological, climatic, land cover and socio-economic characteristics. Moreover, I use the data and global flood risk (catastrophe) model provided by JBA risk, a standard industry model that can simulate flood risk at such large scales.

1.7 Thesis outline

In Chapter 2 I discuss challenges geoscientists face in the application of ML methods in the field and present a novel method for addressing the first research question. That is, I present an interactive framework that put humans in the development loop of ML method. I implement this framework by developing interactive Decision Trees (iDT). With the iDT experts can interact with the DT and make amendments based on their domain/scientific knowledge. I demonstrate with three case studies how iDT overcome problems of current DT algorithms, leading to models with higher interpretability and robustness.

In Chapter 3 I present a structured literature review approach to define plausible variability ranges for input uncertainties. The approach is applied on the Rhine River basin but it is transferable to other regions and similar applications.

Last, in Chapter 4 I perform a global sensitivity analysis of the JBA flood risk model over the heterogeneous Rhine River basin, using the variability ranges from Chapter 2. Once I have identified the dominant input uncertainties over the entire basin, I investigate whether their variability over space can be linked to different characteristics of places within the basin. For example, in places with high urban coverage and high economic value, I may expect flood losses uncertainty to be dominated by the value of the buildings. For this purpose, I use the Interactive Decision Trees method developed in Chapter 2.

Chapter 2: Integrating scientific knowledge into Machine Learning using interactive decision trees

This chapter has been published as a research article in Computers & Geosciences. Slight modifications have been made to better fit the general layout of this Thesis.

Citation: Sarailidis G., Wagener, T., Pianosi, F. (2022) Integrating scientific knowledge into machine learning using interactive decision trees. *Computers & Geosciences*, *170*, 105248. https://doi.org/10.1016/j.cageo.2022.105248.

2.1 Introduction

In flood risk and in geosciences in general we often need to analyse large and complex datasets. In the past few decades, our ability to collect, store and access large volumes of earth systems data has increased at unprecedent rates thanks to improved monitoring and sensing techniques (Hart & Martinez, 2006; Butler, 2007; Karpatne et al., 2017; Zhou et al., 2017), ever growing computational power (Washington et al., 2009), and the development of simulation models that produce large datasets at increasing domain scale and resolution. An example is the CMIP-5 dataset of the Climate Model Intercomparison Project (which contains various climatological variables at daily resolution (1980-2300) with global coverage and over 3 petabytes in size) that has been used extensively for scientific groundwork towards climate assessments (Reichstein et al., 2019). This 'data deluge' has paved the way for the systematic processing and analysis of observational and simulation data, often using Machine Learning or other statistical methods (Reichstein et al., 2019; Karpatne et al., 2019; Sun et al., 2022).

Machine Learning (ML), a term defined by Samuel (1959), is a branch of artificial intelligence (AI) and computer science which focuses on discovering patterns hidden in complex datasets (Bzdok et al., 2017; Reichstein et al., 2019) by imitating the way that humans learn (IBM, 2020). The main purpose of ML is to develop algorithms that can learn from historical data and perform tasks (e.g. predictions and classification) on new input data. The capability of ML methods to automatically extract patterns from large volumes of complex and high-dimensional data have made them an important part of research in many fields, including the geosciences (Bergen et al., 2019, Sun et al., 2022).

In this chapter I focus on a method called Decision Tree (DT) (Breiman et al., 1984), a supervised ML method that is widely used in the geosciences. A DT model is developed

through an automatic algorithm that recursively partitions the space of input variables into subspaces using a set of hierarchical decisions. In Figure 2.1, I show a DT with a schematic representation of the recursive partitioning of the dataset along with basic terms used in this chapter. A DT model is a hierarchical tree structure that comprises nodes and branches. Each node is associated with a logical expression, i.e. a “split”, which consists of the variable and threshold to split, e.g. “Xi smaller $\bar{X}i,j$”. Each node will lead to two branches that correspond to the different possible outcomes of the split. The terminal nodes are called leaves and are associated to either a class or a specific value for the output. DT are thus commonly used for (Flach, 2012):

- Classification: The DT is trained on output data that are categorized under different classes (discrete values or non-numerical categories) and can predict classes for unseen data.
- Regression: The DT is trained on continuous output variables, and it predicts continuous values instead of classes.

Examples of DT applications in the geosciences include, catchment classification (Sawicz et al., 2014; Kuentz et al., 2017), land cover classification (Gislason et al., 2006), studying uncertain factors of simulation models (Almeida et al., 2017; Sarazin, 2018), analyzing rainfall-runoff relationships (Iorgulescu & Beven, 2004; Singh et al., 2014), empirical streamflow simulation (Shortridge et al., 2016), soil mapping (Grimm et al., 2008; Hengl et al., 2017), regionalizing hydrological signatures (Addor et al., 2018), flood risk management (Chen et al., 2021, Paulik et al., 2022).

DTs are quite appealing in the geosciences because geophysical processes often reveal hierarchical structures of controlling variables, and the hierarchical structure of DT with nodes, branches and splits is a straightforward way to capture this. In geoscience applications, DT are particularly appealing for the purpose of organizing spatially distributed entities, such as rivers, catchments or other landscape units, thus demonstrating how large-scale (e.g. climatic) controls interact with small-scale (e.g. land use or geology) controls (Sawicz et al., 2014; Addor et al., 2018).

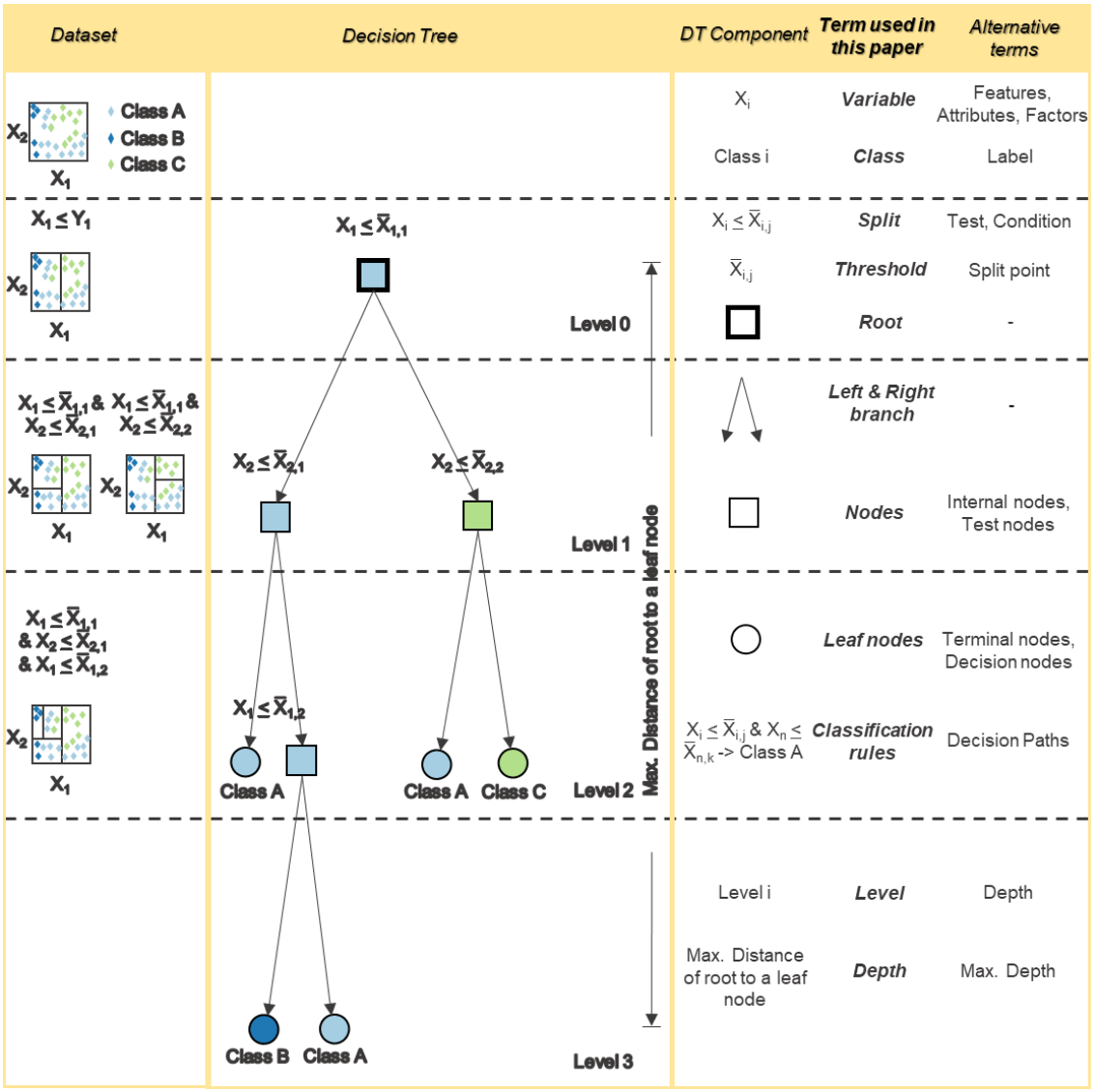


Figure 2. 1 *Schematic representation of a DT and the recursive partitioning of the dataset along with basic terms.* ***Left:*** *A schematic representation of the recursive partitioning of the data space performed by a Decision Tree development algorithm.* ***Middle:*** *A typical Decision Tree.* ***Right:*** *Terminology.*

Despite these advantages, they have limitations (see Figure 2.2) which make their use in the geosciences challenging. I highlight three main challenges that are important to the discussion:

1. Like any statistical tool, DT methods rely on data and consequently their credibility is dependent on the quantity and quality of data available. DT require large amounts of data for training which are not always available (Kirchner et al., 2020). When available, data in geosciences can be biased, complex, uncertain, noisy, heterogeneous and with changing properties (e.g., due to changes in measurement instruments or the data processing algorithms) (Solomatine & Ostfeld, 2008; Faghmous & Kumar, 2014; Beven et al., 2018; Karpatne et al., 2019). Therefore, the

accuracy of DTs deteriorates with decreasing size or quality of the training dataset (Pal & Mather, 2003).

2. DT development relies on statistical metrics and algorithmic decisions aim at statistical optimality, usually measured in terms of classification rate or regression accuracy. However, such statistical optimality does not guarantee that the outcome is physically consistent (Roscher et al., 2020). By physical consistency I mean that a DT should not violate scientific principles (such as conservation of mass) or overlook known physical characteristics of the system investigated. For example, some input variables may have physically meaningful threshold values that may be missed by the DT because other threshold values might produce a statistically better result for the (noisy and biased) dataset used for training. Moreover, most DT algorithms use split rules based on a single variable at each node, whereas combinations of multiple variables may play a significant role in partitioning the data space (Loh, 2014; Almeida et al., 2017).
3. DT complexity may decrease their interpretability and consequently limit their usefulness in geosciences applications. By interpretability I mean the ability by a human expert of making sense of the obtained model (Molnar, 2019), understand how the model works and reaches a specific decision. DT are easier to interpret if they are small. The greater the number of terminal nodes, the deeper the tree and the more difficult it becomes to interpret. (Molnar, 2019; Lipton, 2018). Visualization could also help increase the interpretability of DTs. However, existing visualization techniques mainly focus on displaying information related to the statistical properties of the DT (e.g. impurity, node data points), whereas they do not support the display of information related to the physical properties of the variables – something that would potentially be more useful for geosciences applications (Almeida et al., 2017).

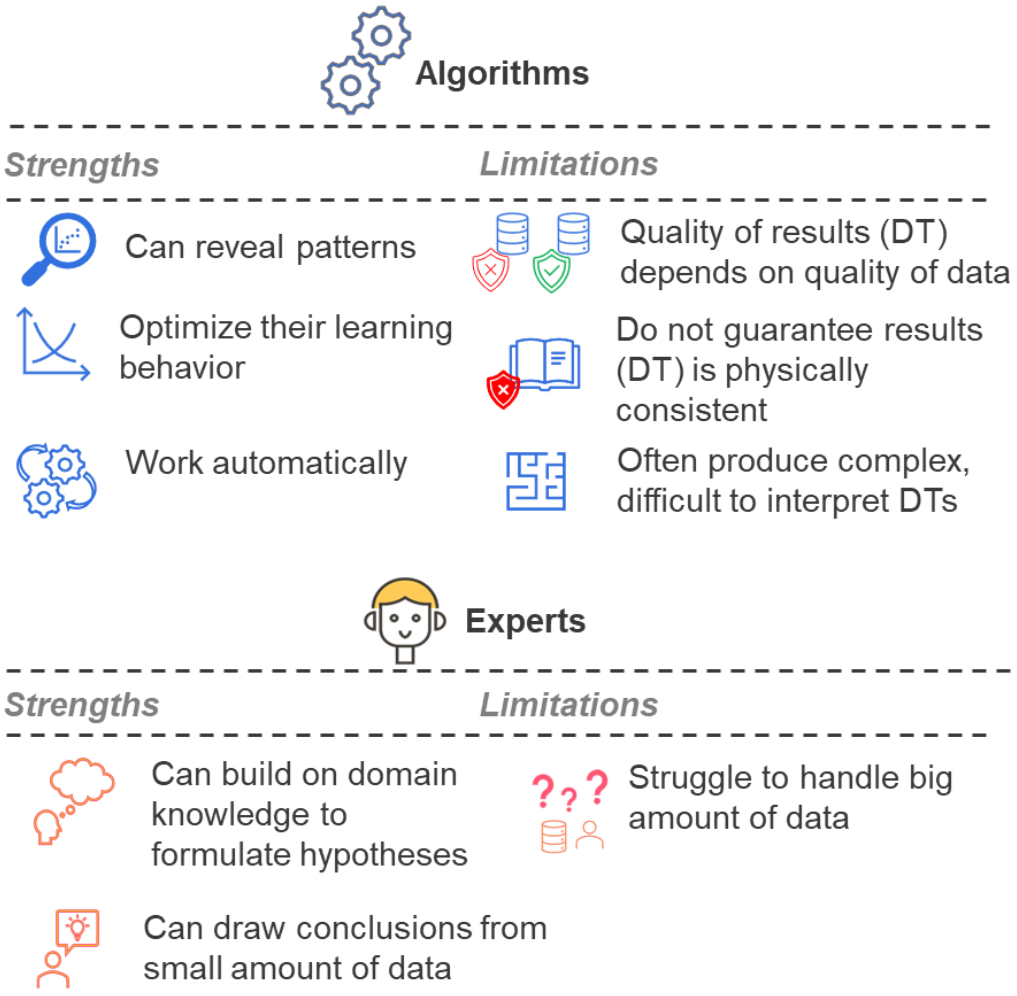


Figure 2. 2 *Strengths and Limitations of Decision Trees Algorithms and experts.*

Integration of human experts in the DT development process – and hence of their domain knowledge and their cognitive ability to formulate hypotheses and theories – may help overcome some of these challenges. For example, experts may have very good knowledge of the physical processes, quantities and phenomena under study and hence be able to define physically meaningful splitting variables and thresholds, or discard DT branches that are physically unrealistic. An example is given in Stein et al. (2020) where a DT model to classify river flood generating processes is built purely based on domain knowledge. In addition, experts can define combinations of input variables that they believe interact in controlling outputs, where current algorithms would not allow for the detection of such combinations. An example is given in Almeida et al. (2017), where expert knowledge enabled integrating multiple variables into a new and physically meaningful factor. Moreover, experts can learn patterns from few datapoints because they have a certain expectation of relevant causal relationships, so they could guide the algorithm to learn from smaller amounts of data, or dataset where a particular output class is under-represented (“imbalanced dataset”) (García & Herrera, 2009). Inclusion of domain knowledge in the

model building process can also increase trust in the modelling results (Solomatine & Ostfeld, 2008). Incorporating scientific knowledge into ML models to improve their physical realism and interpretability has been highlighted as a major challenge and opportunity for ML applications in the geosciences (Read et al., 2019; Sun et al., 2022).

In this chapter, I propose a framework to develop "interactive Decision Trees" (iDTs) that put human experts in the development loop of Decision Trees. The iDT framework establish a two-way interaction between the automatic DT development algorithm and the expert, allowing the expert to manually create new composite variables, changing nodes' splitting variables and thresholds, manually pruning leaf nodes, and visualizing DTs in physically meaningful ways. Past attempts at developing iDT include the works of Ankerst et al. (2000), Han & Cercone (2001), Teoh & Ma (2003), Fails & Olsen (2003), Solomatine & Siek (2004), Mickens et al. (2007), Thanh-Nghi Do (2006), van den Elzen & van Wijk, (2011), Estivill-Castro et al., (2020), Elia et al. (2021). Outside the scientific literature I found two commercial software products that allow users to interact with the DT, Dataiku[1] and IBM SPSS[2]. I discuss briefly why the additional work presented here is warranted despite these previous efforts:

1. To my understanding, the above tools were developed for general use and none of them was tested for geosciences applications. This puts into question whether their interactive functionalities will be applicable and/or useful to overcome the specific challenges discussed above. For example, I ran the tool developed by van den Elzen & van Wijk (2011), but it could not handle the large datasets typical for geoscience applications. The tool by Solomatine & Siek (2004), which was tested on six hydrological datasets, allows for larger datasets, but it is not publicly available.
2. To the best of my knowledge, none of the studies cited above publicly shared the code to run their analyses and this might be one reason why they were not followed up by others or adopted by researchers in our community. The exceptions are: (a) the web application from Elia et al. (2021), which is freely available open source, but is designed for educational purposes; (b) the commercial software Dataiku, which is freely available for academic purpose but not open-source; and (c) the IBM software, which is neither free nor open-source.
3. Finally, in the above tools the main purpose for integrating human expertise in the DT development process is to improve the algorithm' predictive performance. Here instead I argue that interpretability and robustness are at least equally important

[1] https://www.dataiku.com/
[2] https://www.ibm.com/analytics/spss-statistics-software

aspects in geosciences applications. Even to the point that users might accept a reduction in predictive performance if it comes with an increase in interpretability and robustness given new datasets. Hence, I devise and demonstrate a number of visualisation and interaction functionalities that are specifically aimed at increasing DT interpretability, and I also discuss how to measure interpretability (in the context of a specific application – see Case Study 2) beyond simply measuring the DT size (number of layers, number of leaf nodes) and training time (as done in previous studies).

Alongside presenting the iDT framework, I thus also introduce a free, open-source Python package to implement the iDT framework, which I demonstrate using three case studies representatives of typical challenges encountered in the geosciences. In the first case study, I show how color-coding the tree nodes based on their physical meaning produce a physically meaningful visualization, and how the experts can create new composite variables in the training process to better capture existing interactions in the dataset, thus producing a smaller and more interpretable DT. In the second case study, I show how the expert can manually change the splitting threshold values of the tree nodes based on other sources of knowledge, again to increase the DT interpretability. Finally, in the third case study, I show how experts can manually change nodes' splitting variables and thresholds to include under-represented classes in imbalanced datasets and make the DT physically consistent, robust and potentially even more accurate on new datasets.

2.2 Methodology

In this section I describe the framework for establishing interactions between the expert and an automatic DT training algorithm to integrate scientific knowledge in DT development. Moreover, I describe the Python package and the Jupyter Lab Graphical User Interface I developed to implement the framework. Finally, I present my ideas on how to evaluate DT predictive and interpretive performance.

2.3 A framework for interactive construction and analysis of decision trees

Figure 2.3 shows the framework for interactive construction and analysis of DTs and compares it to the classical approach of automatic development. In the classical approach, the analyst prepares the dataset to feed to the ML algorithm, specifies the algorithm's tuning parameters, executes it, and obtains the classification/regression model. In the interactive framework, the analyst (expert) can input their prior knowledge and/or feedback. Specifically, the expert can:

1. Organize and (pre-)process the input datasets, by grouping input variables in a physically meaningful way (such as climate variables, land surface properties, soil properties, etc.). The tree can then be colour coded based on this grouping. The user can also add new composite variables to the input dataset before or after the first algorithm run.
2. Directly manipulate the structure of the DT model, by changing node variables and split threshold values, or by manually pruning the DT or changing leaf node class. This can be useful when the expert is aware of physically meaningful threshold values for certain variables (for example thresholds for climate variables that are used to classify different climate zones) to improve the DT's physical interpretability. Another reason to manipulate the DT structure is the case of an imbalanced dataset, where a certain class is under-represented in the dataset and thus an automatic algorithm may not separately represent that class in the DT. Different tactics have been proposed to overcome this problem, such as resampling (García & Herrera, 2009), synthetic generation (Chawla et al., 2002) or penalized models, although they often are time consuming to implement (Zhou et al., 2017). iDT offers an easier way to overcome the problem by allowing the expert to force the tree to include the under-represented class by manually changing nodes' variable and thresholds to split and/or leaves nodes classes.

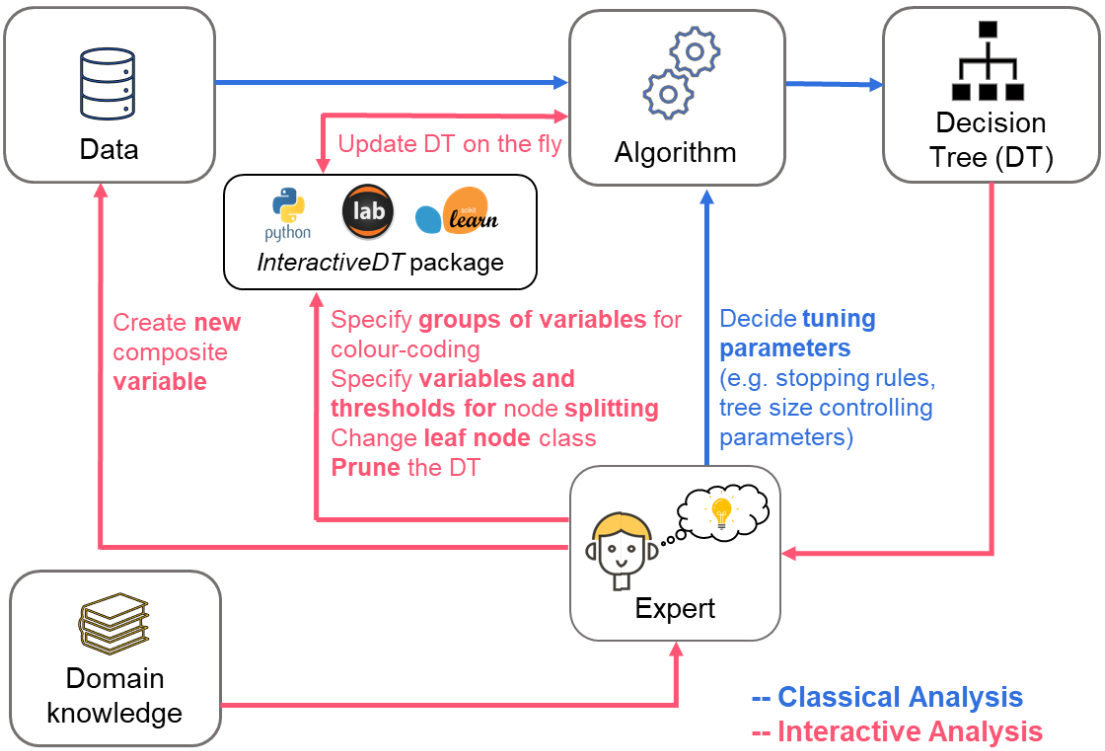


Figure 2. 3 *Classical and Interactive analysis flowcharts. The steps performed to develop a Decision Tree in a "Classical" analysis are coloured in blue and with pink the proposed Interactive analysis.*

2.4 A Python package and graphical user interface in Jupyter Lab for interactive construction and analysis of decision trees

To maximise the reusability, replicability and reproducibility of the proposed approach (Gil et al., 2016; Hutton et al., 2016) I developed and shared an open-source Python package and a GUI in Jupyter Lab for implementing the IDTs framework. The code is available at [3]. I used the sklearn library of scikit-learn package in Python (Pedregosa et al., 2011) that contains the implementation of the tree algorithm (for more details see Appendix A1) as a basis for the interactive tools. I created a new package, called "InteractiveDT", which consists of (1) an "iDT" module containing the functions that enable the expert to interact with the DT or the dataset, and (2) an "iDTGUIfun" module which incorporates these functions into widgets, which are then used in the Jupyter Lab script called "InteractiveDecisionTrees" to create the user interface. Further details about this GUI are also provided in the Appendix A2.

2.5 Evaluating DT predictive and interpretive performance

DT are generally used as predictive tools for classification or regression. Therefore, their evaluation is typically based on statistical metrics of their predictive ability (Lipton, 2018). Examples of such metrics include classification accuracy, confusion matrices, precision, recall, accuracy rate, root mean squared error metrics, and mean error metrics (Pedregosa et al., 2011). However, in geosciences applications, we often would like the DT to be not just a good predictor, but also to be interpretable (Lipton, 2018). In contrast to predictive performance, interpretability is a less well-defined concept and metrics to measure interpretability are not yet well established (Doshi-Velez & Been, 2017). A widely used proxy for interpretability is the complexity of the tree, as it can be reasonably assumed that a less complex tree is easier to interpret (Molnar, 2019; Lipton, 2018). The complexity of a DT can be easily quantified through the number of leaf nodes and/or the depth of the tree (Molnar, 2019). I adopted these simple metrics to evaluate DT interpretability in the first case study.

The need for interpretability is often linked to the use of models to assist scientific understanding (Doshi-Velez & Been 2017). The evaluation of interpretability for scientific understanding though is context specific. In the second case study, I will give an example of a case-specific definition of interpretability, based on the consistency of the DT partitioning with an existing independent classification system of some of the input variables.

[3] https://github.com/Sarailidis/Interactive-Decision-Trees

2.6 Results

2.6.1 Case Study 1 – Color-coding groups of variables and constructing new composite variables to reduce the DT complexity and increase interpretability

The first case study is based on a dataset from a computational landslide study by Almeida et al. (2017), which includes 10,000 combinations of 28 input variables of a slope stability model (the list is given in Table A1 in the Appendix A3.1). These variables are model inputs characterising landscape attributes such as slope geometry, soil and design storm properties and initial hydrological conditions. The model output is the slope factor of safety (FoS), which is typically used to separate the model outputs into two results regarding the stability of a landscape position regarding landslide hazard risk: “stable”, when FoS is above 1, and “failure” otherwise. In Almeida et al. (2017) a standard CART algorithm was used to identify dominant drivers of slope instability. I applied the iDT framework to the same dataset to demonstrate two functionalities of the iDT toolbox: (a) How to increase the visual interpretability of the DT by colour coding variables based on their physical meaning (b) How to better capture interactions between variables by creating new composite variables.

Figure 2.4 shows the statistically optimal DT delivered by the automatic DT algorithm. Nodes are coloured based on Impurity, a default choice in many software packages. Figure 2.4 also shows the graphical interface of the InteractiveDT tool, which enables the user to define groups of input variables and colour code the nodes accordingly. Through this new visualization, it is evident that the first three levels of the tree are dominated by “geophysical properties” and “slope geometry variables”, while levels 4 and 5 are mainly dominated by “design storm properties”. Furthermore, the colour coding helps spotting a repetition of four variables–- cohesion (c_0), thickness of topsoil (H0), rainfall intensity (I) and duration (D)–- at various levels of the tree. Such a repetition suggests that these variables may be interacting with one another to produce slope failures (the tree tries to mimic this interaction). Indeed, a scatterplot of c_0 versus H0 (left hand side in Figure 2.5) shows that combinations of high H0 and low c_o (bottom right) lead to slope failure (black dots). Moreover, it is expected rainfall intensity and duration to interact. Specifically, combinations of high-intensity/short-duration and low-intensity/long-duration rainfall will more likely result in slope failure. This relationship is confirmed in the (log-scale) scatterplot on the right-hand side of Figure 2.5. To capture these interactions, the user may create two composite variables: Soil Ratio, which is the ratio of cohesion and thickness of topsoil (Soil Ratio=c_0/H0), and Storm Ratio, which is the ratio of the logarithms of rainfall intensity and duration (Storm Ratio=-log10(D)/log10(I)). The bottom part of Figure 2.5 shows the new DT delivered by the

algorithm when fed by a training dataset including these two composite variables. Overall, the new DT is “better” than the original one because it is smaller (21 nodes instead of 57, and a depth of 5 layers instead of 8) and hence easier to interpret, and it is more accurate in predicting slope failure (higher number of true slope failures and lower number of false slope stabilities) for both training and test datasets. Evaluation and comparison of the two DTs are summarised in Figure 2.6.

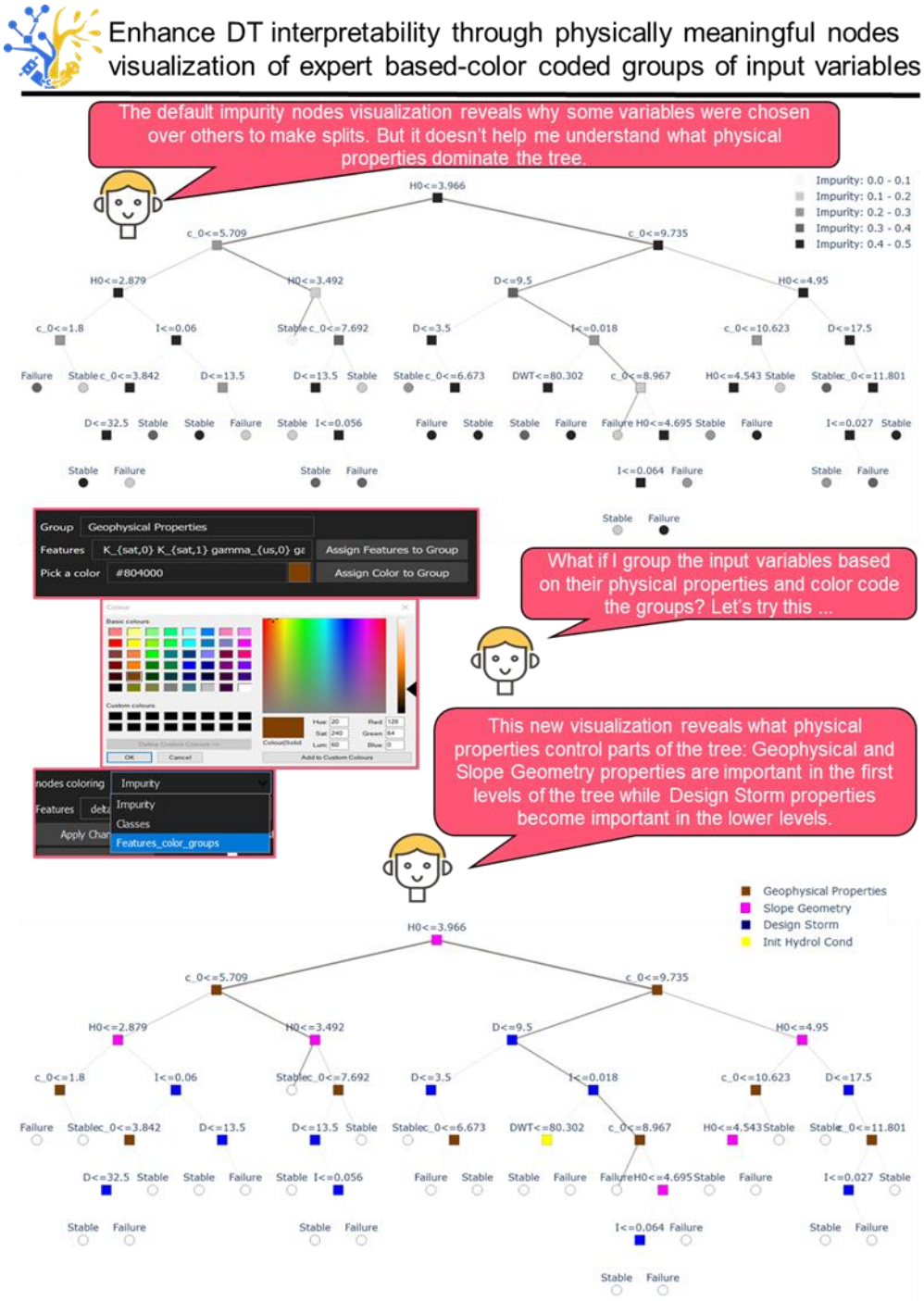


Figure 2. 4 *Showcase of a physically meaningful Decision Tree visualization.* ***Top:*** *Decision Tree for Case study 1 with the conventional nodes coloring approach, which is based on node impurity.* ***Bottom:*** *The same tree is*

shown in the bottom with the proposed alternative nodes coloring, which is based on groups of variables proposed by the user. With this node coloring option, it is evident what kind of variables dominate the tree. The figure also shows a screenshot of the InteractiveDT tool developed to achieve this alternative visualization.

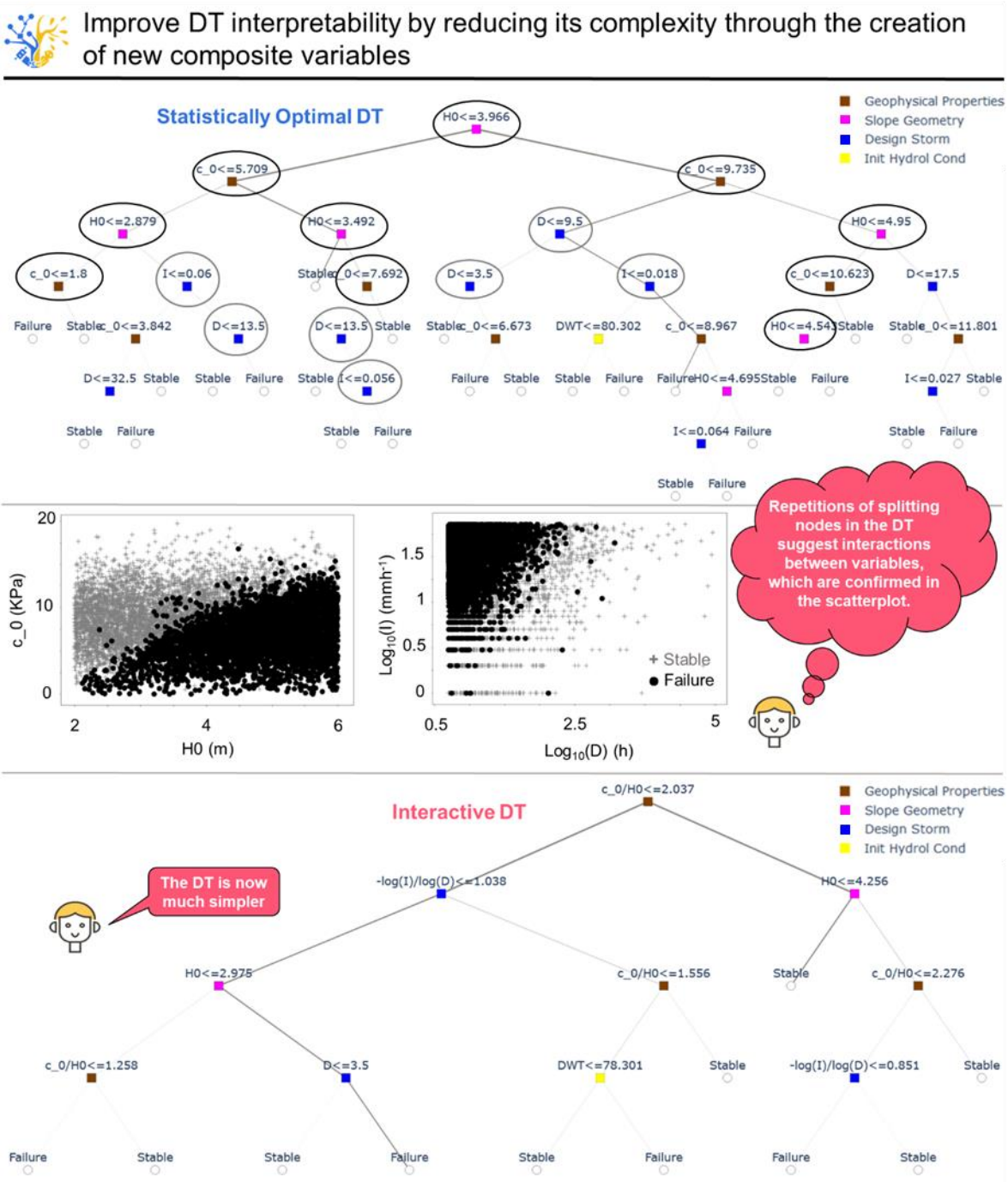


Figure 2. 5 *Showcase the construction of composite variables to reduce Decision Tree complexity and increase interpretability.* ***Top:*** *Initial DT for Case study 1, in which interactions between variables emerged.* ***Middle:*** *Scatter plots of the interacting variables coloured according to whether the associated slope fails (black dots) or not (grey).* ***Bottom:*** *The iDT in the bottom is the new tree after creating two composite variables based on the detected interactions.*

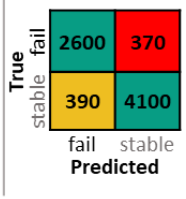
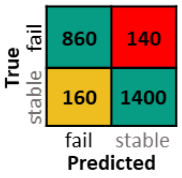
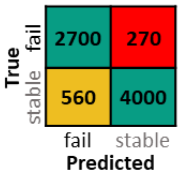
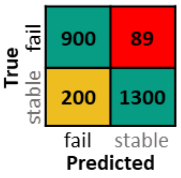
Evaluation criteria	Statistically Optimal DT		Interactive DT	
Classification Performance				
	Train dataset	Test dataset	Train dataset	Test dataset
Classification accuracy	0.898	0.886	0.889	0.885
Confusion matrix				
Interpretability				
Tree Depth (No of levels)	8		5	
Tree size (No of nodes)	57		21	
Variable's **interactions**	Through repetitions		Through composite variables	
Leaf nodes **color-coding**	Based on impurity		Based on physical properties	

Figure 2. 6 *Evaluation of statistically optimal and Interactive DT. Comparison of the statistically optimal and interactive DTs for Case Study 1 based on classification performance (top) and interpretability (bottom).*

2.6.2 Case Study 2 – Increasing interpretability by changing splitting threshold values based on other relevant knowledge sources

The second case study is based on a version of groundwater recharge dataset created by Sarazin (2018) which includes 17,000,000 simulations of 34 input variables of a hydrological model. These variables are model inputs characterizing spatially distributed climate properties, land cover and soil properties of karst landscapes across Europe under current conditions and future climate scenarios. The model outputs are values of annual groundwater recharge, which are then grouped into four classes, namely, C1 (<20 mm/yr), C2 (20 – 100 mm/yr), C3 (100 – 300 mm/yr) and C4 (>300 mm/yr). A DT is built to reveal the key controls of groundwater recharge. To increase the interpretability of the DT I used the iDT framework to manually change some of the nodes' thresholds consistently with a simplified version of Holdridge's life zones classification scheme (Holdridge, 1947). The Holdridge scheme provides a classification of land areas based on annual precipitation and aridity index (i.e. the ratio between potential evaporation and precipitation; Figure A2 in the Appendix A3.2 shows the original and simplified scheme). By imposing that the threshold values for Precipitation (Pm) and Aridity index (AI) in the DT be the same as in the Holdridge chart thresholds, I wanted to explore whether a tree so constructed leads to leaf nodes that map into fewer Holdridge life zones, and as such may be more interpretable, and whether this gain in interpretability comes with a significant loss in classification accuracy.

I generated 15 datasets of 1000 samples each by randomly sampling from the original dataset (of 17,000,000 samples) and hence each dataset is different. I then split each dataset into training and test sets (75% and 25% of the dataset size respectively). For each training dataset I derived a statistically optimal (SO) and an interactive (iDT) decision tree. To derive the SO decision tree, I tried different combinations of the algorithm tuning parameters (splitting criterion based on “Gini impurity” or “entropy”, maximum number of leaf nodes varied from 15 to 25, maximum impurity decreases of 10-5, 10-6, 10-7) and retained the best SO tree based on 10-fold Cross Validation strategy. To derive the corresponding iDT, I used the iDT framework to manually change all the splitting thresholds for Pm and AI to the closest Holdridge chart threshold values. The closest threshold could sometimes be quite far, and the choice of changing it, is subjective. But in some other cases even a small change e.g., changing AI threshold from 1.1 to 1, could make a big difference in terms of interpretability.

Figure 2.7 shows an example of a statistically optimal DT (top) and the corresponding iDT (bottom), focusing on the specific branch where I manually changed the thresholds for Precipitation (Pm) and Aridity index (AI), and hence the resulting leaf nodes. Next to the tree branches, I show the Holdridge life zones (HLZs) that the leaf nodes are mapped into. In the statistically optimal DT (top), the three leaf nodes map into 13 and 5 HLZs respectively. In the iDT (bottom), the number of HLZs is reduced to 4, 2 and 5 after changing one precipitation threshold from 229.82 to 250 and another one from 527.54 to 500 (in line with the HLZ classification). Of particular interest is the leaf nodes labelled C2, which define conditions under which groundwater recharge is low. In the statistically optimal tree, such conditions appear in 13 different climatic zones, while in the iDT they can only appear in two climates: thorn/steppe or thorn woodland. This drastic reduction opens up the opportunity for the expert to find meaningful explanations of why those two particular climatic zones exhibit lower recharge and the implications of this finding. While it is beyond the scope of this chapter to go into such explanation, my argument is that the opportunity to develop it in the first place would have not been present if using the statistically optimal tree. Since class C2 was associated to a variety of different climatic zones.

Figure 2.8 shows the classification and interpretability performance of all 15 statistically optimal DTs (one per each of the 15 datasets sub-samples) and associated iDTs (each obtained from the statistically optimal DT by changing Pm and AI thresholds). Regarding classification performance, the differences are not pronounced, which means the changes made by the expert did not lead to a significant loss of performance. As expected, the statistically optimal DTs always show a slightly higher classification accuracy in the training

sets. Interestingly though, the iDTs outperform the statistically optimal trees in most cases (9 out of 15) in the test sets. Interpretability performance was quantified through the number in climatic zones classes C2 and C3 classes can be mapped to. Overall, the plot shows that the number of HLZs associated to leaf nodes of classes C2 and C3 tends to decrease. In conclusion, this example shows that incorporating other knowledge sources in the DT development by manually changing the splitting thresholds produces iDTs with a clearer link to that knowledge, and hence higher interpretability potential, at no significant loss in classification accuracy. The sample datasets and the resulting training and test datasets used in this case study were different but of the same size. However, I would expect the variation in the classification accuracies between SOT and iDT to decrease with increasing dataset sizes because there is more information for the models to learn the underlying general patterns. I would expect that the conclusions on the interpretability performance should be unaffected by the dataset size because the definition of interpretability is related to the number of climatic zones the leaf nodes are mapped to and so it's not dependent on the dataset size used to construct the DT model.

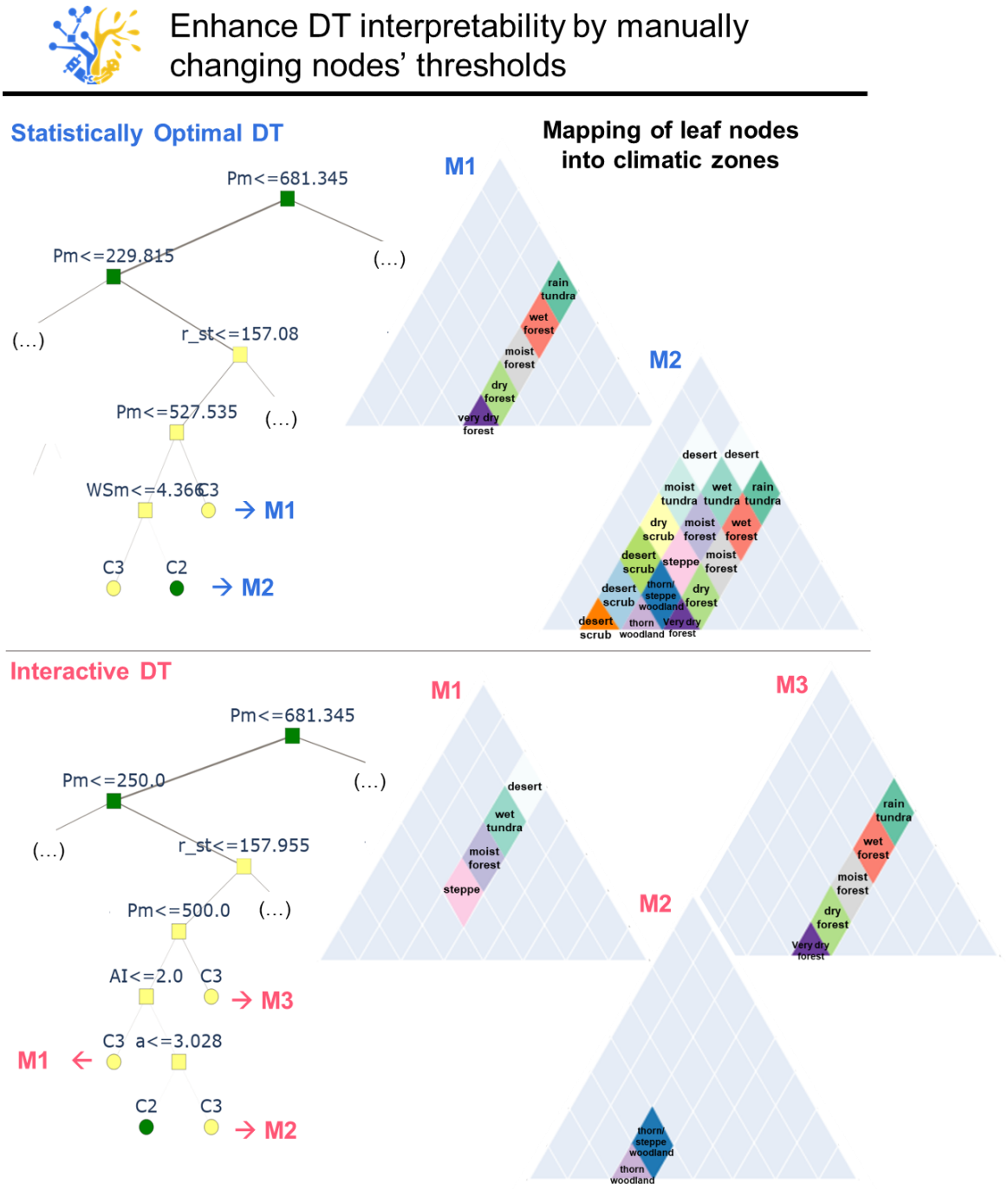


Figure 2. 7 *Comparison of leaf nodes mapping to Holdridge Life Zones for the Statistically optimal and Interactive Decision Trees.* ***Top:*** *Detail of the statistically optimal DT for Case Study 2.* ***Bottom:*** *iDT after manually changing the thresholds for Precipitation (Pm) and Aridity index (AI). For the leaves nodes of each DT I plotted the Holdridge scheme and highlighted the diamonds that the leaves can be mapped to. The variables appearing in the two trees are: Pm: precipitation [mm/y], r_st: Stomatal resistance [s/m], AI: Aridity Index [-], a: Spatial variability coefficient, Wsm: Mean wind speed [m/s].*

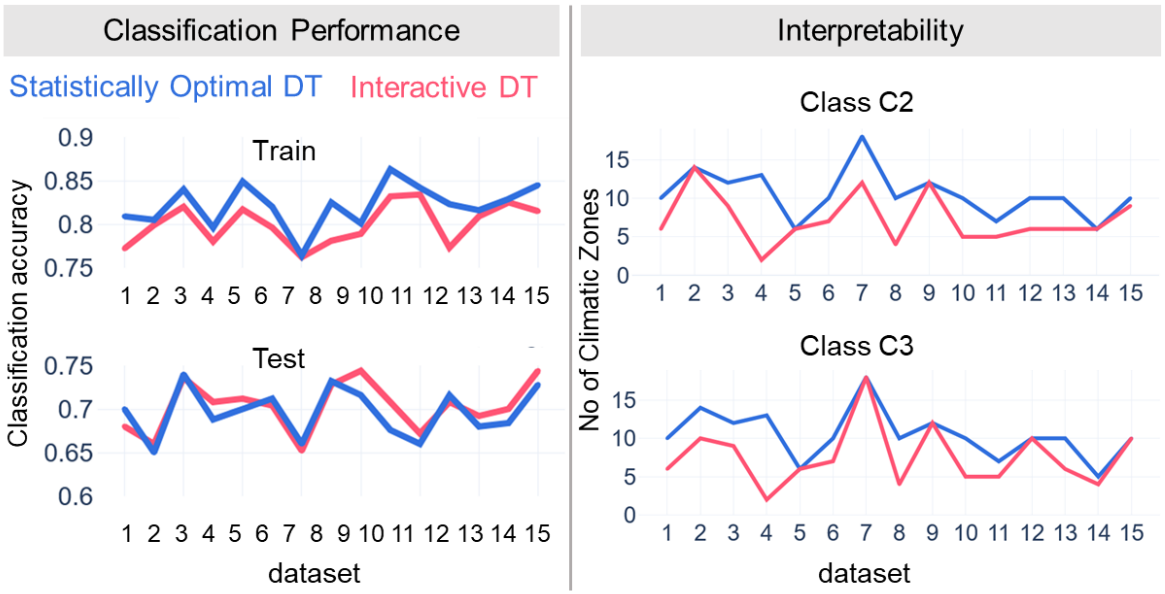


Figure 2. 8 *Evaluation of Statistically Optimal and Interactive Decision Trees. Comparison of the statistically optimal and interactive DTs for Case Study 2 based on their classification performance and interpretability.*

2.6.3 Case Study 3 Manually changing nodes' variables and threshold values to include under-represented classes in imbalanced datasets

This case study is an example of application of iDT in cases where certain classes are under-represented in a dataset, a situation known as "imbalanced datasets". I again used the dataset from Sarazin (2018) as in section 2.6.2, and randomly generated 5 subsample datasets of increasing sizes (1000, 5000, 10000, 50000 and 100000 samples). I then split each subsample dataset into a training and a test set (75% and 25% of the dataset size respectively) and randomly removed data points that belonged to class C2 from the training dataset. Therefore, the training sets contained only few data points of class C2 (<2%). Similarly, to Sec. 3.2, for each dataset I trained a Statistically Optimal (SO) decision tree and then derived an iDT by manually changing the nodes' variables and thresholds until the iDT included the unrepresented class C2 in some of its leaf nodes. In some cases, I also manually changed the class of a leaf node to class C2. For example, in Figure 2.9 on the left I show a part of the SO tree obtained for sample dataset 2. It is known from Sarazin (2018) that low recharge class C2 should appear for low precipitation values, but the algorithm fails to include the C2 class in the SO tree as the class is under-represented in the training dataset. Hence, based on the splitting variables and thresholds of the DT found in Sarazin (2018) I manually changed the threshold in the split node "Pm<=639.075" to "Pm<=300" and the node variable in the split "Vr<=201.14" to "Pm<=65" in the DT, so to create a branch in the tree that specifically explore low precipitation cases. In response to these manual changes, the algorithm created a leaf node for class C2 in the iDT (top right of Fig. 2.9). The change induced a loss of classification accuracy in the training dataset (see Figure 2.9, case

'2') but an increase in performance on the test dataset against unseen data. Moreover, the confusion matrices indicate that iDT is more capable in correctly classifying data points into Class C2 as shown in Figure A.3 in the Appendix A3.3. A similar trend is found for all other datasets: as expected, SO trees perform better in the training sets but iDTs outperform SO trees in in test set, particularly for smaller datasets.

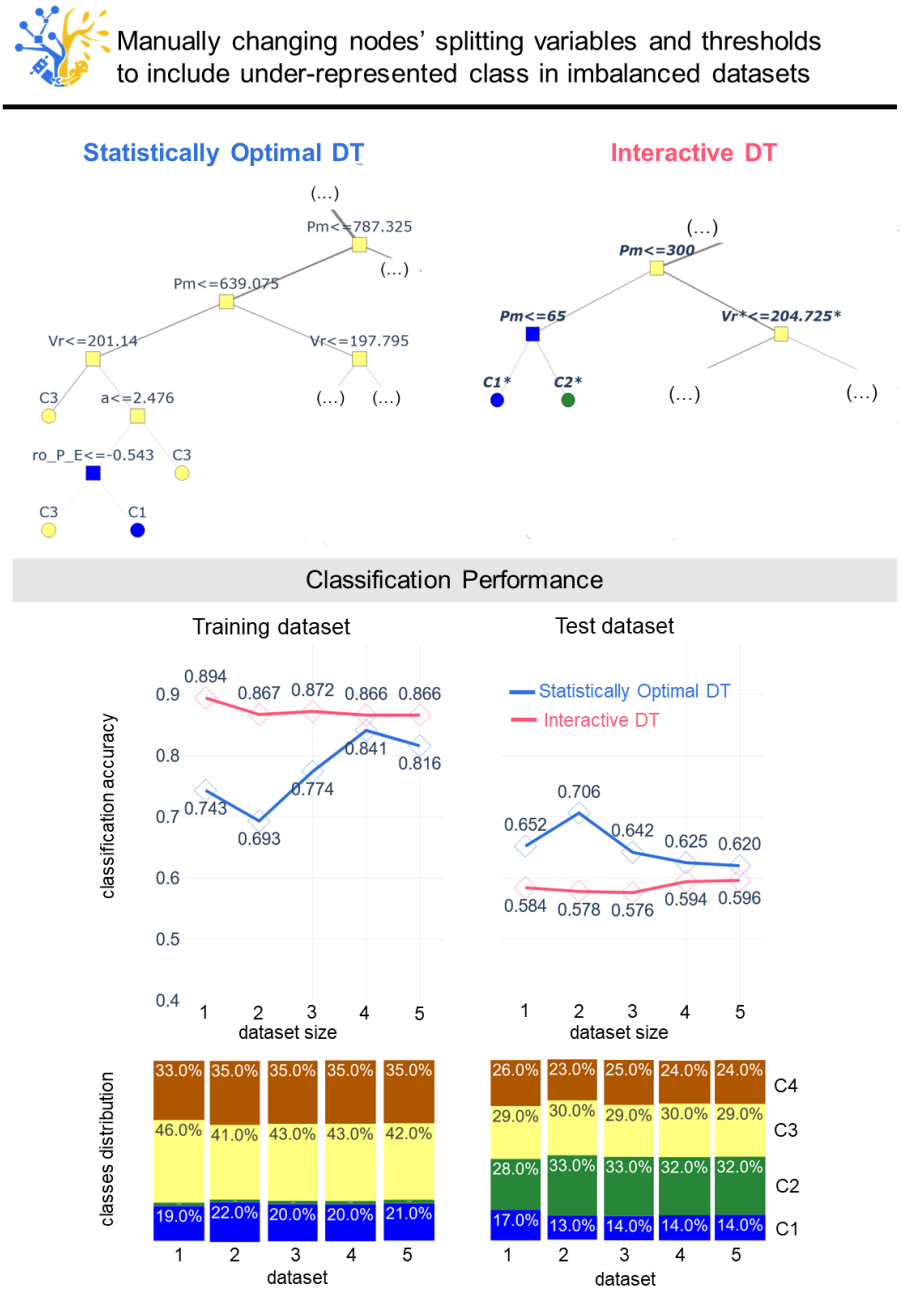


Figure 2. 9 *Showcase of ensuring physical consistency in imbalanced datasets by manually changing splitting variables and thresholds.* ***Top left:*** *Detail of the statistical optimal DT for sample dataset 2.* ***Top right:*** *Interactive DT for the sample dataset 2. In the iDT, the variables and threshold values in italic are those manually changed by the user, and those marked with an asterisk are changed by the DT algorithm in response to the manual*

*changes. **Bottom**: Panel showing the classification accuracies on the training and test sets, and the distribution of the four output classes (C1-C4), in each of the 5 datasets (bottom).*

2.7 Conclusions

How I can incorporate scientific knowledge into ML models to improve their physical realism, their robustness and their interpretability remains a major challenge and opportunity for ML applications in the geosciences (Read et al., 2019; Sun et al., 2022). To address this problem, I propose a framework for the construction and analysis of interactive decision trees (iDTs) for application in the geosciences. I created an open-source implementation of iDT in Python and Jupyter Lab, which I hope will encourage the use of iDT in future research applications. I demonstrated the value of iDT approach in three case studies that represent typical challenges encountered in applications of decisions trees in the geosciences. I found that the proposed iDT framework supports the development of decision trees that are easier to visualise and interpret in a physical sense. In the second case study, I find that manual adjustment of splitting thresholds can lead to a more physically meaningful tree with almost no loss in classification performance. In the third example, I show how experts can build a more robust and physically consistent DT in cases of imbalanced datasets that can generalize better on unseen data. Even though manually changing the nodes' variables and threshold values based on domain knowledge to consider an under-represented class deteriorated the classification accuracy in training sets, it improved it in test sets.

In Chapter 4, I will use again iDT to analyse the relationship between dominant input uncertainties of a flood risk model, and the system characteristics over a large spatial domain. As will be shown in that Chapter, iDT will help identifying an interpretable tree and deal with under-represented categories.

Chapter 3: Characterization of input uncertainties in flood risk modelling: application to the Rhine River basin

3.1 Introduction

Uncertainty and Sensitivity Analysis is now used in flood risk modelling to understand the impact of input uncertainties on the flood risk estimates uncertainty (Apel et al., 2004, Apel et al., 2008, Merz & Thieken 2009, De Moel & Aerts 2011, De Moel et al., 2014, Sampson et al., 2014, Kaczmarska et al., 2018). Uncertainty analysis (or, output uncertainty quantification) deals with the question of determining the uncertainty in the model output given the level of uncertainty in the model inputs. Sensitivity analysis (or, uncertainty attribution) deals with the question of which input uncertainty mostly contributes to the output uncertainty.

A first step in any uncertainty and sensitivity analysis application is the selection of input uncertainties and their characterization, that is, to assign each input uncertainty with a probability distribution function (PDF) or a variability range or a list of plausible values. These choices represent our confidence/knowledge on what values we think are most likely to be true for a variable and how they are distributed. For example, triangular (Wagenaar et al., 2016, Egorova et al., 2008), beta (Sampson et al., 2014), normal (Apel et al., 2008) and uniform (Saint-Geours et al., 2015, Duha Metin et al., 2018, Wagenaar et al., 2016) PDFs have been used in the past to characterize input uncertainties. A difference of the uniform distribution in respect with the rest is that it assumes that only the range that separates values into plausible and implausible is known.

To characterize input uncertainties in an objective manner, good quality datasets (Bomers et al., 2019, Wing et al., 2020) or expert knowledge/judgement (Cooke, 1991, O'Hagan et al., 2006, Krueger et al., 2012, Morris et al., 2014) could be used. But when it comes to flood risk assessment this is challenging. Good quality datasets are not always available or accessible. For example, regarding the value of exposed assets, economic databases like market values, construction costs etc. and/or exposure market portfolios could be used but they are usually owned by (re)insurance companies and thus are not accessible (Sampson et al., 2014). Besides accessibility issues, scarcity of datasets is another challenge. For example, flood damage data are rarely available (except for data rich countries like US, UK, Netherlands, Germany) to quantify the variability in vulnerability curves (Wing et al., 2020, Bernhoven et al., 2020). There is also the uncertainty that observations of hydrologic

extremes are particularly uncertain. Coxon et al. (2015) studied the uncertainty in observed discharges in gauging stations in England and Wales. They found that uncertainty intervals widely range from 10% to 397%. Regarding the option of expert knowledge elicitation, experts are aware of the existence of input uncertainties in models, but they may find it difficult to express these uncertainties in terms of probability. For example, it is well known there is considerable uncertainty in the vulnerability curves (Merz et al., 2010, McGrath et al., 2019, Wing et al., 2020) but it is still unclear how the damage changes depending on the flood damaging characteristics (e.g., flood depth, velocity, or the building types).

Hence, flood risk modellers usually face the problem that little information is available to guide their choices of input uncertainty distributions (Moret et al., 2017). Scientists in other fields facing similar challenges have suggested to characterize input uncertainties through the definition of variability ranges instead of PDFs. For example, Moret et al. (2017) propose a method to define variability ranges by collecting information for each input uncertainty based on different criteria.

In this chapter, I present a structured approach to define the variability range of the uncertain input parameters of a flood risk model. To exemplify the approach, I focus on the following three input uncertainties: the value of residential buildings, the damage ratio values of the vulnerability curves and the return period of flood events. These input uncertainties are present in almost every flood risk analysis. Moreover, they correspond to different flood risk components (hazard, vulnerability, exposure) and will be used in the sensitivity analysis in Chapter 4, helping us understand the relative importance of these 3 components. The approach is based on the synthesis of quantitative and qualitative information from literature sources. I looked at studies/reports where input uncertainties were defined (for a specific region(s) within the Rhine River basin), either by giving ranges of variability or a list of possible options. Then I extracted and combined these characterisations to define plausible ranges for the entire basin. I apply the approach to the Rhine River basin, however, it is transferable to similar applications and the reported ranges can be used as a starting point in future applications in the same and/or similar areas.

3.2 Methodology

3.2.1 Rhine River basin

The Rhine River basin (shown in Figure 3.1) is a major European river with an average discharge of about 2,900 m^3/s which is among the highest in Europe (Merz & Thieken 2009).

It starts from high mountainous areas to rocky canyons and lowland plains and passes through major European cities of 8 different countries (Switzerland, Liechtenstein, Austria, France, Germany, Belgium, Luxembourg and the Netherlands). The river can be divided into 8 sections/sub-basins. *High Rhine* contains the origins of the Rhine and the river that leaves Lake Constance in a westerly direction and takes in the river Aare. *Alpine Rhine* section includes the confluence of the Anterior and Posterior Rhine which flows in a south-north direction until it reaches Lake Constance. The *Alpine and High Rhine* are in steep, high mountainous areas covered by glaciers. The *Neckar* sub-basin is located near Mannheim and it is where the river Neckar River flows into the Rhine. River Neckar has its source at a height of 706 m over sea level and flows over steep slopes. Cities and ports located in Neckar sub-basin include Plochingen, Stuttgart, Heilbronn, and Mannheim. The *Upper Rhine* is the section where the river changes direction from West to North. Important cities located here include Basel, Strasbourg, Colmar, Mulhouse. *Moselle-Saar* sub-basin is on the west of *Upper Rhine* and it flows through France, Luxembourg and Germany. *Middle Rhine* is the section where the river leaves the Mainz basin and enters the western part of the Rhenish Massif. In this part, there are meanders that have cut canyons of 200–300 depth into the rocks. *Main* is the sub-basin where the Main River flows into Rhine. Important cities in this sub-basin are Mainz and Frankfurt. *Lower Rhine* is the part where it enters the North German Plain. It falls from 50m to 12m and it is a typical lowland river. The Lower Rhine is characterized by intense industrial activities and a high density of population. Major ports and cities located here include Cologne, Düsseldorf, Neuss and Duisburg. Finally, the *Rhine Delta* roughly coincides with the political borders of Germany and the Netherlands, and it is the largest delta in Europe. The Rhine delta is situated at the southern margin of the North Sea Basin (Preusser, 2008; Merz & Thieken, 2009). Other large European cities that Rhine passes through are Karlsruhe, Bonn, Leverkusen, Arnhem, Utrecht, Rotterdam.

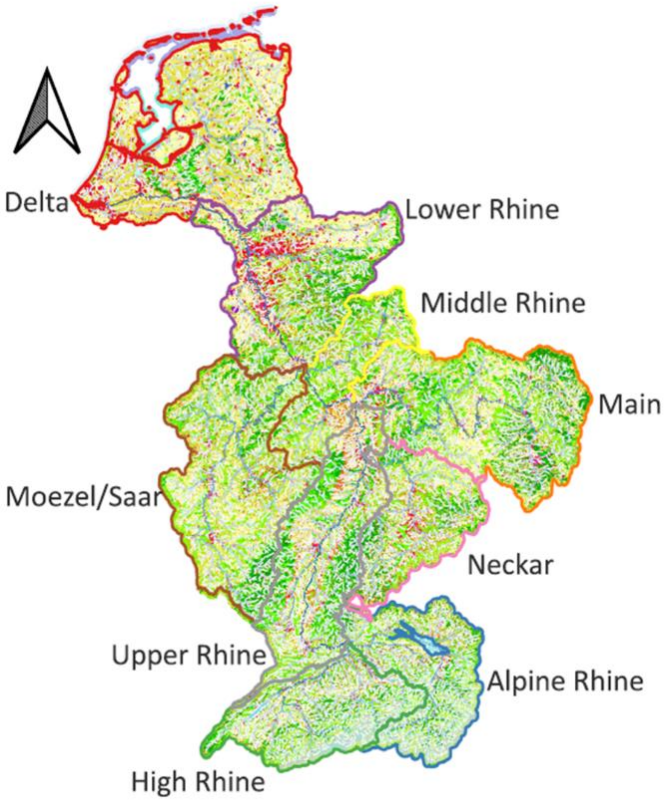


Figure 3. 1 *Study Area (Rhine River Basin)*

3.2.2 Selection of input uncertainties

A list of input uncertainties in flood risk modelling is shown in Table 3.1 (for more information on sources of uncertainty for each component see Gerl et al., 2016, Pittore et al., 2017, and Beven et al., 2018). I have selected to quantify the uncertainty in the value of residential buildings, the damage ratios of the vulnerability curves and the return period of the flood events (highlighted in bold letters in Table 3.1). They appear in three different components (hazard, vulnerability, exposure) of the flood modelling risk chain, so comparing their relative importance in Chapter 4 will enable us to compare the relative importance of the 3 components. Besides, they are quantities that are used in almost every flood risk analysis. Residential assets are considered an important component of each nation's wealth (Piketty & Zucman, 2014) and they account for the biggest share of damages induced by floods (Paprotny et al., 2020). Moreover, every vulnerability model uses an estimate of the residential asset's value, either expressed per unit of area or in terms of individual buildings (Paprotny et al., 2020), to calculate flood losses. The vulnerability component is considered as an important driver of uncertainty in flood losses estimation (Jongman et al., 2012a, Wagenaar et al., 2016, Wing et al., 2020). Currently more than half of the available vulnerability models worldwide use curves that express the damage in relative terms, that is, through damage ratios. Moreover, the majority uses only flood depth to relate flood impact to

a damage (Gerl et al., 2016). Finally, an extreme value model is used in every flood risk analysis to estimate the return period of flood events (Beven et al., 2018). It starts with selecting the annual maximum discharges of the observed discharges, or peak values that exceed a certain threshold (Schendel & Thongwichian, 2017). A probability distribution function is then fitted to the peak values. The fitted distribution is then used to estimate discharges for various return periods. These can then be used as input to inundation models to simulate flood events (depths and extents) of various return periods. Usually, a set of flood events with different return periods is used to estimate losses for a range of flood events.

Table 3. 1 *List of input uncertainties for each flood risk component. In bold, I highlight the input uncertainties I have selected to quantify.*

Flood Risk Component	Type of Uncertainty	Input Uncertainty
Exposure	Exposed Assets Uncertainties	***Asset Value*** Asset Location Number of Assets Asset Footprint Area Asset Geometric features (e.g., height) Financial Sector (e.g., residential, commercial, industrial, etc.) Level of detail (e.g., individual objects or aggregated land use classes)
	Disaggregation Uncertainties	Disaggregation method Proxy dataset (e.g., population, nightlights)
Vulnerability	Vulnerability Curves	**Damage Ratios** Shape of the curve Flood damaging characteristics (e.g., flood depths, velocity, contamination etc.) Level of details (e.g., individual objects or aggregated land use classes) Model concept (deterministic, probabilistic)
	Defences	Standard of Protection Protected Area Location
Hazard	Extreme Value modelling uncertainties	**Return Period of flood events** Assumptions of extreme value statistics (e.g., homogeneity, independence, stationarity) Choice of sample (Annual Maximum Series, Peaks over threshold) Choice of Probability distribution function Choice fitting method Length of data records
	Event Selection uncertainties	Rules used to define a flood event
	Inundation Modelling Uncertainties	Definition of the river network Hydraulic Model (1D, 2D) Inaccuracies in datasets (e.g., DEM) Inaccuracies in channel geometry Channel parameters (e.g., roughness coefficient) Model assumptions (e.g., fixed channel capacity)

3.2.3 A systematic approach to explore the literature on the quantification of input uncertainties.

In this section I suggest a structured approach to characterize input uncertainties. I am suggesting that it is possible to look at how past studies have quantified the variability of relevant input uncertainties, even if their scale of application differs, and synthesize the collected information to extract variability ranges for the input uncertainties and scale of application of interest in this thesis. To my knowledge, there has been no previous attempt to do such literature review, with the exception of Moret et al., (2017) in the energy sector. They suggested to characterize input uncertainties based on different criteria. Among the criteria for example are the use of historical datasets and/or expert judgement, which however in flood risk modelling is challenging to use as explained in the introduction.

My approach starts with the literature search for each input uncertainty chosen for the case study. To search the literature I used the Google Scholar, Scopus, and Semantic Scholar engines. I used the key words *Uncertainty, Variability Flood risk, Rhine, Europe*, for every input uncertainty. Additionally, I used the keywords *residential building value, flood damage assessment, asset value* for the residential buildings' value, *flood damage model, flood damage assessment, vulnerability functions, depth-damage functions, damage ratios* for the vulnerability curves, and *flood frequency curves, extreme value modelling, peak discharges, return period*. To choose the relevant papers I read the abstract and looked at the figures and tables.

The next step is to process the information retrieved from each literature source. This could include unit transformation, referencing values to the same time period or scaling values to nearby areas. If additional information is required in the processing, it could be retrieved either from existing open databases or based on a literature search. Asset values for example can be expressed either in square meters or individual objects. Square meters values could be transformed to individual objects values using indexes that express the average area of the single asset. Moreover, it is quite often that they are referenced to a specific year or time period. Hence, when comparing values from different sources it is necessary to reference them in the same year. This could be done using indexes such as price indexes.

The final step is to estimate the minimum and maximum values. Each study provides its own characterisation of uncertainty for a specific area within the study area. The final minimum

and maximum values for the study area are estimated by averaging the minimum and maximum values from these individual characterizations of uncertainties found in each study using equation (3.1):

$$\bar{X}_{i,min} = \frac{\sum_{s=1}^{S} X_{i,min,s}}{S}$$
$$\bar{X}_{i,max} = \frac{\sum_{s=1}^{S} X_{i,max,s}}{S} \qquad (3.1)$$

Where:

$\bar{X}_{i,min}$, $\bar{X}_{i,max}$, are the final minimum and maximum values for input uncertainty i.

$X_{i,min,s}$, $X_{i,max,s}$, are the minimum and maximum values for input uncertainty i, for study s.

S is the total number of studies.

There are two cases: studies that provide a range of variability and studies that provide a list of possible values. In the former case one can directly apply equation 3.1. In the latter case, one first needs to derive minimum and a maximum value from that list before applying equation 3.1.

In the following sub-sections, I exemplify the approach with an application on the Rhine River Basin and focusing on the three selected input uncertainties.

3.3 Results

3.3.1 Residential Buildings Values

Literature Search

I found nine studies (listed in Table 3.2) that reported minimum and maximum Residential Buildings Value (RBV) for areas within the Rhine River basin. The scale of application for the different studies may range from local (e.g., dike ring, catchment, city) to large (e.g., country).

The studies provided estimates of RBV for urban, agricultural and industrial areas. I only considered the values provided for urban areas. The studies did not provide estimates of RBV depending on the building type except for Paprotny et al., (2020), and hence, I did not make such distinctions in my analysis. Moreover, the studies provided estimates of RBV in market, replacement and/or depreciated values which reflect the value of the buildings differently. Depreciated values represent the reduced value of the assets after considering

the loss in their value that naturally occurs through time. They reflect a value at the actual time of a flood and are useful to estimate the damage to the national economy (Messner et al., 2007). Replacement costs reflect the costs which would be necessary to replace/rebuild the buildings at the time of the event and are more relevant for the insurance companies or individuals (Messner et al., 2007). Market values represent the value of the assets on the real estate market. They reflect the total value of the building (including the land and contents) (Messner et al., 2007).

Table 3. 2 *List of studies, their scale, area and type of values, considered in the quantification of uncertainty of Residential Buildings Value.*

Studies	Scale	Area	Type of values
De Moel et al. (2011)	Dike Ring	South Netherlands	Market values
Jongman et al. (2012a)	City	Eilenburg (Germany), Carlisle (UK)	Replacement, Depreciated, Market values
Cammerer et al. (2013)	Catchment	Northwestern Austria	Replacement values
Jongman et al. (2014)	Country	The Netherlands	Market values
Wagenaar et al. (2016)	Dike Ring	Southwest Netherlands	Replacement values
Huizinga et al. (2017)	Country	Europe	Depreciated values
Lüdtke et al. (2019)	Country	Europe	Replacement values
Paprotny et al. (2020)	Country	Europe	Replacement values
Sieg et al. (2022)	Country	Germany	Market values

For this input uncertainty I extracted variability ranges at country level and more specifically the 8 countries that Rhine River flows over, averaging the values provided by the nine studies.

Processing the retrieved information

In some cases, estimates of residential building values were given in Eur/m^2. I converted them into Eur/Building per country by using the average building area of each country.

$$RBV_{min,r,s} = RAV_{min,r,s} * BA_r \ , \quad RBV_{max,r,s} = RAV_{max,r,s} * \ BA_r \qquad (3.2)$$

Where:

$RBV_{min,r,s}$, $RBV_{max,r,s}$ is the minimum and maximum value in Eur/Building for region r in study s

$RAV_{min,r,s}$ $RAV_{max,r,s}$ is the minimum and maximum value in Eur/m^2, for region r in study s

BA_r is the building area for region r estimated by averaging values provided by JBA (personal communication), Huizinga et al. (2007) and Eurostat (2022a).

In cases where ranges for residential buildings values were not available for a region, I calculated them from estimates for a nearby region, using the construction costs indexes of the two regions, as follows:

$$RBV_{min,r,s} = RBV_{min,j,s} * \frac{BCC_{r,s}}{BCC_{j,s}} \ , \quad RBV_{max,r,s} = RAV_{max,j,s} * \frac{BCC_{r,s}}{BCC_{j,s}} \qquad (3.3)$$

Where:

$RBV_{min,j,s}$, $RBV_{max,j,s}$ are the minimum and maximum values available for the nearby country j,

$BCC_{r,s}$ is the building construction costs for country r retrieved from Eurostat (2022b)

$BCC_{j,s}$ is the building construction costs for country j retrieved from Eurostat (2022b)

I selected 2011 as the reference year. In studies where residential building values were estimated at a different year, I rescaled them using equation 3.4:

$$RBV_{min,r,s} = RBV_{min,r,s,y} * \left(\frac{HP_{r,2011}}{HP_y}\right), \quad RBV_{max,r,s} = RBV_{max,r,s,y} * \left(\frac{HP_{r,2011}}{HP_y}\right) \qquad (3.4)$$

Where:

$RBV_{min,r,s,y}$, $RBV_{max,r,s,y}$ is the minimum and maximum residential building value for region r and year y for study s, respectively

$HP_{r,2011}$ is the house price index for country r for the year 2011 retrieved from Eurostat (2022c).

$HP_{r,y}$ is the house price index for country r for the year y retrieved from Eurostat (2022c).

Estimation of minimum and maximum values

I estimated the final minimum and maximum values per country using equation 3.1. I used the values from all the available studies without converting to a specific type of values (Market, Depreciated or Replacement). The results are shown in Figure 3.2 (and in Table B.1 in the Appendix B).

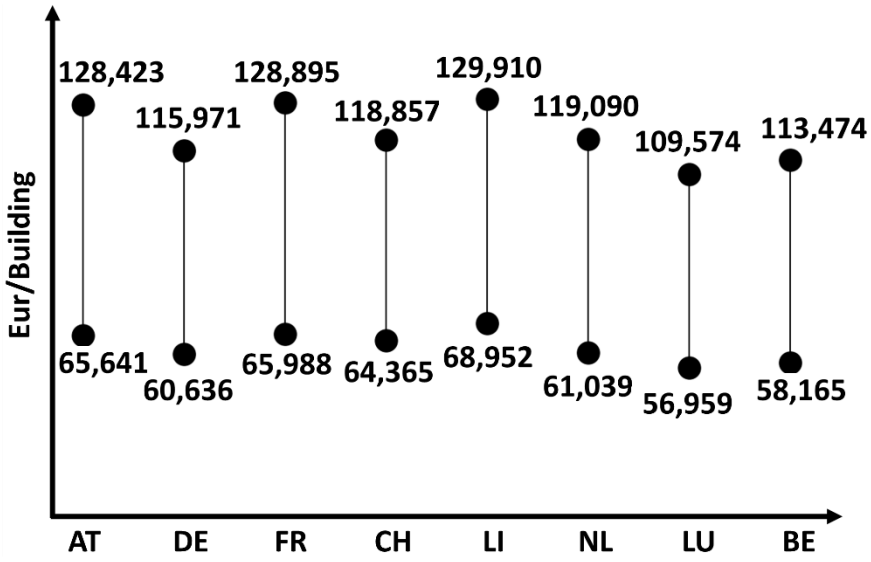


Figure 3. 2 *Minimum and Maximum Residential building values (in Euros) per country in the Rhine River basin*

3.3.2 Damage ratio values of the vulnerability curves

Literature search

I found eight studies listed in Table 3.3 each using a different set of vulnerability curves. In these studies, one can see that while the scale of application of the vulnerability curves ranges from local (e.g., river reach, dike ring and/or city) to large (e.g., country and/or continent) the same vulnerability curves can be used. Hence, despite the differences in scale among the studies I used the curves I found within them to extract variability ranges for the vulnerability curve for the whole Rhine.

Table 3. 3 *List of studies, their scale, areas, and vulnerability curves, considered in the quantification of uncertainty of vulnerability curves. The letters in the columns correspond to the following curves a: Rhine Atlas, b: MURL, c: Hydrotec, d: Damage Scanner, e: Flemish, f: Netherlands later, g: Polynomial, h: FLEMO, i: HAZUS, j: Multi Coloured Manual, k: JRC (Germany), l: linear, m: square root, n: HIS-SSM, o: Tebodin/Billah, p: JRC (UK), q: JRC. With red colour I indicate the curves I did not consider as they were developed for countries not belonging to the Rhine River basin.*

Studies	**Scale**	**Area**	**Vulnerability Curves**																
			a	**b**	**c**	**d**	**e**	**f**	**g**	**h**	**i**	**j**	**k**	**l**	**m**	**n**	**o**	**p**	**q**
Apel et al. (2009)	City	Germany	+	+	+					+									
Bubeck & De Moel (2010)	Country	Netherlands	+			+													
De Moel & Aerts (2011)	Dike Ring	Netherlands	+				+	+											
Elmer et al. (2010)	Country	Germany		+	+				+	+									

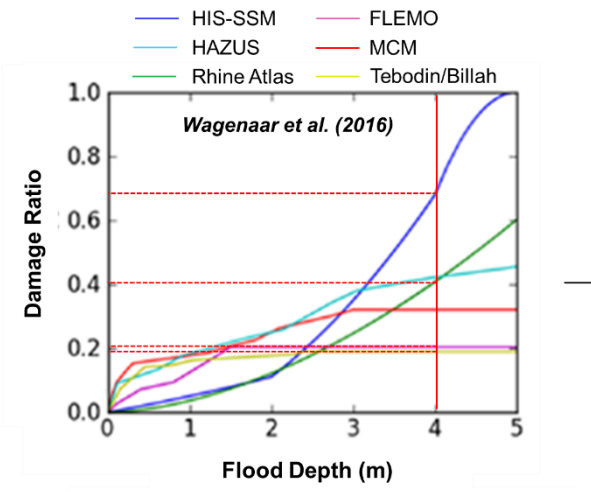
Jongman et al. (2012a)	City	Eilenburg (Germany), Carlisle (UK)	+			+	+			+	+	+	+					+	
Cammerer et al. (2013)	Catchment	Northwestern Austria	+	+	+				+	+				+	+				
Wagenaar et al. (2016)	Dike ring	Southwest Netherlands	+							+	+	+				+	+		
Huizinga et al. (2017)	Continental	All continents																	+

Moreover, these studies include in total 17 vulnerability curves. Cammerer et al. (2013) tested the transferability of different curves in residential areas and found that curves derived from similar regions performed much better than the curves from dissimilar regions. Hence, for this case study I discarded the curves which were developed for other countries than the ones through which the Rhine River flows. These were the HAZUS curve (developed for the US), Multi Coloured Manual curves (UK) and the ones developed by the Joint Research Centre (JRC) for the UK. Of the curves I considered, only the Rhine Atlas was developed specifically for the Rhine River basin, whereas the FLEMO and MURL models were developed for Germany, the Flemish for Belgium, the HIS-SSM, Damage Scanner, Netherlands Later and Tebodin for the Netherlands, and JRC and JBA for the whole Europe.

Estimation of minimum and maximum values

First, I estimated minimum and maximum vulnerability curves for each study. To do that I used the different vulnerability curves shown in figures within each study. In Figure 3.3 I exemplify this with the study of Wagenaar et al. (2016). For each flood depth (e.g., red vertical line at flood depth 4m) I extracted[4] the damage ratio value for each curve (e.g., red dashed horizontal lines). I recorded the values in a table and then calculated the minimum and maximum values, and range at each flood depth. The range is defined as the difference between the maximum and minimum values.

[4] I used the Graph Grabber 2.0.2 software to extract the values from the published figures.



Wagenaar et al. (2016)					
Flood Depth (m)	**1**	**2**	**3**	**4**	**5**
HIS-SSM (NL)	0.06	0.12	0.34	0.69	1.00
Rhine Atlas (DE)	0.05	0.12	0.25	0.41	0.60
Flemo (DE)	0.13	0.20	0.20	0.20	0.20
Tebodin/Billah	0.17	0.19	0.195	0.195	0.195
Min	0.05	0.12	0.195	0.195	0.195
Max	0.17	0.2	0.34	0.69	1
Range	0.12	0.08	0.145	0.495	0.805

Figure 3. 3 *An example (based on a figure retrieved from Wagenaar et al., (2016)) for estimating minimum and maximum vulnerability curves.*

The final minimum and maximum values (shown in Figure 3.4) are calculated by averaging the minimum and maximum values of all studies using equation 3.1. The final range at each flood depth is shown in Table 3.4.

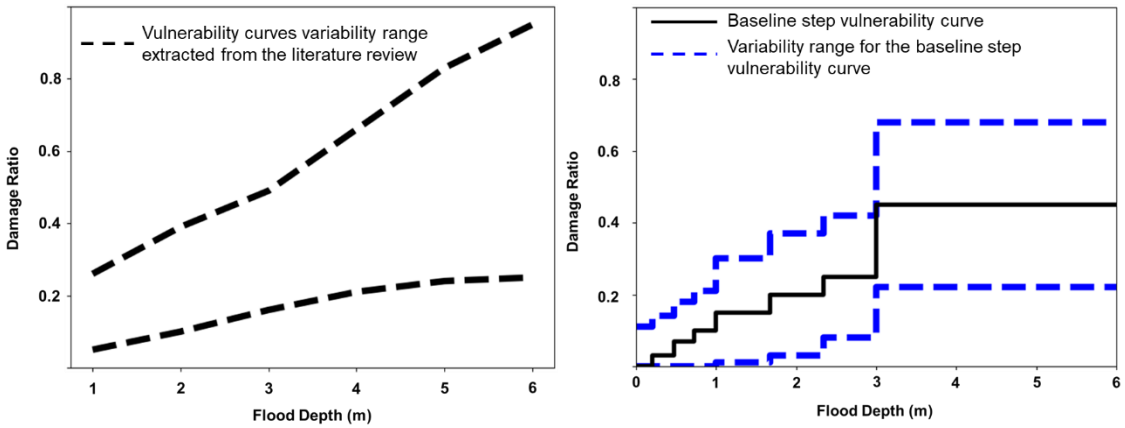


Figure 3. 4 *Vulnerability curves variability ranges.* ***Left:*** *Minimum and Maximum vulnerability curves resulted from the literature.* ***Right:*** *Minimum and Maximum (blue dashed lines) vulnerability curves for the step vulnerability curve of the flood risk model I used in Chapter 4. The blue dashed lines were extracted based on the ranges of Table 3.4 and applying equation 3.5.*

Table 3. 4 *Variability range for damage ratios at various flood depths*

Flood depth (m)	Range
1	0.22
2	0.29
3	0.34
4	0.45
5	0.58
6	0.70

Table 3. 5 *Perturbation range for the baseline vulnerability curve function used in Chapter 4. The 1st column shows the flood depths and the second column the corresponding minimum and maximum damage ratio values.*

Flood Depth (m)	Perturbation Range (Min-Max)
0-0.2	0-0.11
0.2-0.5	0-0.14
0.5-0.7	0-0.18
0.7-1	0-0.22
1-1.7	0.01-0.3
1.7-2.34	0.03-0.34
2.34-3	0.08-0.42
>3	0.22-0.68

In the sensitivity analysis application of Chapter 4 I characterized the uncertainty in vulnerability curves by perturbing the baseline step function (black line in the right plot of Figure 3.4) of JBA Risk Management Ltd flood risk model within a range (blue dashed lines in the right plot of Figure 3.4). To define this range of perturbation I used the ranges in Table 3.4 and equations 3.5:

$$Vuln_{min} = dr_i - \frac{Rn_i}{2}$$
$$Vuln_{max} = dr_i + \frac{Rn_i}{2} \quad (3.5)$$

Where dr_i is the default damage ratio value at flood depth i

Rn_i is the range at flood depth i reported in Table 3.4

The ranges I defined in Table 3.4 are per 1 meter of flood depth. While the steps of the baseline function are in higher resolution (<1 meter). When applying equation 3.5, for each step of the function I used as Rn_i the range of the flood depth interval the step belongs to in Table 3.4. For example, for every step with depth <1m, Rn_i in equations 3.5 will be equal to 0.22. The resulting range of perturbation for each step is shown in Table 3.5. One can notice that while the resulting range from the literature in the left plot of Figure 3.4 is increasing for flood depths >3m, the range of perturbation in the right plot of Figure 3.4 does not increase. This happens because in the baseline function the last step contains all flood depths >3m and so this step got the same range. Moreover, in the first 4 steps the minimum damage ratio value is always zero. This happens because when applying equation 3.5 to the baseline damage ratio would result in negative values which is unrealistic and so I decided to put the minimum damage ratio equal to zero.

3.3.3 Return period of the flood events

Literature Search

For this input uncertainty, I found five studies in total (shown in Table 3.6). Three of them performed extreme values analysis based on observed discharges for the Cologne Station. Each study used datasets of different record length. The other two did the same for the Lobith station, again using datasets with different record lengths. In this case it is sensible to average estimates that refer to the same station. Thus, the spatial unit for this input uncertainty is the gauging station. More specifically, I used the estimates from Merz & Thieken (2005), Merz & Thieken (2009), Apel et al. (2008) for Cologne and Toonen et al. (2015) and Bomers et al., (2019) for Lobith.

Table 3. 6 *List of studies, their gauging station, dataset and time periods, considered in the quantification of uncertainty of vulnerability curves.*

Studies	Gauging Station	Dataset	Time Period
Merz & Thieken. (2005)	Cologne	Annual Maximum series	1880-1999
Merz & Thieken (2009)	Cologne	Annual Maximum & Peaks over Threshold series	1846-2004
Apel et al. (2008)	Cologne	Annual Maximum series	1961-1995
Toonen et al. (2015)	Lobith	Annual Maximum series	1901-2011
Bomers et al. (2019)	Lobith	Annual Maximum series	1901-2018

Estimating the minimum and maximum values

The studies included figures providing estimates of minimum and maximum discharges for various return periods. I used these figures to estimate minimum and maximum Return Period for various discharges. In Figure 3.5[5] I exemplify this with the study of Bomers et al. (2019). For various discharges, e.g., red horizontal lines, I estimated the minimum and maximum return periods, e.g., red dashed vertical lines. I then plotted the minimum and maximum return periods for the various discharges.

Then, using equation (3.1) I calculated the minimum and maximum values for each gauging station using the estimates from the corresponding studies. The results are summarized in Figure 3.6 (and in Table B.2 in Appendix B).

[5] I used the Graph Grabber 2.0.2 software to extract the values from the published figures.

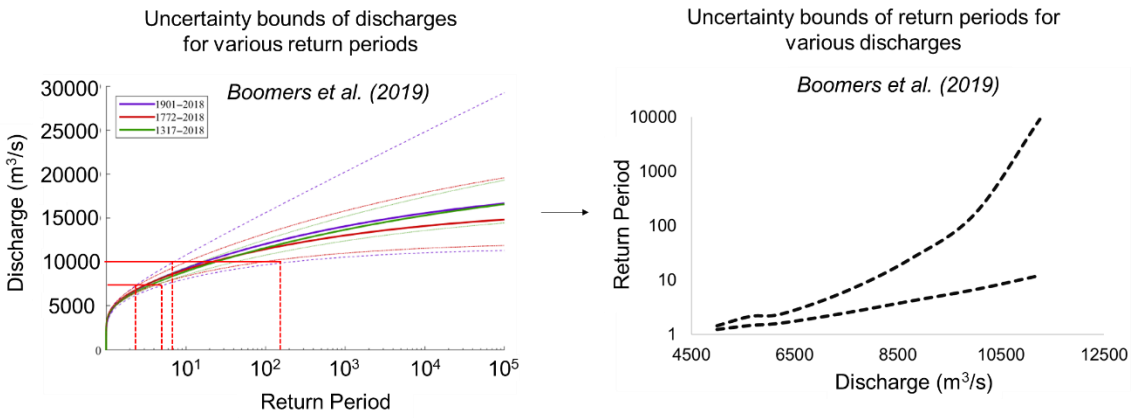


Figure 3. 5 *An example (based on a figure retrieved from Bomers et al. (2019)) for estimating min and maximum values of return periods for various discharges.*

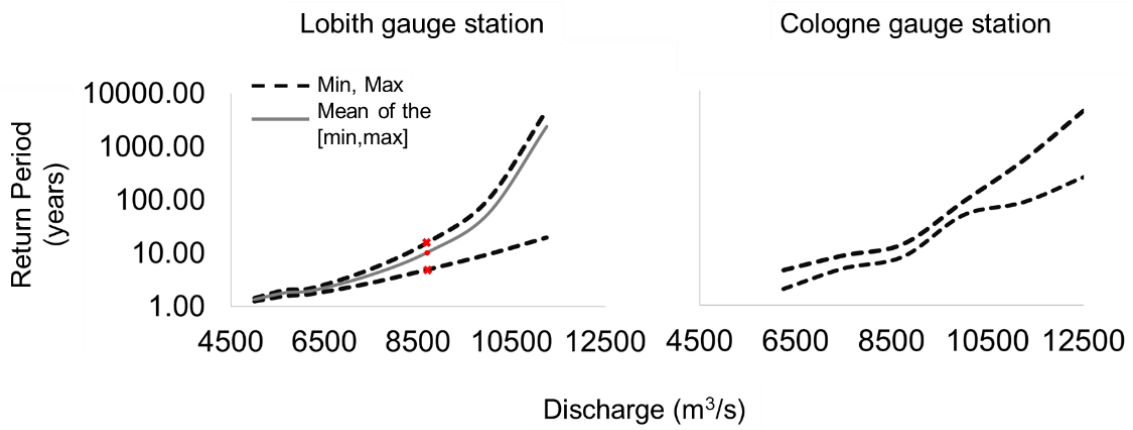


Figure 3. 6 *Minimum and Maximum values for the return period (in years) depending on the peak discharges for the Lobith and Cologne gauging stations.*

These graphs provide variability ranges for return periods depending on the discharge. Additionally, I extracted variability ranges for various return periods without having to pass through a discharge value. I summarize them in Table 3.7. In Figure 3.6 and the Lobith plot, I exemplify how I calculate the range for the 10-year return period. The grey solid line is the mean of the variability range (black dashed lines). I found where on this line the 10-year return period lies (red dot). Then, following a vertical line down and up until intersecting lower and upper black dashed lines I calculated the min and max values for the 10-year return period respectively (red crosses). I did the same for the rest of the return periods shown in Table 3.7.

Table 3. 7 *Minimum and Maximum values of Return Period for various return periods for the Lobith and Cologne gauging stations.*

Lobith										
Return Period	**2**	**5**	**10**	**15**	**20**	**50**	**100**	**200**	**500**	**1500**
Min RP	1.75	3	4.5	5.2	6.4	8	10	11	12.5	15
Max RP	2.4	7	15	25	34	88	200	385	894	3200
Cologne										
Return Period	**2**	**5**	**10**	**15**	**20**	**50**	**100**	**200**	**500**	**1500**
Min RP	1.5	4	8.4	12	16.2	23	33	48	86	-
Max RP	2.6	6	12	18	24.6	75.4	164	371	1014	-

3.4 Discussion

The resulting residential buildings values are generally low in respect with current values. One explanation is that most of the studies considered are more than five years old and all values are referenced back to year 2011. Hence, any recent temporal variations in economy that may affect building values were probably not accounted for. For example, none of the studies considered current high inflation rates which can significantly impact the value of a building. Moreover, as expected the resulting variability ranges differ, but not significantly, among the regions. The differences existing among the countries can be justified by temporal and spatial variations. Temporal variations exist due to economic trends, investments, inflation rates. Spatial differences may exist due to differences in wages, material costs, etc., (Merz et al., 2010). Moreover, there is great heterogeneity among the residential properties' characteristics (Eurostat, 2013). They can vary in terms of their structural type (e.g., detached, semi-detached, etc.), construction material (e.g., concrete, steel, etc.) and age.

Regarding the damage ratios, the ranges are wide for all the different flood depths considered and they are increasing with increasing flood depths. These wide variability ranges stem from the differences existing in the various models considered. The vulnerability curves models were developed using different approaches (e.g., empirical, expert judgement, synthetic). For example, FLEMO was developed based on flood damage data, the Tebodin model using expert judgement, the Damage Scanner, the Flemish, Rhine Atlas and JBA models using an approach that combined flood damage data and information retrieved from a literature review. Moreover, the different models even though they have been used in case studies in Rhine, they were originally developed for various regions as mentioned in the previous section. Finally, another reason is that all models that were used to estimate the variability ranges are based only on flood depth to translate the flood impact to damages. According to the literature, even though flood depth is the main contributor to

flood damages it does not explain the total variability in observed damages (Gerl et al., 2016). For example, Merz et al. (2004) used a flood damage database for residential buildings and a non-parametric regression model based on flood depth to show that there is great variability in flood losses for a water depth of $\geq$ 1m, and that the flood depth explains only a part of that variability. Thieken et al. (2005) analysed the impact of different flood characteristics to residential buildings losses and found that variability in losses increases with increasing flood depths, which might explain why the differences among the models are more pronounced for larger flood depths. Other flood characteristics like duration, velocity and water contamination are deemed to also contribute to the total variability of flood losses.

Finally, regarding the Return Period of the flood events there is large uncertainty for large return periods ($\geq$100 years). One explanation is the short record lengths (around 100 years) used in the local studies to perform the extreme value analysis. Extrapolation of the measured discharges from such records to return periods much larger than the record length results in large uncertainty for the predicted discharges (Bomers et al., 2019) as shown in Figure 3.6. The values shown in Table 3.6 are extracted based on the discharges shown in Figure 3.6 hence it is sensible that there is large uncertainty. Moreover, I couldn't extract variability ranges for the whole Rhine River basin because I collected information only for two gauging stations. The variability ranges extracted for these two stations are representative for the local areas but probably not for the whole Rhine River basin. Thus, I deemed that the sample of two stations was too small to extrapolate to the whole Rhine River basin. For the sensitivity analysis application of the next chapter, I will define my "own" variability range for the Rhine River basin, and I will check how this compares to the ranges I calculated for the two stations.

3.5 Conclusions

In flood risk modelling our knowledge of input uncertainties is limited and the probability distribution are practically unknown. How we quantify the input uncertainties in such cases remains an open challenge, yet a crucial step in uncertainty and sensitivity analysis to understand the robustness of risk model outputs. In this chapter I presented a methodological approach for addressing this challenge. I assumed that independent and uniform prior distributions are the most appropriate distribution type given the lack of knowledge. A literature review is then performed to determine minimum and maximum values for these uniform distributions. I exemplified the approach with an application to the Rhine River basin for three input uncertainties in flood risk modelling. I defined variability ranges for the residential buildings value and the vulnerability curves. These ranges will be

used in a sensitivity analysis application in the next chapter. Regarding the return period of the flood events, I extracted variability ranges only for two gauging stations because I could not find relevant information for more stations. The reported ranges for the two stations can still be useful for local scale flood risk applications. This approach can be transferred to similar applications and areas. In the next chapter I explore how the uncertainties defined here propagate through a flood risk model of the Rhine River.

Chapter 4: Linking the relative importance of input uncertainties of a flood risk model to River Rhine spatial characteristics

4.1 Introduction

Comprehensive flood risk assessments are necessary to better manage the risk associated with floods (European Commission, 2007, United Nations, 2005, 2015). Nowadays, numerous flood risk models are available to simulate flood risk (usually expressed in terms of average annual losses or selected quantiles of losses over a year) at various scales, from catchment to regional or even global scale (Ward et al., 2015, Trigg et al., 2016, Kaczmarska et al., 2018). They involve a complex modelling chain that estimates risk as the product of the hazard, exposure and vulnerability (Duha Metin et al., 2018; Beven et al., 2018). These chains may differ among the various models, but some basic components are common. A typical modelling chain is shown in Figure 1.1 in the introduction chapter. Each element in the chain can contain numerous aleatory and/or epistemic input uncertainties, that propagate and contribute to the final uncertainty in risk estimates (Apel et al., 2004; Merz & Thieken, 2009; Beven et al., 2018).

Different studies across Europe have tried to quantify output uncertainty of flood risk models and attribute it to its various sources along the modelling chain. I have summarized the conclusions of these studies in Figure 4.1. Green dots mark the study locations and their size is proportionate to the size of the study domain. The boxes are coloured according to which input uncertainty was found to dominate uncertainty in flood loss (risk) estimates. I see that studies reached different conclusions regarding what uncertainty mostly influences flood risk estimates within a particular study domain.

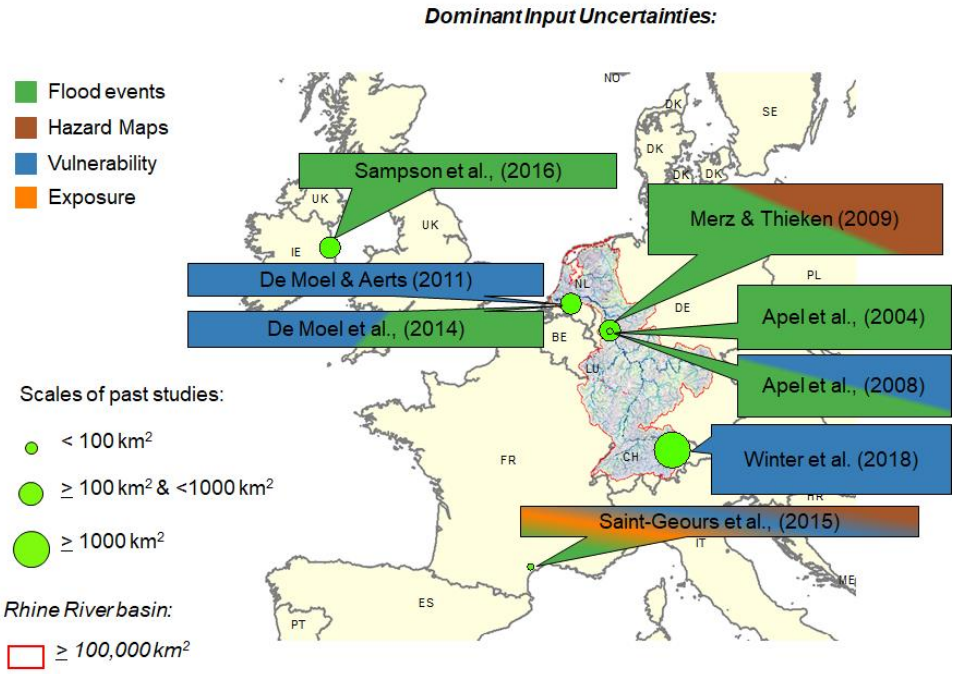


Figure 4. 1 *Summary of local scale uncertainty attribution studies results in the Rhine River basin and Europe. The boxes are coloured according to the conclusion of each study as to which input uncertainty dominates the uncertainty in flood loss estimates.*

There are two possible explanations for these ambiguous outcomes. One explanation may be that the studies used different flood risk models and characterization of input uncertainties, which can significantly impact the sensitivity analysis results (Pappenberger et al., 2008; Pianosi et al., 2016). For example, Merz & Thieken (2009) used a peaks over threshold model to estimate peak discharges, which they then transformed into flood depths and extents using two different 2-dimensional inundation models. Apel et al. (2008), instead, used five different extreme value distributions to estimate peak discharges and a 1-dimensional inundation model to simulate flood depths and extents. These studies also differ in the characterization of input uncertainties. For example, Merz & Thieken (2009) and de Moel & Aerts (2011) used different vulnerability curves to quantify the uncertainty in flood loss estimates. Saint-Geours et al. (2015) used a single depth-damage curve and defined a [-50%, 50%] variability range around their default curves. Apel et al. (2004) applied a randomizing procedure on the three-parameter polynomial function representing the vulnerability curve used to model damage estimates.

The second explanation is that the studies reviewed above were conducted in different cities and catchments over a large region (ranging from the Mediterranean to central and Northern Europe). Hence the differences in the dominant input uncertainties could be due to the climatic, hydrological or socio-economic differences across the study sites.

In this chapter, I study the question of uncertainty attribution across a much larger domain than those investigated in the previous studies, namely the heterogeneous region of the Rhine River basin (described in Section 3.2.1). The goal is to provide evidence on how the importance of input uncertainties varies across places with different climatic, hydrologic, land cover and socio-economic characteristics. To this end, I use an industry flood risk (catastrophe) model provided by JBA Risk Management Ltd (hereafter JBA) which is capable of simulating flood risk across such a large region. In this way, I can compare results from different sub-regions in the domain while using the same model, methods and data everywhere. I use a combination of different statistical methods to analyse the sensitivity of model outputs to the input uncertainties with the aim to:

1. Identify the dominant input uncertainties within the spatial domain.
2. Find the system characteristics that explain the spatial variability of the dominant input uncertainties

4.2 Methodology

4.2.1 Loss Calculation

Like all Flood Risk Models (FRM) (or catastrophe models, Mitchell-Wallace et al., 2017), JBA FRM uses multiple components and inputs to estimate flood risk in terms of economic losses induced by floods (Figure 4.2). The estimation starts by calculating the loss for a single exposed asset and a single flood event according to Equation 4.1:

$$L_{s,t} = EV_s \times DR\left(FD_{s,t}\right) \qquad (4.1)$$

Where $\mathrm{L}_{s,t}$ is the loss at exposed asset s for event t, EV_s is the asset value, FDs,t is the flood depth and $DR(.)$ is the vulnerability curve that returns the damage ratio as a function of the flood depth.

Equation 4.1 is applied for all the exposed assets in the spatial domain and for a range of flood events over a 10,000-year period. Such a large time period is used so that one can robustly calculate statistics of the loss frequency distribution. In this Chapter I will consider two: the Average Annual Losses (AAL), i.e., is the average of the per-year losses across the a range of flood events, and the Loss Exceedance Curves (LEC), i,e, the level of annual losses that is exceeded with given probability. These loss statistics at individual asset point can then be spatially aggregated over user-defined spatial units. In this project, I aggregated losses at Catastrophe Risk Evaluating and Standardized Target Accumulation (CRESTA)

high resolution zones (CRESTA, 2022). This is a standard spatial aggregation unit widely used in the (re)insurance sector (Grossi & Kunreuther, 2005, Mitchell-Wallace et al., 2017, Kaczmarska et al., 2018). For a CRESTA Zone (denoted by *C* hereafter), the Average Annual Loss AAL_c is calculated using the following equation:

$$AAL_c = \sum_{s \in C} \sum_t p_{s,t} L_{s,t} \qquad (4.2)$$

Where $p_{s,t}$ is the annual probability of event *t* happening at the location of asset *s*.

The level of annual losses that is exceeded with given probability in the C-th CRETSA Zone, $LEC_C(p)$, is defined as:

$$LEC_C(p) = X \quad such\ that \quad \mathbb{P}(L_{C,Y} \geq X) = p \qquad (4.3)$$

where *p* is the probability value chosen by the user, and $L_{C,Y}$ is the expected annual loss in Zone C in year Y, calculated as:

$$L_{C,Y} = \sum_{s \in C} \sum_{t \in Y} p_{s,t}\, L_{s,t} \qquad (4.4)$$

Note that the probability *p* appearing in Equation (4.3) can equivalently be expressed (and will be expressed throughout this chapter) as a Return Period (RP) (in years):

$$RP = 1 / p \qquad (4.5)$$

4.2.2 Calculation of flood depths, damage ratios and exposed asset value

The flood depth at each location of the study domain for a given flood event (FDs, t) can, in principle, be estimated by running a flood inundation model with river flow and/or rainfall as forcing inputs characterising that event. A large number of flood events is necessary to approximate the probability of exceedance $\mathbb{P}(L_{C,Y} \geq X)$ from the empirical distribution function for $L_{C,Y}$. However, running the inundation model for millions of different flood events at such large scales is computationally prohibitive. To overcome this problem, the flood inundation model is first used to derive Flood Hazard Maps for a limited set of return periods (in this case 20, 50, 100, 200, 500 and 1,500 years) and then flood depths for events of any other return period are obtained by interpolating through the flood depths of the appropriate hazard maps (see Fig. 4.2). For example, if the exposed asset is attributed with a return period of 147 years for an event, then the model will interpolate between the depths of the maps with return periods 100 and 200. In the Appendix C1 I provide more information on JBA flood risk model and on the flood events and hazard maps generation.

The flood events considered for the calculations of AAL and/or LEC are pre-computed and stored in a Flood Event Set. It contains millions of plausible flood events (observed and simulated in a 10,000-year period). Each event is assigned with a return period and its location, year and duration. Typically, the event sets are generated using a historical record of observed events in which synthetic events, generated using physical (e.g., Global Circulation Models) and statistical models (e.g., extreme value models, Heffernan & Tawn, 2004) are added with the purpose that the final record will cover events that has not been yet observed (Grossi & Kunreuther, 2005, Lamb et al., 2010, Keef et al., 2013, Mitchell-Wallace et al., 2017, Kaczmarska et al., 2018).

The estimated flood depth (FDs,t) is then used to read the damage ratio (DR) from the vulnerability curve. The damage ratio values range between 0 (no damage) and 1 (total loss). Here step functions were used (as also done for instance in Apel et al., 2009, Cammerer et al., 2013, Sairam et al., 2021). The use of step functions means that it is assumed that a range of sufficiently similar flood depths all lead to the same damage ratio. Other studies have used different shapes for the vulnerability functions (Bubeck & de Moel, 2010, De Moel & Aerts, 2011, Jongman et al., 2012a, Wagenaar et al., 2016). I will further discuss the implications of this choice in Section 4.4.

The estimated damage ratio value is multiplied by the value of the exposed asset (EVs) which is retrieved from the exposure portfolio. The exposure portfolio contains all the relevant information regarding the exposed assets (e.g. location, total value etc.). Here I fabricated an exposure portfolio based on a market portfolio provided by JBA Risk Management. The spatial resolution of this market portfolio is at CRESTA zones level, as it is usually the case in flood risk modelling (Mitchell-Wallace et al., 2017, Kaczmarska et al., 2018). I then disaggregated it at coordinate level based on the number of exposed assets in each CRESTA zone and by sampling against the distribution of a proxy dataset (population).

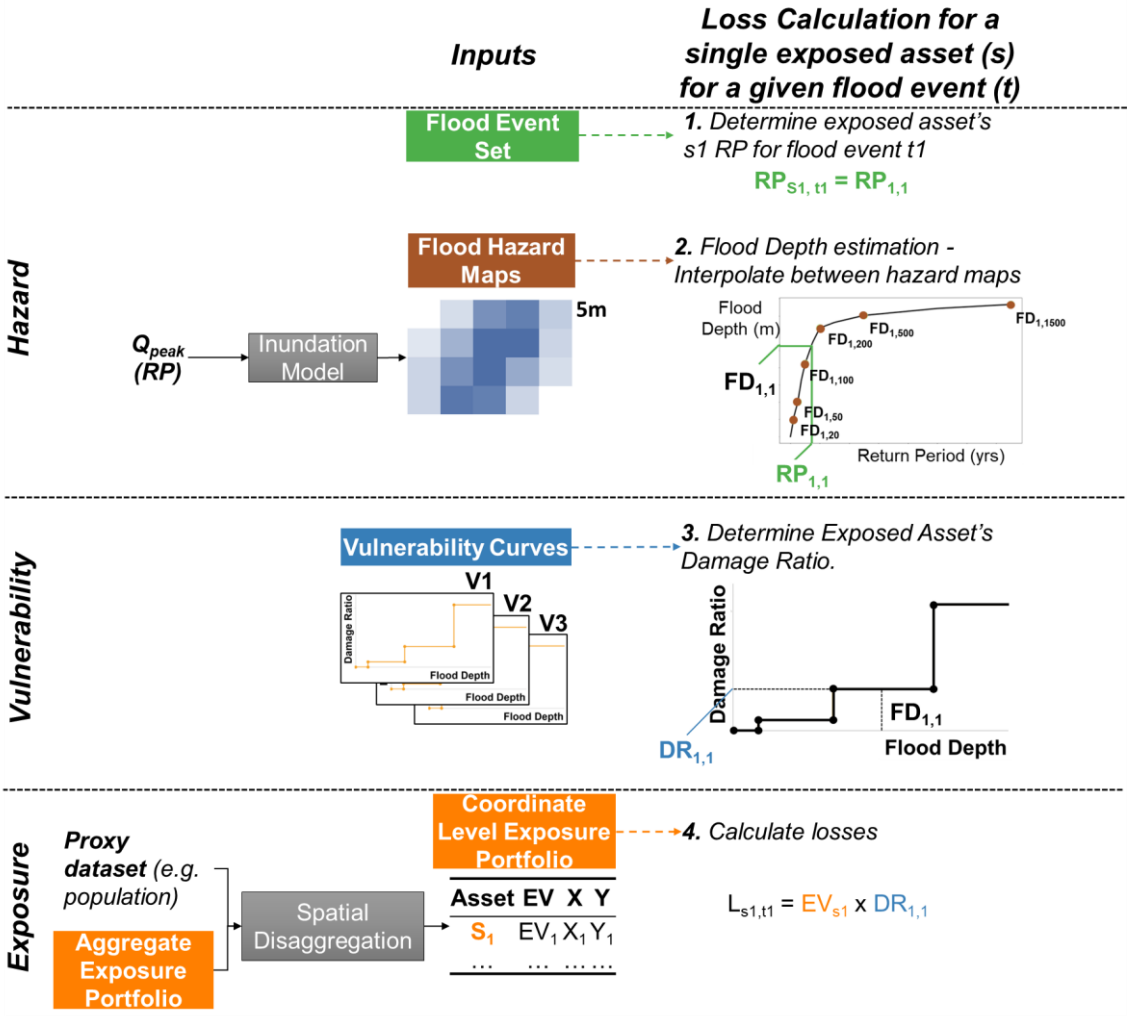


Figure 4. 2 *Schematic representation of loss calculation for a single exposed asset for a single event.*

4.2.3 Global Sensitivity Analysis

I selected to analyse the uncertainty associated with four inputs: (1) the value of the exposed assets, which in this case coincides with the residential buildings' value (RBV) as the assets are all residential; the Damage Ratio values of the vulnerability curves; (3) the return period of the hazard maps; and (4) the return period of the flood events in the event set. Note that these input uncertainties refer to different flood risk components (exposure, vulnerability and hazard). Thus, performing sensitivity analysis will enable us to compare the relative importance of the three components. It is worth noting that by changing the return period of the Hazard maps both the flood depths and extents are perturbed.

I used global sensitivity analysis (GSA) to formally assess the relative importance of these input uncertainties on the AAL and the LEC at different return periods. The four steps of

GSA were described in the introduction and in Figure 1.2. Here I briefly describe how I applied each step to this case study:

1. *Identification and characterisation of input uncertainties*. For the characterization of the uncertainty in residential buildings' value and the vulnerability curves I adopted the ranges reported in Chapter 3 and summarized in Figure 4.3. For the return period of the hazard maps and of the flood events, in Chapter 3 I could not extract variability ranges for the whole Rhine basin, hence here I decided to apply a +-50% uniform perturbation of the return periods from their default values. In Table 4.1 I show how this compares to the ranges I extracted from the literature for the two stations of Cologne and Lobith. In Lobith the +-50% uniform perturbation underestimates the range for return periods $\geq$15 years, closely estimated the 10-year return period range and overestimates the range for lower return periods (<10 years). In Cologne, there is an underestimation for return periods $\geq$100 years, a good fit for the 50-year return period and an overestimation for the return period $\leq$20 years.
2. *Sampling inputs' variability space*. I randomly sampled combinations of input uncertainty values from the input variability space using a random uniform distribution and the Latin Hypercube Sampling strategy. I generated a sample of size 400.
3. *Model execution against each input value combination.* I executed the model against each input combination.
4. *Calculation of sensitivity indices*. Finally, to calculate sensitivity indices I used the PAWN method (Pianosi & Wagener, 2015, Pianosi & Wagener, 2018). In the PAWN method, the sensitivity of the output y to the input factor x_i is quantified by measuring the distance between the unconditional cumulative distribution function (CDF) of y that is obtained by varying all inputs simultaneously, and the conditional CDF obtained when all inputs vary but x_i. Operationally, if one has a dataset of input-output samples (e.g., generated by latin hypercube sampling as in point 2), the PAWN sensitivity indices are approximated by splitting the range of variation of each input uncertainty x_i into n equally spaced intervals I_k and using the following equations:

$$\hat{S}_i = mean_{k=1,\dots,n}\, KS(I_k) \qquad (4.6)$$

$$KS(I_k) = max_y \left| \hat{F}_y(y) - \hat{F}_{y|x_i}(y|x_i \in I_k) \right| \qquad (4.7)$$

Where $\hat{F}_y(y)$ and $\hat{F}_{y|x_i}(y|x_i \in I_k)$ are the unconditional and conditional CDFs of the output y and KS is the Kolmogorov-Smirnov (KS) statistic that measures the distance between the CDFs (Kolmogorov, 1933, Smirnov, 1939).

The sensitivity indices vary between 0 and 1: the higher the value of the sensitivity index the more sensitive the output to the input uncertainty. PAWN has been shown to effectively estimate sensitivity indices with a relatively low number of model evaluations (Pianosi & Wagener, 2018) which is useful in this context to use a lower sample size and reduce the computational time of this experiment. To assess the robustness of the estimated sensitivity indices values, bootstrapping (Efron & Tibshirani, 1994, Archer et al., 1997) was used to repeat the calculation of the sensitivity indices against a prescribed number of random resamples of the original input-output dataset, yielding a statistical distribution of the sensitivity indices.

Table 4. 1 *Comparison of the return period ranges resulting from the +-50% uniform perturbation from default values with the return period ranges extracted from the literature (in Chapter 3) for the gauging stations of Cologne and Lobith.*

Lobith										
Return Period	**2**	**5**	**10**	**15**	**20**	**50**	**100**	**200**	**500**	**1500**
Min RP	1.75	3	4.5	5.2	6.4	8	10	11	12.5	15
Max RP	2.4	7	15	25	34	88	200	385	894	3200
Cologne										
Min RP	1.5	4	8.4	12	16.2	23	33	48	86	-
Max RP	2.6	6	12	18	24.6	75.4	164	371	1014	-
Uniform +-50% perturbation										
-50%	1	2.5	5	7.5	10	25	50	100	250	750
+50%	3	7.5	15	22.5	30	75	150	300	750	2250

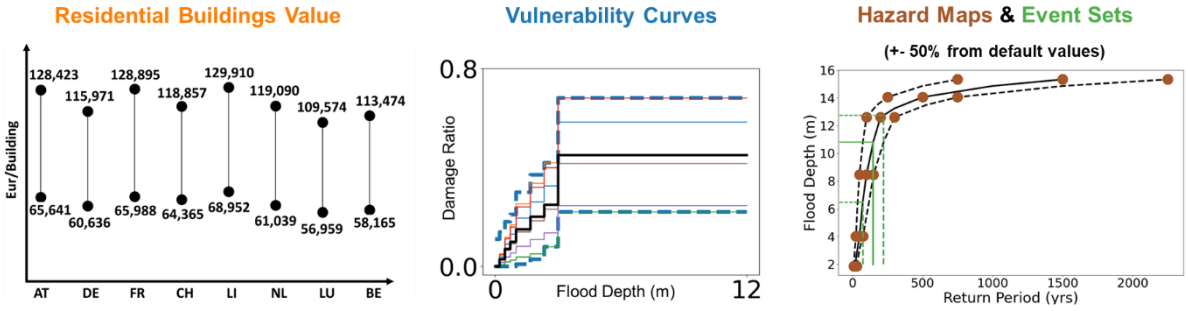


Figure 4. 3 *Variability ranges for each input uncertainty.*

4.2.4 Definition of dominant input uncertainties

I calculate sensitivity indices to rank the input uncertainties with respect to their relative contribution to the output uncertainty and determine the most important ones (dominant input uncertainties) in each spatial unit. However, as mentioned in Sec. 4.2.3, by using bootstrapping a distribution of sensitivity indices is actually derived for each input uncertainty in each unit. An intuitive approach to define the dominant input uncertainty may be to

compare the means of the bootstrap resamples. In Appendix C2 I provide reasons why this approach may not work well. Here I propose a new way to define the dominant input uncertainty that considers the full distribution.

1. For each bootstrap resample, I determine the dominant input uncertainty as the one with highest sensitivity index.
2. For each input uncertainty (say the i-th), I calculate the frequency (FR_{x_i,s_i}) with which it is identified as dominant across bootstrap resamples:

$$FR_{x_i,s_i} = \frac{N_{x_i}}{M} \qquad (4.8)$$

   Where N_{x_i} is the number of bootstrap resamples in which the i-th input uncertainty has the highest sensitivity index, and M is the total number of bootstrap resamples (=500 in this case).
3. An input uncertainty is said to be dominant in a given CRESTA zone if its frequency FR_i is the highest and exceeds the second highest by at least a prescribed difference (0.3 in this case which was a judgement call). If instead the difference between the first and second highest frequencies is less than 0.3, I deem both input uncertainties as dominant.

4.2.5 Linking dominant input uncertainties with system characteristics

The last step of the analysis is to investigate whether I can link the dominant input uncertainty in a given CRESTA zone with any of its characteristics. To do this, I used a range of different variables representing climatic, hydrological, socio-economic and land cover properties for each CRESTA zone (shown in Table 4.2). I included hydrological variables characterizing river properties such as length, average discharge and stream order, because I investigate riverine floods. I also included a categorical hydrological variable which indicates the river section or sub/catchment the CRESTA zone belongs to, because they are quite different according to the description of the study area in Chapter 3 and it also serves as a proxy for the topography. More specifically, Lower Rhine and the Delta are dominated by flat terrains while the rest of the sections and sub-catchments by steep terrains. Hydrological variables were retrieved from HydroSHEDS (HydroRivers) dataset (Lehner & Grill, 2013). Additionally, I included variables describing the flood depths within the CRESTA zone because flood depth is one of the main damage driving flood characteristics. These were retrieved from JBA's flood maps. I included socio-economic variables such as population and residential buildings' value because they are the main exposure characteristics that influence flood losses estimation. Population was retrieved from the World Population Dataset (WorldPop, 2020) and the residential buildings' value

from the exposure portfolio. As climate can influence floods, I included a variable representing the Koppen Climate Class (Beck et al., 2018) of the CRESTA Zone. Finally, I included a categorical variable which represents land cover properties of each CRESTA zone, because land cover can potentially greatly impact floods in an area (Rogger et al., 2017). I modelled CRESTA Zones' land cover properties using CORINE dataset (Copernicus, 2018). The datasets used to retrieve the values for each variable have been frequently used in the past in flood risk applications: HydroSHEDS (Alfieri et al., 2013, Lindersson et al., 2021), WorldPop (Smith et al., 2019, Lindersson et al., 2021, Bernhoven et al., 2022), CORINE (Wünsch et al., 2009, Jongman et al., 2012a, Paprotny et al., 2018).

Table 4. 2 *Summary of the variables representing the hydrological, socio-economic, land cover and climatic properties of the study area.*

Property	Source	Variable Name	Description	Units	Range of Values	Dataset format
Hydrological	Lehner & Grill (2013)	RivLen	River length	Km	[0-103]	Shapefile
		AvDisch	Average discharge	m^3/s	[0-2427]	Shapefile
		MaxStOr	Maximum river order indicator at a CRESTA zone according to Strahler ordering system	-	[1-7]	Shapefile
		HydBas	River section or sub-catchment where the CRESTA zone belongs to.	-	1: High Rhine 2: Alpine Rhine 3: Upper Rhine 4: Neckar 5: Main 6: Middle Rhine 7: Moezel/Saar 8: Lower Rhine 9: Delta	Shapefile
	JBA Flood Hazard Maps	MajFD	Most frequent flood depth value	m	[0-12]	Raster
		MeanFD	Average Flood Depth	m	[0.02-5.8]	Raster
		MaxFD	Maximum Flood Depth	m	[0-12]	Raster
		FD3	Percentage of flood depths over 3m	-	[0-100]	Raster
Socio-Economic	Chapter 2 values	MeanVal	Mean value of residential	Euros	[276,027-52,629,147]	csv

			buildings			
	WorldPop (2020)	Pop	Number of People	-	[4-1,804]	Raster
	CRESTA (2022)	Area	Area of the CRESTA zone	Km^2	[0.2-235]	Shapefile
Climatic	Beck et al. (2018)	ClimClass	Koppen Climate Class	-	15: Temperate, no dry season, warm summer 26: Cold, no dry season, warm summer 27: Cold, no dry season, cold summer 29: Polar, tundra	Raster
Land Cover	Copernicus (2018) (CORINE)	ARTF	Percentage of artificial surfaces	-	[0-100]	Raster
		AGR	Percentage of agricultural surfaces	-	[0-100]	Raster
		FRST	Percentage of forests and semi natural areas	-	[0-100]	Raster
		WETLN	Percentage of wetlands	-	[0-100]	Raster
		WB	Percentage of water bodies	-	[0-100]	Raster

I analyse the potential links between these variables and the dominant input uncertainties by using the interactive Decision Trees (iDT) algorithm presented in Chapter 2. The dataset I fed to the DT consists of the values of the above variables for each cresta zone as inputs and the dominant input uncertainty of each CRESTA zone as outputs. It is worth to mention that each CRESTA zone is associated with one dominant input uncertainty. Hence, there are four different classes (the dominant input uncertainties: Residential Buildings Value, Vulnerability Curves, Hazard Maps and Event Set) in total for this case study. The resulting DT model(s) will show the combination of variables (system characteristics) lead to each class (dominant input uncertainty).

4.3 Results

4.3.1 Uncertainty in Average Annual Losses is dominated by the uncertainty in the damage ratios

The results for the AAL are summarized in Figure 4.4. The map on the left shows the Rhine river basin, divided into CRESTA Zones which are coloured according to the dominant input uncertainty on Annual Average Loss (AAL) predicted for that zone. In white colour I represent the zones where no data were available and therefore no losses and sensitivity indices were calculated. The bar plot in the top right corner summarises the total number of CRESTA zones in which a specific input uncertainty is dominant. Figure 4.4 shows that uncertainty in AAL is dominated by the uncertainty in damage ratio values with very few exceptions.

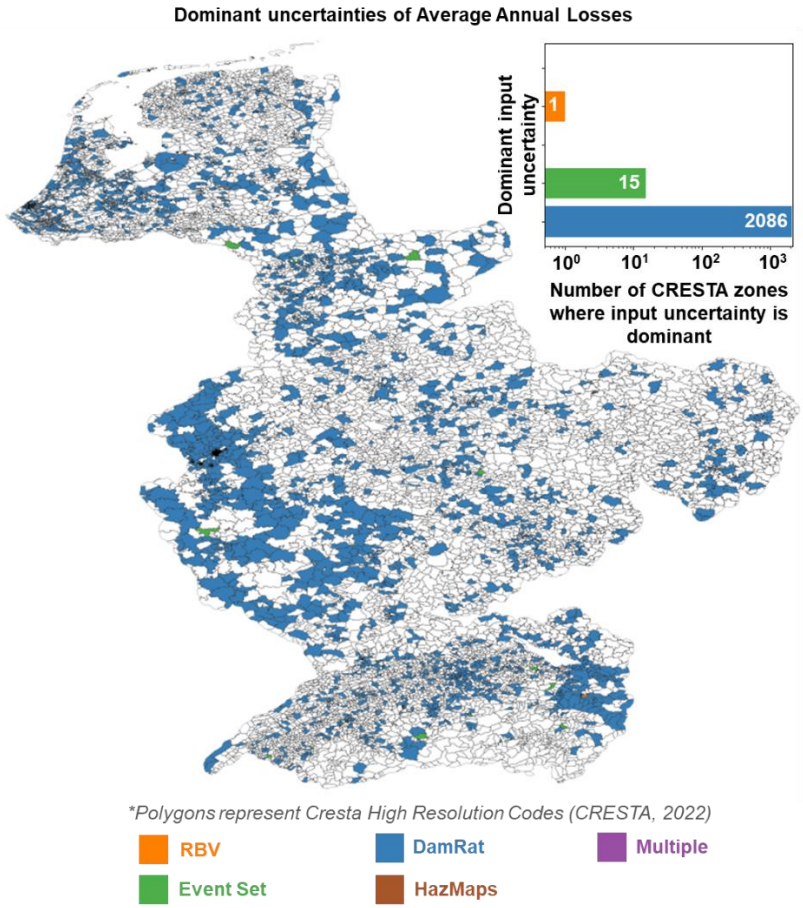


Figure 4. 4 *Map of Rhine River basin with dominant input uncertainties on Annual Average Losses (ALL) at CRESTA zones (CRESTA, 2022). The bar plot in the top right corner summarises the total number of CRESTA zones in which a specific input uncertainty is dominant. RBV: Residential Buildings Value, DamRat: Damage Ratio, HazMaps: Hazard Maps*

4.3.2 Dominant input uncertainties change while moving from lower to larger return periods in the Loss Exceedance curves

The results for Loss exceedance curves (LEC) for six different return periods are summarised in Figure 4.5. The LEC uncertainty is dominated by damage ratio and event set uncertainties for small return periods (≤ 20 years). There are also see some cases where Hazard Maps and Exposed Value (EV) uncertainties dominate the LEC uncertainty. This might be explained by the fact that at lower return period, losses are likely to be dominated by small and frequent flood events where the impact of localized features is greater. Thus, uncertainties in estimates of flood depths and/or value of individual assets become more important for the estimation of losses. As the return period increases, the damage ratio becomes dominant in more and more CRESTA zones while the influence of other input uncertainties progressively fades away. The number of CRESTA zones with multiple dominant input uncertainties (purple bar) also decreases with increasing return period. These may be explained by the fact that, at large return periods, losses are dominated by floods with large inundation extents where many assets experience similar flood depths and thus flood loss uncertainty is mainly driven by the uncertainty in the damage ratios.

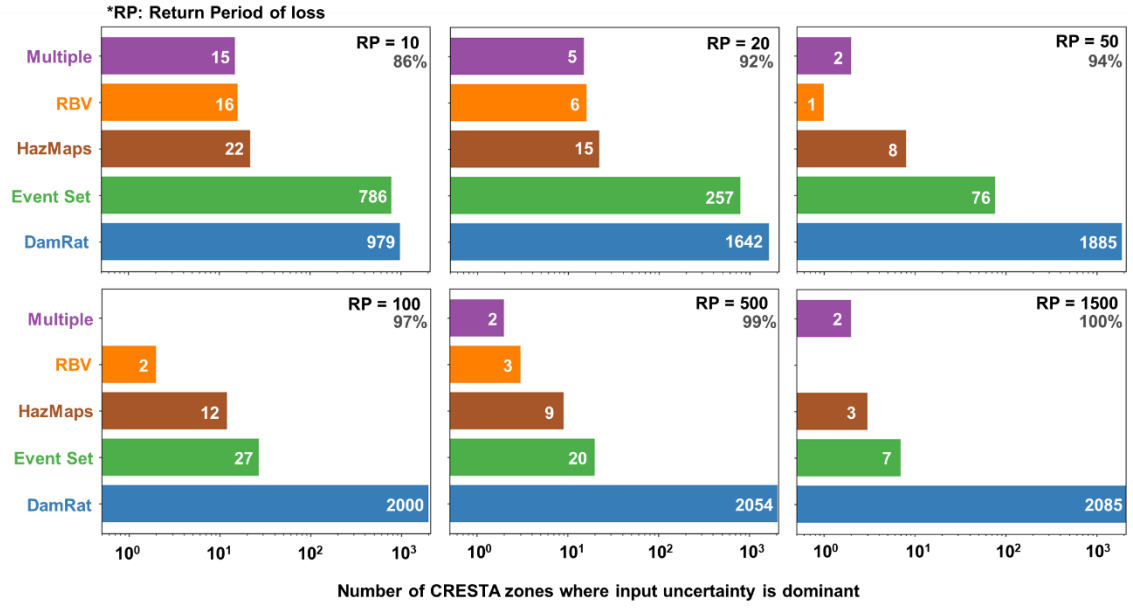


Figure 4. 5 *Bar plots showing the number of CRESTA Zones where each input uncertainty dominates the uncertainty in Loss exceedance curves (LEC) for different return periods (in years). The percentage in the top right of each plot indicates the number of CRESTA zones sensitivity indices were calculated because flood loss occurred.*

4.3.3 An Interactive Decision Tree can link each dominant input uncertainty to a combination of system characteristics

This part of the analysis is based on the results of LEC with return period of 10 years. I chose this output because it is the one showing greater variability in the dominant input uncertainties. I trained both a conventional Decision Tree (here called a Statistical Optimal Tree, to highlight that the conventional DT algorithm aims for statistical optimality) and an interactive Decision Tree using the method of Chapter 2. Both trees are trained to use the explanatory variables in Table 4.1 to predict the dominant input uncertainty into each CRESTA zone. It is worth mentioning that I trained several Statistically Optimal DTs (see Appendix C4) because there is an inherit randomness in the algorithm (e.g., it randomly selects the data to split the dataset into train and test sets) which may lead to the generation of different DTs during the training process. However, in this case this randomness had little impact on the emerging patterns in the different trained DT and their associated accuracies.

For the sake of simplicity, I start by analysing and comparing only the leaf nodes of the different DTs (instead of their entire structure). Specifically, the top panel of Figure 4.6 shows the leaf nodes of a randomly picked statistically optimal DT (the full DT is shown in the Appendix C4 in Figure C.3). The content of the leaf nodes is visualized with a set of stacked coloured bars which show the distribution of dominant input uncertainties in the CRESTA Zones falling in that node. If one colour prevails in the leaf node (e.g., green in the first node on the left) it means that the leaf node mainly contains CRESTA Zones with the same dominant input uncertainty (e.g., Event Set). In other words, the purity of the leaf node is high, and one can confidently classify that leaf node as representative of places where that particular input uncertainty is dominant. When instead there is no clearly prevailing dominant uncertainty (to prevail, it needs to represent $\geq$ 70% of the CRESTA zones included in a leaf node), the leaf node is very impure and is not really representative of any specific category of places. Leaf nodes of this kind are labelled as “undefined” in Fig. 4.6. Under each bar three numbers are noted. The first number indicates how many CRESTA zones fall into that leaf. The second number indicates how many of the CRESTA zones in that leaf are associated with the dominant input uncertainty prevailing in that leaf. The third number is the percentage of CRESTA zones associated with that dominant input uncertainty with respect to the total over the entire Rhine River basin. For example, looking at the first bar in the statistically optimal DT, the numbers show that there are 185 CRESTA zones in that leaf node, of which 150 have Event set (green) as dominant input uncertainty, which is the 27% of all the CRESTA zones in the Rhine River basin where event set was found to be dominant. Hence, when the latter number is small (for example, less than 5%) then the leaf node explains only a “tiny” portion of the overall variability of the dominant input uncertainty,

meaning further analysis of these leaves is probably not particularly meaningful. These cases are represented in Fig. 4.6 with empty bars.

In summary, the statistically optimal DT has 13 leaf nodes, of which however only 4 enable a clear classification of the prevailing dominant input uncertainty and at the same time explain a sufficiently high proportion (ie. >5%) of zones associated with that dominant input. These nodes represent places where either the Damage Ratios (blue) or the Event set (green) are the dominant uncertainties, but no meaningful node is found associated to Hazard Maps or the Value of Residential buildings. This means that this statistical optimal DT may be further analysed to look for the explanatory variables that make the uncertainty in Damage ratio or Event set be dominant, but it would not provide us with any possible explanation of where and why other input uncertainties are dominant.

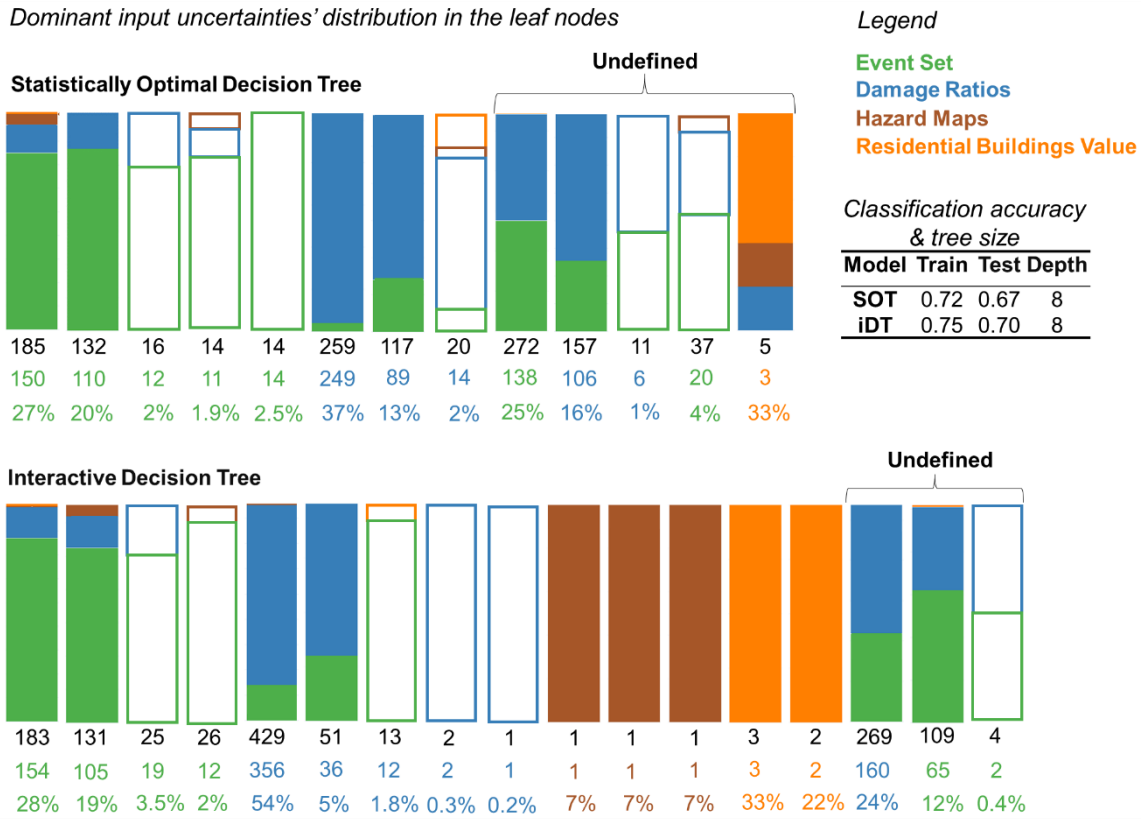


Figure 4. 6 *Evaluation of statistically optimal and interactive Decision Trees. The coloured bars represent the distribution of dominant input uncertainties within the leaves nodes of the statistically optimal (top) and interactive (bottom) Decision Trees. Below each bar, the first number indicates how many CRESTA zones fall into that leaf. The second number indicates how many of the CRESTA zones in that leaf are associated with the dominant input uncertainty prevailing in that leaf. The third number is the percentage of CRESTA zones associated with that dominant input uncertainty with respect to the total over the entire Rhine River basin.*

The bottom part of Figure 4.6 shows that leave nodes of the interactive DT (iDT) obtained using the toolbox presented in Chapter 2 (the full DT is shown in the Appendix C4 in Figure C.3). Using the statistically optimal DT as a starting point I obtained the iDT in an interactive "trial and error" approach: I was manually expanding or pruning tree branches either by changing splitting variables and thresholds, and/or changing the number of leaf nodes. I stopped manual interactions with the DT when I got a model that represented all dominant input uncertainties in its leaf nodes. It is obvious that, in general, the purity of the iDT nodes is higher than in the statistically optimal DT. Also, in the case of iDT, out of 17 leaf nodes, the 11 that are meaningful (i.e., neither blanked or "undefined") span over all four input uncertainties. In fact, besides the 4 leaf nodes where Event set or Damage ratios are dominant, there are another 5 nodes representative of places where either hazard maps (brown) or residential buildings' value (orange) are the dominant input uncertainty. In absolute numbers, these leaf nodes may contain few CRESTA zones, but in relative terms these few CRESTA zones explain a significant percentage of the total variability for these dominant input uncertainties across the river basin. This is important because it means that further analysis of the iDT structure (presented in the next subsection) will help determine the characteristics that explain where any of the four input uncertainties is dominant. Last, it is worth noting that these useful results yielded by the iDT is not obtained at the expenses of overall classification accuracy. Actually, the iDT is slightly more accurate than the SOT in both the training and test sets - and this is obtained with the same tree depth 8 layers in both cases).

4.3.4 Topography, degree of urbanization and residential buildings' value are key characteristics to explain the spatial variability of dominant input uncertainties.

Figure 4.7 shows the full iDT whose leaves were reported in Figure 4.6. For the sake of simplicity, the inconclusive leaves (those labelled as "undefined" in Figure 4.6. and those that explain less than 5% of the total variability for a dominant input uncertainty) are not reported in Figure 4.7. The earlier a variable appears in the tree structure, the more important it is for classifying the dominant input uncertainties. Hence, the key characteristics are the river section/sub catchment (*HydBas*), the mean value of the residential buildings at a region (*MeanVal*) and the degree of urbanization (*ARTF*) because they dominate the DT at the top three levels. Other useful variables are the river length and the statistics of flood depths because they dominate the tree in the lower levels. The percentage noted at each leaf is the third number in Figure 4.6. In Table 4.3 I have summarized the conclusions drawn from the iDT as to what system characteristics lead to each input uncertainty. Each summary

statement in the table is labelled with a number which points to the respective leaves in the iDT of Figure 4.7.

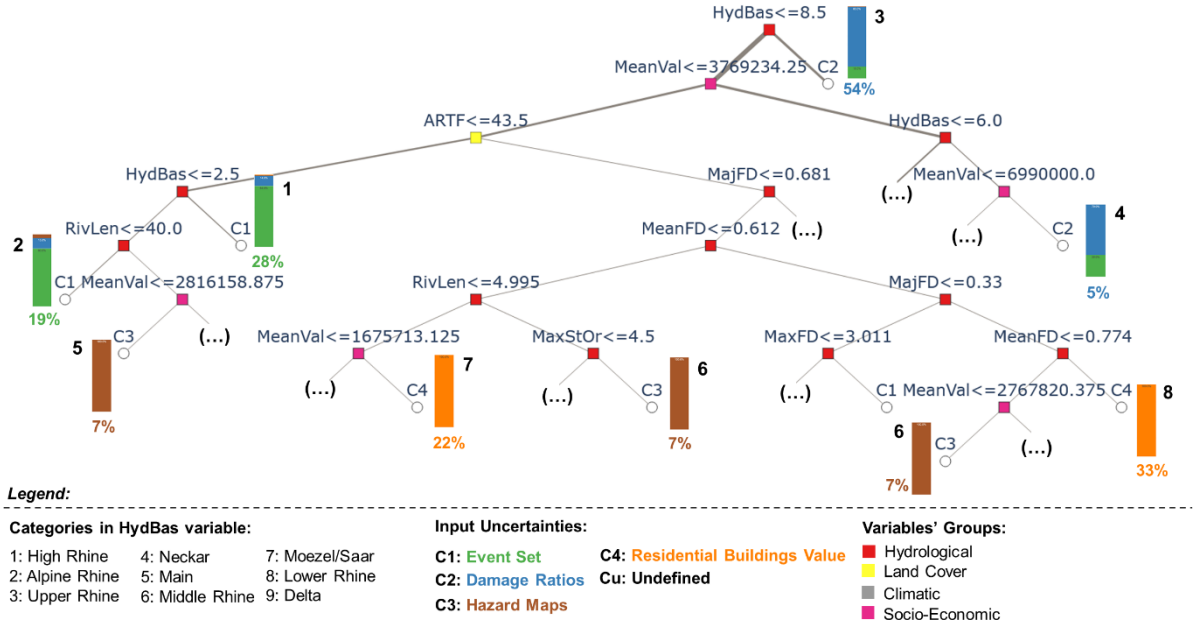


Figure 4. 7 *Interactive Decision Tree. The leaf nodes are assigned with a percentage which refers to the third number in figure 4.6 and expresses the fraction it represents with respect to the total cases of the corresponding dominant input uncertainty. The numbers in black bold letters point to the corresponding summary in Table 4.3*

Table 4. 3 *Summary of the characteristics of the places where each input uncertainty is likely to dominate the flood losses uncertainty. The terms High, Alpine, Upper, Middle. Lower Rhine, Neckar, Main, Moezel/Saar and Delta refer to the Rhine River sections and sub-catchments shown in Figure 3.1.*

Event Set	Damage Ratios	Hazard Maps	Residential buildings' Value
1. In Upper, Middle & Lower Rhine, and Neckar, Main, Moezel/Saar with low residential buildings values and urbanized areas **2.** In High and Alpine Rhine, with low residential building values, urbanized areas and medium to small river lengths	**3.** In Delta River section **4.** In Moezel/Saar and Lower Rhine with medium to high residential buildings values	**5.** In High and Alpine Rhine, with low residential building values, urbanized areas and large river lengths **6.** In all regions but delta, with low residential building values, highly urbanized areas, small flood depths, large range of river lengths and low slopes.	**7.** In all regions but delta, with low residential building values, highly urbanized areas with small flood depths and river lengths. **8.** In all regions but delta, with low residential building values, highly urbanized areas and medium flood depths

Topography seems to be a key factor in differentiating the dominant input uncertainties. The leaf nodes on the right-hand side of the iDT (labelled with number 3 and 4 in Fig. 4.7) show that more than half of the CRESTA zones where damage ratios uncertainty is dominant belong to regions with flat terrains, i.e. either the Delta (node 3 with HydBas = 9) or the

Lower Rhine and Moezel/Saar (HydBas =7,8). In these regions the water from a flood event, unaffected by cross sectional changes, is free to extend further, thus leading to large flood extents with low variability in flood depths (Merwade et al., 2008). In such cases any changes to flood depths are likely to lead to similar flood depths that cause similar damage. Hence, uncertainty in damage ratios emerges as the dominant input uncertainty. On the other hand, iDT classifies 19% of the CRESTA zones where Event Set uncertainty is dominant in regions with steep, e.g., High and Alpine Rhine (node 2 with Hydbas =1,2) and urbanized terrains. In these areas, water is confined in slopes (or from artificial surfaces), leading to lower flood extents but with greater variability in flood depths. In such cases, changes in flood depths are likely to lead to different damages. Hence, uncertainty in the values related to more localized features (e.g. flood depths estimation) emerge as the dominant input uncertainty. This conclusion is confirmed by the scatterplots in Figure 4.8. The scatterplots show the mean terrain slope (1st row) and the variance of flood depths (2nd row) in CRESTA zones that are dominated by Event Set (left column) and Damage ratios (right column) uncertainty. These CRESTA zones correspond to cases 1,2 and 3,4 in Table 4.3. Where Damage ratios are the dominant input uncertainty (right column), the mean slope and flood depths variance in CRESTA zones are significantly lower than where Event Set is dominant. Looking at the mean slope scatterplot for Damage ratios (top right panel), the CRESTA zones with higher mean slope (highlighted by the red rectangle) are those belonging to the Moezel/Saar basin, which has bit higher slopes in comparison with the Delta and Lower Rhine River sections where the rest CRESTA zones belong to.

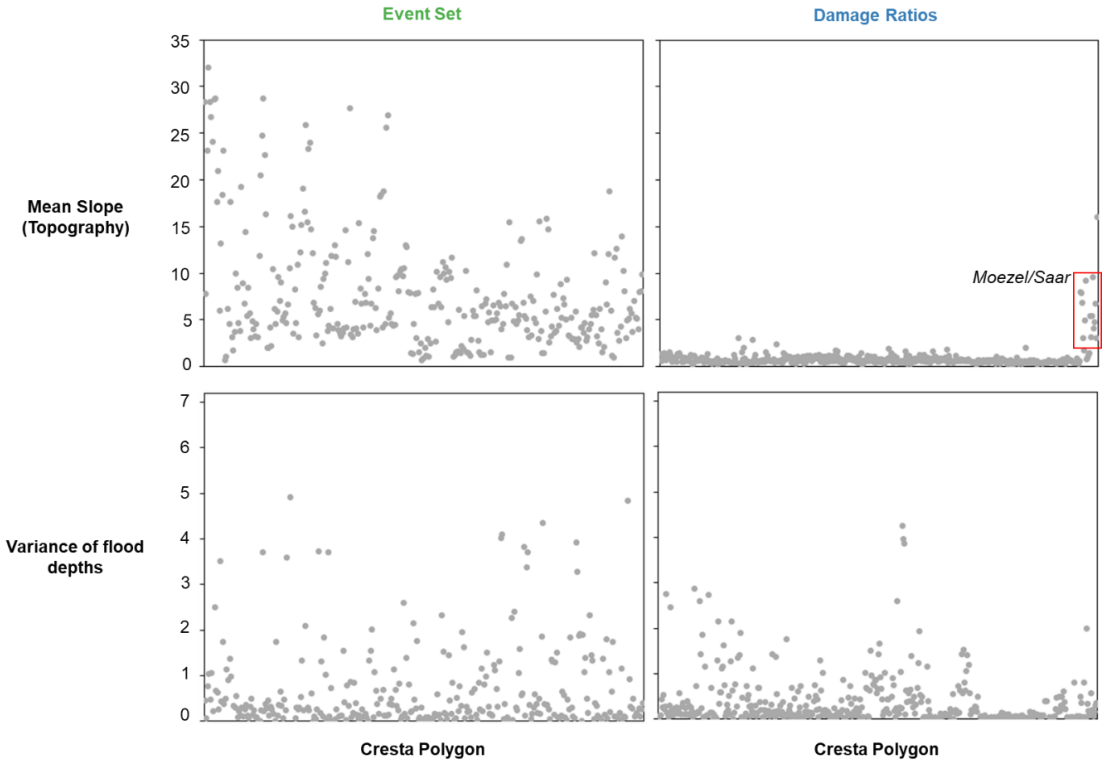


Figure 4. 8 *Scatterplots of the mean slope (1st row) and variance of flood depths (2nd row) for CRESTA zones that are dominated by Event Set (Left column) and damage ratios (Right Column) uncertainty. The selected CRESTA zones correspond to cases 1,2 and 3,4 in Table 4.3.*

Moreover, as expected the residential buildings' value is dominant in highly urbanized places (nodes 7,8 with ARTF>43.5) where a greater number of exposed assets is likely to be located. Additionally, one can see that some nodes on the branches leading to CRESTA Zones with residential buildings value as dominant uncertainty (orange leaves) make splits on flood depths variables. The splitting thresholds indicate that relatively small flood depths are associated with such places (node 7). This is explained by the fact that the variability range for those depths is smaller in comparison with larger flood depths.

According to the methodology, when perturbing the hazard maps both flood depths and extends are changing. Hence, uncertainty in flood hazard maps is likely to dominate in a larger range of regions. Indeed, based on the iDT Hazard Maps uncertainty can dominate in steep terrains like High and Alpine Rhine (node 5, with Hydbas =1,2) with large river lengths and/or in flatter terrains (node 6, with MaxStrOrd>4.5) with relatively small flood depths and a large range of river lengths.

4.4 Discussion

4.4.1 Uncertainty in vulnerability component significantly impacts flood loss estimation

Sensitivity analysis of average annual losses (AAL) highlights the importance of the vulnerability component on flood loss estimation. It is sensible because vulnerability curves are the means to translate the physical hazard to flood loss. Hence, even if we assume that the flood depths and residential buildings' value estimation is perfectly accurate if there is significant uncertainty in the vulnerability curves then the losses will ultimately be over or under-estimated. This result aligns with many previous studies. Wing et al. (2020) argues that even for very accurate hydraulic models damage estimates can be poor due to the use of vulnerability curves. Freni et al. (2010) demonstrated that using detailed hydraulic models will not necessarily improve flood damage estimates due to significant uncertainty in vulnerability curves which they see as the bottleneck in flood loss estimation. Thus, they suggest that efforts on uncertainty reduction should first be focused on the improvement of the vulnerability curves. The Lighthill Risk Network (2019) report ranked the vulnerability curves and the potential bias created by damage ratios as top priority challenge to be addressed in the flood risk modelling industry among a list of flood model topics. De Moel & Aerts (2011), in a case study in Netherlands also concluded that the uncertainty in the estimated AAL is driven by the uncertainties in the vulnerability component.

Besides the damage ratio values, the results show that the shape of the function, that is the steps in the function, has a great impact on flood loss estimation too. Every change in the value of damage ratios affects losses but this is not true for the flood depths. Flood depths lead to different losses only if the change in their value is combined with a change in the step of the vulnerability function. If not, the damage ratio will be the same and ultimately the losses too. Hence, when using step functions the resolution/width of the steps needs to be carefully determined. For example, if the steps are too wide, the effects of uncertainty in flood depths can be dampened and potentially lead to biased conclusions.

Sensitivity analysis results of the Loss Exceedance Curves (LECs) show that dominant uncertainties change with the chosen return period. Uncertainty in the event set, in the residential buildings' value and in the hazard maps are more likely to control LECs at low return periods, i.e. when losses are driven by small and frequent flood events. Damage ratios uncertainty instead dominates the LEC uncertainty at large return periods when extreme flood events are important. These results are in line with other findings from the literature. Kaczmarska et al., (2018) also concluded that the impact of the vulnerability component on flood losses increases as the return period of flood losses increases.

Moreover, the results for the low return period LEC, agree with most of the local scale studies: De Moel et al. (2014), in the Delta section and Apel et al. (2004) & (2008), and Merz & Thieken (2009), in the lower rhine section. Except Winter et al. (2018) in the Alpine Region where they concluded that the vulnerability component and not the flood events uncertainties dominate the LEC uncertainty. The differences might be explained by methodological differences existing between the two studies. Firstly, Winter et al. (2018) used different rules (e.g., time interval of peak discharge above a certain threshold) for the event definition. Another reason could be the different extreme value model used to estimate the return period of the events. Finally, Winter et al. (2018) generated flood events based on 1,000 years simulation period, while in this case study the events are generated based on 10,000 years simulation period. This means that in this case study a much larger range of flood events is considered. The length of the simulation period is deemed crucial for flood risk analysis (Kaczmarska et al., 2018).

4.4.2 The importance of investigating dominant uncertainties across large spatial domains

Applying sensitivity analysis to a large-scale FRM allowed us to gain new insights regarding the dominant controls of flood losses. To the best of my knowledge this is the first study to provide evidence on how the different system characteristics (hydrological, socio-economic, land cover) can explain (to some degree at least) why certain input uncertainties dominate in different places. Topography seems to be a key factor to differentiate flat places (e.g., deltas) where LEC uncertainty is dominated by uncertainties in the vulnerability component, from steep regions where hazard uncertainty (Hazard Maps and Event Set) is dominant. This is also consistent with previous studies. Devitt et al. (2021) investigated the spatial variability in the uncertainty of 100-year flood events' magnitude estimates of large-scale flood hazard models across the USA. They found that topographic characteristics (such as elevation and mean per cent slope) has a strong influence on the spatial variability of the predictive performance of flood magnitude estimates. The authors identified contrasting patterns in the prediction performance of models for flat terrains (e.g. deltas) versus high, steep and complex terrains (e.g. mountainous areas). Lindersson et al. (2021) investigated the model agreement of three different global flood hazard models across variable geographic conditions (e.g., topography). They found poor inter-model agreement in steep and flat regions. As expected, the degree of urbanization is a key factor for places where the residential buildings' value uncertainty dominates the losses. De Moel & Aerts (2011) investigated the effect of different flood risk components uncertainties to the overall uncertainty in flood risk estimates within a dike ring in Netherlands. They also concluded that

uncertainties in residential buildings' value are especially important for the urbanized areas which also accounted for a significant part of the study case's loss. Other hydrological characteristics like river length and the statistics of flood depths seem to be more relevant for input uncertainties that are associated with more localized features (event sets and residential buildings' value)

4.4.3 The importance of using scientifically informed data-based methods to analyse large datasets

In flood risk applications, I often have to analyse imbalanced datasets where one or more variables are under-represented. There are two problems with models built on unbalanced datasets. Firstly, they are poorly generalizable on unseen data (Faghmous & Kumar, 2014). Second, if the interest is on the under-represented class, the model might ignore it given that fitting the small number of data does not influence the overall performance statistic, hence rendering the model less informative. Both problems are highlighted in this case study. The statistically optimal Decision Tree (DT) performed poorly on unseen (test set) data because the DT model did not include the under-represented dominant input uncertainties. The algorithm would treat them as statistically insignificant due to their low representation in the dataset. Hence, the resulting model was not very informative as my efforts to understand which local characteristics are linked with these dominant input uncertainties were compromised.

Common approaches for dealing with imbalanced dataset were mentioned in Chapter 2 and some of them have been used in flood risk applications too. For example, Park & Lee (2020) used a resampling technique to deal with imbalanced datasets in a coastal flood risk application. However, these approaches can often be time consuming (Zhou et al., 2017). Here I showed that one can interactively build a DT model to overcome the challenge of imbalanced datasets.

4.4.4 Suggestions for improving input uncertainty characterization

The choices made on the characterization of damage ratio and return period uncertainty may partly explain why they dominated the loss uncertainty in most places within the study domain. In Sec. 4.4.1 I explained that the use of a step function for the vulnerability curve may dampen the effect of varying flood depths, and in Sec. 4.2.3 I showed that the +-50% homogeneous perturbation overestimates the variability range for small and medium return periods.

In the context of this application, I believe there are two directions for further research that are particularly important to investigate. Firstly, the use of a smooth vulnerability curve instead of a step function could potentially lead to a greater variability in the dominant input uncertainties and thus allow to explore further links with local characteristics. Second, it is worth investigating the use of a perturbation range that increases with return period. This could potentially lead to greater variability in the dominant input uncertainties for losses with large return period. These losses are dominated by extreme flood events and therefore it is possible that the role of uncertainty in the characterisation of these extreme events is underestimated with the current approach.

My choice to perturb the return period of the hazard maps and events in the event set, was driven by the computationally expensive nature of the hazard component. The assessment of uncertainty all along the modelling chain requires the implementation of a comprehensive uncertainty quantification scheme (Deroche, 2023). The hazard component is a bottleneck to this effort because running multiple simulations under different parameterizations is computationally expensive and therefore not always an option. An idea would be to use the hazard component (e.g., flood maps) from the various available flood risk models to assess the uncertainty in the hazard component. The challenge here is that some of them are proprietary and so not easily accessible. It's currently unclear how to best represent the uncertainty in the hazard component and this is a topic of interest beyond the context of this application.

4.5 Conclusions

Uncertainty is inherent in every flood risk model. One way to increase the value of flood risk estimations is to better understand the dominant controls in flood risk estimations, and how and why they vary in space. I answered these questions for the heterogeneous Rhine River basin by presenting a GSA application on a flood risk model. First, I identified the dominant input uncertainties (among uncertainty in the flood depth estimates, vulnerability curves and exposure dataset) within the spatial domain on two key model outputs, the Average Annual Loss (AAL) and the Loss Exceedance Curve (LEC). Then, using interactive Decision Trees I linked the dominant input uncertainties in each spatial unit within the domain with its hydrological, climatic, and socio-economic properties.

The methodological approach and results presented in this chapter are useful for model development and to help prioritize efforts for uncertainty reduction in flood risk modelling. They can be used by (re)insurers and government agencies contributing to more informed

decisions. For example, AAL and LEC are standard metrics that are used for risk portfolio management and/or solvency calculations. The results indicate that to reduce the uncertainty in those metrics estimation, efforts should focus on the vulnerability component. Moreover, for tasks like underwriting, pricing and exposure management which are performed at post-code level the results provide evidence on which input uncertainty modellers should focus. For example, it seems that for flat terrains vulnerability component uncertainties strongly influence the uncertainty in flood loss estimations while the hazard and exposure component uncertainties are more likely to be influential in steep and urbanized places.

5 Conclusions and future research

5.1 Chapter summaries

In this Thesis, I aimed at improving our understanding of uncertainty quantification and attribution in flood risk modelling. My efforts were guided by three main research questions.

How can we incorporate scientific knowledge into data-based methods?

In flood risk modelling (and in geosciences in general) there is often the need to analyse large and complex datasets. Machine Learning offer promising tools such as Decision Trees, that can support this effort. Decision Trees have been widely used in the geosciences to automatically extract patterns from complex and high dimensional data. However, like any data-based method, the application of decision trees is hindered by data limitations, such as significant biases, leading to potentially physically unrealistic results. In Chapter 2 I developed interactive DT (iDT) that put humans in the loop to integrate the power of experts' scientific knowledge with the power of the algorithms to automatically learn patterns from large datasets. I created an open-source Python toolbox that implements the iDT framework. Users can interactively create new composite variables, change the variable and threshold to split, prune and group variables based on their physical meaning. I demonstrate with three case studies how iDT overcomes problems with current DT thus achieving higher interpretability and robustness of the result.

How can we define the input variability space when we are too uncertain about our inputs?

To quantify the uncertainty in flood risk estimates one needs first to characterize the input variability space. This requires the modeller to make subjective statements on what values are possible and not. In mathematical terms these choices are represented by the use of appropriate probability distribution functions or a list of possible values for the input uncertainties. However, in the context of flood risk modelling, this is challenging because input uncertainties are poorly known and constrained, and the quantity and quality of available data that could be used to infer probability distribution functions are often poor. In Chapter 3 I develop an approach to characterize input uncertainties in cases where we are too uncertain about the input uncertainties. I suggest using uniform distribution functions and defining plausible variability ranges for the input uncertainties through a systematic literature review approach. I demonstrated this approach by defining variability ranges for the residential buildings' value, the damage ratios of the vulnerability curves for the Rhine River

basin and the return period of flood events for two gauging stations. The variability ranges were then used in a sensitivity analysis application. Besides, these quantities are used in every flood risk analysis. The approach is transferable to similar applications and areas.

How the importance of input uncertainties changes in places with different hydrology, climatology, land-cover, and economy?

In flood risk modelling it is essential to know which uncertainty sources mostly control risk estimates so to guide efforts for model improvement, as well as to help risk managers make better decisions. Past efforts to attribute the output uncertainty of flood risk models have reached conflicting conclusions. This may be because these studies used different risk models and different uncertainty and sensitivity analysis approaches; or, that they were conducted at relatively small (catchment and/or city) scale, in places with different climatic, hydrological, and socio-economic characteristics. In Chapter 4, I investigated dominant uncertainties of a flood risk model across a much larger scale, namely the entire Rhine River basin, and explored whether dominant uncertainties at specific places can be linked to their physical or socio-economic characteristics. In particular, I analysed two model outputs: the Average Annual Losses (AAL) and Loss Exceedance Curves (LECs). For each output, I first identified the dominant input uncertainties (among uncertainty in the flood depth estimates, vulnerability curves and exposure dataset) in each spatial unit of the modelled domain; and second, I linked those dominant input uncertainties to the characteristics of the spatial units. I find that uncertainties in the vulnerability component dominate the AAL. The dominant uncertainties for the LECs change with the return period of loss, with vulnerability becoming increasingly important with increasing return period. I used interactive decision trees to link the dominant input uncertainties to hydro-climatic and socio-economic characteristics of the places. Using this method, it was possible to extract useful insights for the dominant input uncertainties in the datasets even for the under-represented ones. Topography (flat versus steep terrains), degree of urbanization and economic value of the buildings are key characteristics for determining how dominant uncertainties change spatially within the study domain.

5.2 Thesis Contributions

In this section I summarize the three main contributions of this thesis.

1. I have presented a methodology to perform Global Sensitivity Analysis (GSA) of a complex flood risk model at large domains. By applying this methodology to the Rhine River Basin, I showed that despite the computationally expensive nature of

GSA it is possible to propagate uncertainties across large domains, identify which input uncertainty dominates the uncertainty in risk estimates at different places and learn some general lessons on what controls the spatial variability of dominant input uncertainties.

2. Moreover, I have shown that in cases with little knowledge or data availability it is possible to characterize input uncertainties by collecting and combining information from the literature on how past studies have characterized relevant input uncertainties.
3. I developed a method that establishes an interactive approach between experts and the algorithm for building ML models. With this approach the experts can gain a deeper understanding of the models they build and the cases they investigate, and they can constrain the process of knowledge discovery with their domain knowledge.

5.3 Overarching remarks

In this Thesis I faced cross-cutting issues which I tackled and presented in separate chapters and here I discuss how they are connected.

The imbalanced datasets were the first cross-cutting issue tackled in this Thesis. We frequently encounter such datasets in our field because the phenomena we are studying are usually controlled by only few processes (Wagener & Pianosi, 2019). Indeed, the sensitivity analysis application presented in Chapter 4 confirms the above point. The dominant controls of flood loss within Rhine River basin resulted in an imbalanced dataset.

Finding patterns in imbalanced datasets using current automated ML proved to be a daunting task. My attempts in Chapter 4 to identify links between local characteristics and dominant input uncertainties with a conventional ML approach (Decision Trees) were unsatisfactory because the resulting models were missing the under-represented dominant input uncertainties due to their low numbers in the dataset. In fact, the automated algorithm built the Decision Tree by expanding only the branches that improved the overall accuracy based on statistical metrics, which meant under-represented classes did not appear in the DT. But these minor classes were physically important to us. Even though it is a well-known problem in geosciences (Oommen et al., 2011) not much progress has been made in developing methods that tackle this problem. Existing approaches were discussed in Chapter 2 but most of them are time consuming (Zhou et al., 2017). A lesson I learned is that interaction between humans and ML algorithms discussed in Chapter 2 can significantly improve and ease the process of finding patterns in an imbalanced dataset. As it was

unknown to us a priori what the characteristics linked to the under-represented dominant uncertainties could be, it was important to use a tool to explore this. The interactive DT toolbox was very handy to us because it established an expert-algorithm iterative learning process. With us attempting changes (e.g., through manual pruning and changing of splitting variables and thresholds) and the algorithm updating on the fly I was able to guide the algorithm to expand branches that were physically (and not statistically) sensible to us, that is, those that I believed they could contain the underrepresented classes. Thus, with just few changes I was able to end up with a model that represented all classes and not just the major ones.

The second cross cutting issue discussed in this Thesis was how to create a comprehensive scheme to investigate the dominant controls of flood losses uncertainty. This was tackled in Chapters 3 and 4 and I found particularly challenging to identify, select and quantify the input uncertainties (Chapter 3) for the global sensitivity analysis of the chosen flood risk model (Chapter 4). From my interactions with my industry collaborators at JBA Risk Management and experts at GFZ Potsdam, I realized that even though it is known there is considerable uncertainty in every risk component it is unclear how this could be best translated in mathematical terms for sensitivity analysis applications, and there is not an established methodology for this step. Moreover, the available information in the literature was fragmented. I thus developed an approach that can build on this fragmented information (e.g., information available on different places and scales) from various studies to propose variability ranges of selected input uncertainties for the entire Rhine River Basin. It may not always be possible to find enough information in the literature, but this is a first step in overcoming this daunting and complex problem. In Chapter 4 I used the ranges I found in Chapter 3 to perform a sensitivity analysis and identify the dominant controls of flood losses uncertainty.

5.4 Future Research

In this Thesis I have looked at how and why uncertainty in flood risk modelling changes in space. I investigated dominant input uncertainties in a large and heterogeneous spatial domain and presented an approach to explain their variability by linking it to the characteristics of the places. The results are useful to better understand the spatial patterns of flood risk uncertainty. But one limitation worth discussing is that the methodological approach presented in this Thesis doesn't consider the temporal aspect of risk. It implicitly assumes that hazard, exposure and vulnerability remain stable through time and so the same is true for flood risk estimations and the associated uncertainty.

However, the three flood risk components (hazard, exposure, vulnerability) are all changing in time: they have changed over the past and will likely be subject to further changes in the future (Duha Metin et al., 2018, Kreibich et al., 2017). There are several anthropogenic or natural causes for these changes, e.g., climate, land use, flood plain changes, socio-economic trends, flood prevention management policies (Schanze, 2012).

For example, Merz & Thieken (2009) reported that flood hazard along the Rhine has changed due to river training networks, retention measures, the construction of weirs. Slater & Villarini (2016) studied flood hazard trends in the United States and found increases in flood hazard in the Midwest and decreases in the Gulf coastal plain, south-eastern United States and California. Their results suggest that these patterns cannot be explained by the large-scale spatial precipitation patterns alone, but also from more localized factors like shifts in basin wetness that occur at the land surface and subsurface over annual time scales (due to climate and human activity). Blöschl et al. (2017) studied the impacts of climate change on the timing of European floods and found temporal and spatial patterns (e.g, earlier than usual spring snowmelt floods due to increased temperatures throughout north-eastern Europe, Delayed winter floods around the North Sea and in some places in Mediterranean coast due to delayed winter storms associated with polar warming, Earlier than usual winter floods in western Europe which are happening due to earlier soil moisture maxima.

Exposure and vulnerability are also changing. Europe has experienced significant increases in population (>130%) and wealth (>2000%) during the period 1870-2016 and thus increasing trends in the number of people affected by floods and related economic losses (Paprotny et al., 2018). Jongman et al. (2012b) investigated global exposure to river and coastal flooding using population and a land-use based methods and found increasing trends in population and wealth exposure to floods between 1970-2010. Jongman et al. (2015) found declining vulnerability (expressed in terms of mortality and loss rates) in river floods between 1980 and 2010. This pattern is consistent in low- and high-income countries and associated with increasing gross domestic product. Tanoue et al. (2016) studied long term trends in vulnerability (expressed in terms of mortality and loss rates) at global scale and found consistent decreasing trends (in loss rates) for high income countries (classified in terms of gross domestic product per capita) and an inverted U-shape pattern for low- and middle-income countries. The U-shape pattern reveals that at first there is an increase in risk because the effect of adaptation strategies is slower than the increase in assets. Then risk decreases once the effect of adaptation is more pronounced.

Additionally, the three risk components interact in multiple ways leading to non-linear effects on flood risk. Di Baldassarre et al. (2018) describes how urban development close to the river in protected areas can lead to non-linear increases in flood risk. The presence of structural flood protection tends to create incentives to build closer to the river and therefore increases flood exposure. The presence of structural flood protection creates a sense of safety reducing the awareness and preparedness to floods increasing vulnerability. This phenomenon known as the “levee effect” can offset part of the intended benefits of structural flood protection and, paradoxically, flood risk can even increase in the medium–long term after the introduction or reinforcement of a structural flood protection. Another example of a non-linear effect of structural adaptation are the comparable (in terms of hazard) 2002 and 2013 floods in Elbe River in Germany. After the 2002 flood event, Saxony decided to adapt to floods by investing on structural flood defences. The installed defences in Saxony withstood the hydraulic load in the 2013 flood event by routing it downstream to Saxony-Anhalt where increased levee failures were recorded in comparison with the 2002 flood event. This resulted in increased flood loss (risk) for this area fuelling a public debate on the benefits of structural adaptation policies (Vorogushyn et al., 2018).

Investigating such temporal changes and interactions is important to understand how the different flood risk modelling components affect flood risk. Changes in the three risk components will continue to occur in the future in ways that are impossible to predict and thus associated with significant uncertainty. For example, the negative impacts of climate change are escalating at a faster rate than scientists predicted few years ago, according to the latest report from a United Nations climate panel (IPCC 2022, Tollefson, 2022). Improving our understanding of the impacts of climate change on the frequency and severity of flood losses for different return periods and in different regions of the world is a top concern for governments, regulators and the (re)insurance industry and the public (Lighthill Risk Network, 2019, AXA Future Risks Report, 2022, Mitchell-Wallace, 2022, Ellison, 2023).

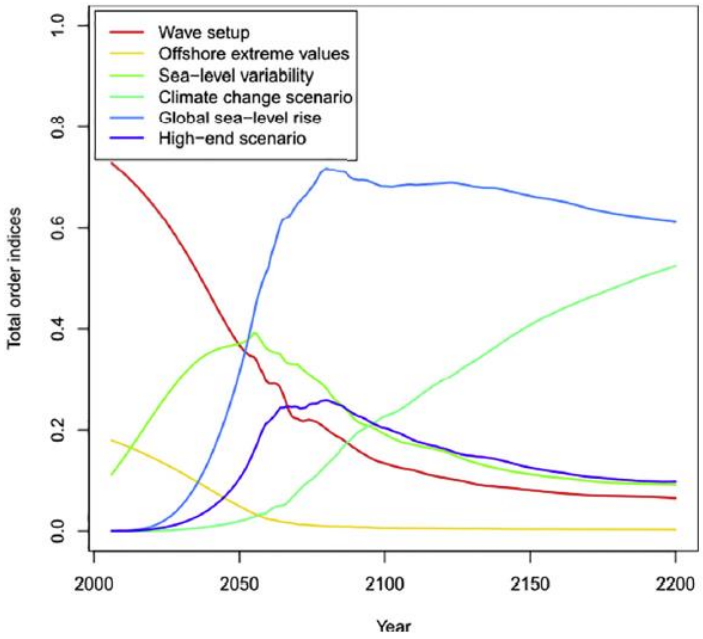


Figure 5. 1 *Evolution of sensitivity indexes over the time, for five input uncertainties related to coastal defence vulnerability. Source: Le Cozannet et al., (2015)*

An important question for future research thus is: How do the drivers of flood risk change in time? A time-varying version of the uncertainty and sensitivity analysis techniques used in this Thesis, could help investigate such question (Wagener & Pianosi 2019). For example, Le Cozannet et al. (2015) studied the changing frequency of coastal flood events. They presented an application of time varying global sensitivity analysis to quantitatively assess the relative importance of various input uncertainties over the coming decades. The following figure (retrieved from the paper) shows how the sensitivity of predicted coastal defence vulnerability (namely the yearly probability of exceeding the threshold height of coastal defences) varies with time. The variation in dominant input uncertainties over time is significant. For example, global climate change scenario starts dominating after 2070 while offshore extreme values are not influential after then. Up to 2050 the dominant factor is the 'wave set-up' parameter, which accounts for sea level rise induced by wave breaking. This local process is determined by the near-shore coastal bathymetry and is often neglected in coastal hazard assessments studies. This type of analysis helps revealing the changing effect that certain processes can have over time. Failing to incorporate such patterns may lead to invalid conclusions and to an overestimation of the effects of input uncertainties at short, mid and/or long-term period.

Future research should explore the applicability of a similar approach to flood risk modelling in order to investigate the temporal dynamics of uncertainty in flood risk estimation.

Appendix A – Supporting information for Chapter 2

A1 Decision Tree Algorithms and CART algorithm

Decision trees fall in the category of non-parametric supervised learning methods. There are various algorithms developed and applied over the years. Classification and Regression Trees (CART) proposed by Breiman et al. (1984) is of particular interest in Chapter 2 because it is used as a basis for the construction of the DT model in the interactive methods. CART algorithm is used to build models for non-linear predictions from a dataset.

Description of CART algorithm:

The DT structure described in the Introduction section of Chapter 2 is constructed in a top-down fashion based on a greedy algorithm that searches through all the possible combinations to find the optimal split. This is the variable and threshold that will lead to pure nodes. The two most popular criteria that have been used for this purpose and the ones implemented here are information gain and gini index. CART algorithm allows trees to fully grow and then a pruning technique is applied. It uses 10-fold cross validation to estimate error rates and the tree is reduced to the size that has the lowest cross-validation error estimate. CART can handle missing values using surrogate splits (Breiman et al. 1984; Loh, 2014).

A1.1 Mathematical Formulation:

Given vectors of independent variables xi ∈ R^n, i=1, …, l and a label vector (which I call classes) y ∈ R^l.

If Q are the data at node m then for each candidate split θ=(X_i, $\bar{X}_{i,m}$) where X_i represents the variable and $\bar{X}_{i,m}$ the threshold, one can partition the dataset into $Q_{left}(\theta)$ and $Q_{right}(\theta)$ subsets:

$$\begin{aligned} Q_{left}(\theta) &= (x,y) | X_i \leq \bar{X}_{i,m} \\ Q_{right}(\theta) &= Q - Q_{left}(\theta) \end{aligned} \qquad \text{(A.1)}$$

The impurity at node m can be estimated using an impurity function I.

$$G(Q,\theta) = \frac{n_{left}}{N_m} I\left(Q_{left}(\theta)\right) + \frac{n_{right}}{N_m} I\left(Q_{right}(\theta)\right) \qquad \text{(A.2)}$$

The variable and threshold that minimises the impurity will be selected:

$$\theta^* = argmin_\theta G(Q, \theta) \qquad \text{(A.3)}$$

Then, there is a recursive partitioning of the subsets: $Q_{left}(\theta^*)$ and $Q_{right}(\theta^*)$ until the maximum allowable depth is reached, $N_m < min_{samples}$ or $N_m = 1$.

A1.2 Impurity criteria

The nodes are split in a way that the new nodes will be as pure (homogeneous) as possible. Consequently, it is a usual practice to adopt a metric that will measure the impurity of the nodes (e.g., how many values are of the same class or not). A general equation to compute the impurity is shown below:

$$IG(D_p, f) = I(D_p) - \sum_{j=1}^{m} \frac{N_j}{N_p} I(D_j) \qquad \text{(A.4)}$$

Where: F is the variable to perform the split

Dp dataset of the parent node

Dj dataset of the jth child node

I is the impurity measure

Np is the total number of samples in the parent node

Nj is the total number of samples in the jth child node

Therefore, the optimal split is for the variable and corresponding threshold that maximizes the above metric.

Commonly used impurity metrics are entropy, gini index and minority class. Here only the entropy and gini index will be explained as they are used for the construction of the trees.

a. *Gini Index*:

$$\mathrm{I_G(t)} = 1 - \sum_{\mathrm{i=1}}^{\mathrm{m}} \mathrm{p(i|t)}^2 \qquad \text{(A.5)}$$

Where: m is the number of classes

p(i/ t)is the fraction of data points in the node t belonging to category i.

If Gini=0 then the node is pure and the samples in that node are perfectly separated and belong only to one class. If Gini=0.5 the node is impure and the samples are misclassified.

b. *Entropy:*

$$\mathrm{I_H(t)} = -\sum_{\mathrm{j=1}}^{\mathrm{m}} \mathrm{p(i|t)} \log_2 \mathrm{p(i|t)} \qquad \text{(A.6)}$$

Where: m the number of classes

p(i|t) is the proportion of the samples that belong to class m for a node t.

If Entropy=0 the node is pure and all samples of the node belong to the same class. If Entropy is maximal the nodes is impure and the samples are misclassified.

A1.3 Cost Complexity Pruning

Breiman et al. (1984) proposes a pruning algorithm called minimal cost-complexity to prune a tree and avoid overfitting. The basic parameter of this algorithm is the cost-complexity parameter a≥ 0. This parameter is used to define the cost complexity measure, Ra(T) of a given tree T:

$$R_a(T) = R(T) + a|T| \qquad \text{(A.7)}$$

Where $|T|$ is the number of terminal nodes in T

$R(T)$ is normally the misclassification rate in the terminal nodes. However, scikit-learn uses for $R(T)$ weighted impurity of the terminal nodes instead.

Minimal cost-complexity pruning finds the subtree of T that minimizes $R_a(T)$. Cost-complexity for a single node is defined as $R_a(T) = R(T) + a$. T_t is a branch with root node t. The impurity of a node is greater than the sum of impurities of its terminal nodes. However, the cost complexity measure of a node, t, and its branch, T_t, can be equal depending on a. As effective cost complexity parameter a_{eff} of a node t is defined the value of a where $R_a(T_t) = R_a(T)$ or $a_{eff}(t) = \frac{R(t) - R(T_t)}{|T| - 1}$. The non-terminal node with the smallest value of a_{eff} is the weakest link should be pruned.

A2 Graphical User Interface of the web-based tools for interactive construction and analysis of decision trees.

The Graphical User Interface (GUI) of the tool (Figures A.1) is comprised of three tabs:

1. The first tab is named “Pre-processing Stage” and contains widgets (e.g. text boxes, dropdown menus etc) for:
    a. Defining classes,
    b. Grouping the available features (and assign colours to the groups) and
    c. Selecting important features.
2. The second tab is named “Interactive Decision Tree” and contains widgets to enable the expert to:
    a. Control the tree size and structure.
    b. Create new variables and feed it back to the algorithm.
    c. Manually change variables and thresholds to split
    d. Manually prune nodes and/or tree branches
    e. Manually change leaf nodes classes
3. The third tab is named “Evaluation Metrics and Plots” and contains widgets to enable the expert to:
    a. Calculate the classification accuracy
    b. Plot confusion matrices

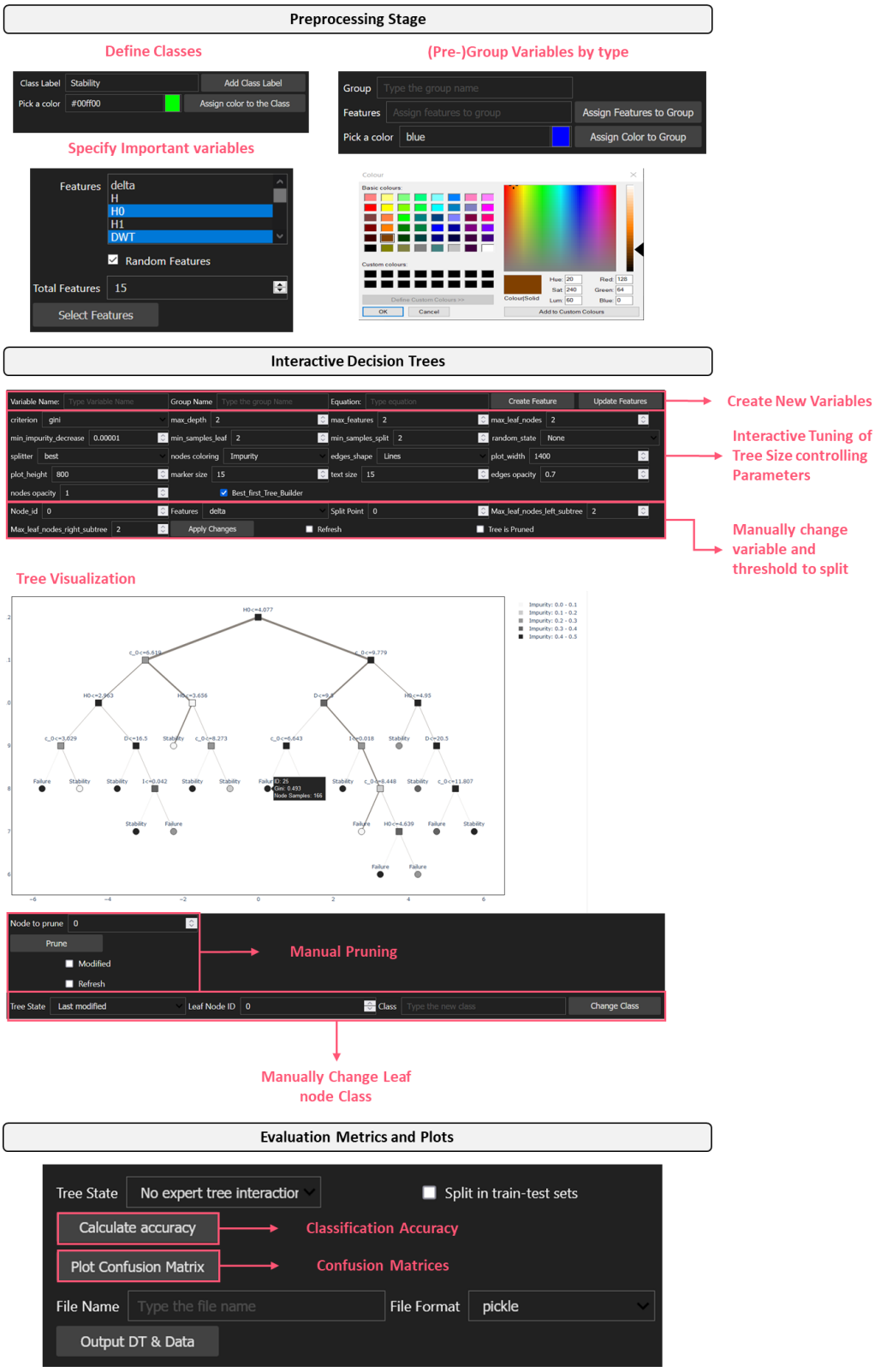


Figure A. 1 *The three tabs of the Interactive Decision Trees Graphical User Interface.*

A3.1 Case studies Supporting Information

A3.1 Case Study 1

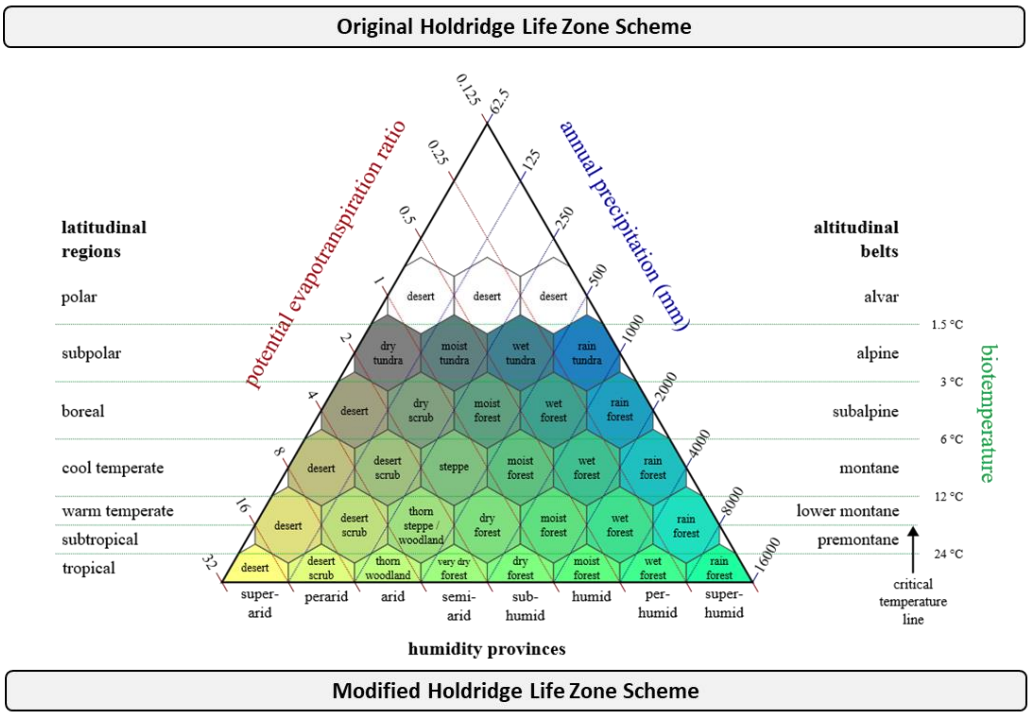
In Table A.1 the groups of variables, the variables belonging to each group and the colour of the group are given.

Table A. 1 *Input Features (CHASM parameters) divided in groups and color coded*

Variable	Symbol	Colour
Geophysical Properties		
Saturated Hydraulic Conductivity:	K_sat_0, K_sat_1	**Brown**
Dry Unit Weight:	gamma_us_0, gamma_us_1	
Saturated Moisture Content:	theta_s_0, theta_s_1, theta_s_2	
Residual Moisture Content:	theta_r_0, theta_r_1, theta_r_2	
Van Genuchten suction-moisture curve α:	alpha_0, alpha_1, alpha_2	
Van Genuchten suction-moisture curve n:	n_0, n_1, n_2	
Effective Cohesion:	c_0, c_1	
Initial Surface Suction:	psi_init	
Effective Friction Angle:	phi_0, phi_1	
Slope Geometry		
Slope Angle:	delta	**Purple**
Thickness of Soil:	H, H0, H1	
Design Storm		
Rainfall Intensity	I	**Blue**
Rainfall Duration	D	
Initial Hydrological Conditions		
Water Table Height	DWT	**Yellow**

A3.2 Case Study 2

In Figure A.2 the original and Modified Holdridge Life scheme is shown.



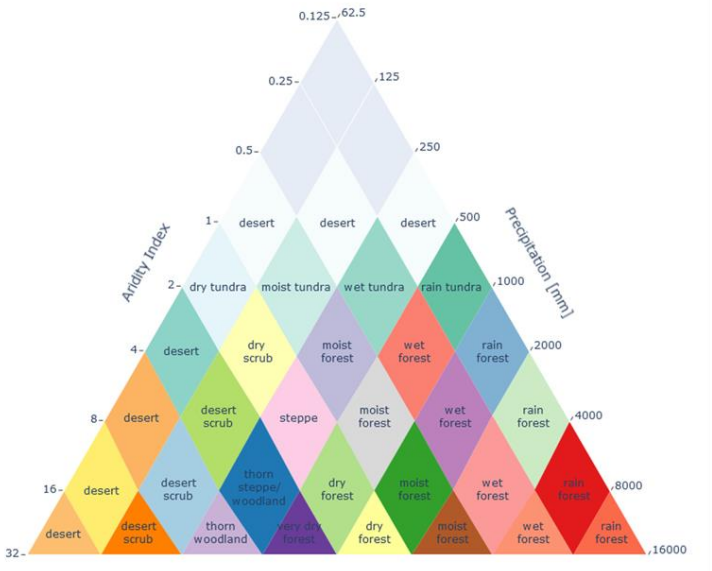


Figure A. 2 *Holdrigge Life Zone Schemes. Top: Original Holdridge Life Zone Scheme, Bottom: Modified Holdridge Life Zone scheme.*

A3.3 Case Study 3

In Figure A.3 the confusion matrices for the statistically optimal and interactive DTs for the training and test sets are shown.

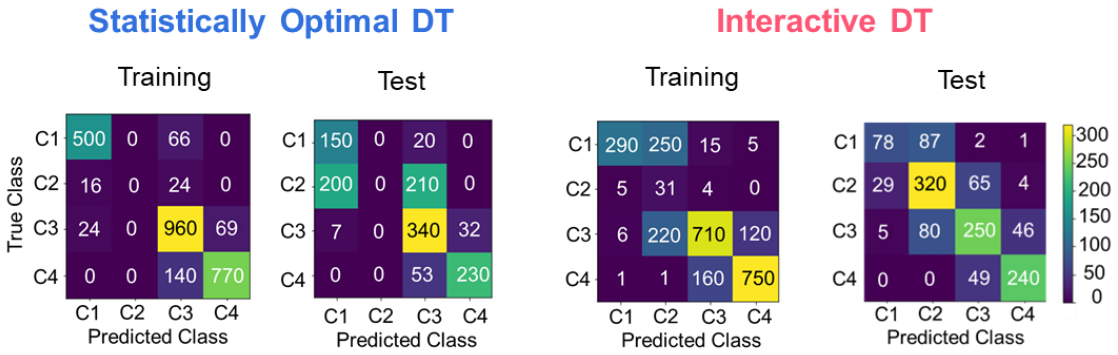


Figure A. 3 *Confusion matrices for statistically optimal and interactive DT for training and test sets.*

Appendix B – Supporting Information for Chapter 3

Table B. 1 *Minimum and Maximum values and Range in Euros per Residential building for the regions Rhine flows over*

Region	Min (Eur/Buil)	Max (Eur/Buil)	Range (Eur/Buil)
Switzerland	64,365	118,857	54,492
Liechtenstein	68,952	129,910	60,958
Austria	65,641	128,423	62,782
France	65,988	128,895	62,907
Germany	60,636	115,971	55,335
Luxembourg	56,959	109,574	52,615
Belgium	58,165	113,474	55,309
Netherlands	61,039	119,090	58,051

Table B. 2 *Minimum and Maximum values and Range for the return period (in years) depending on the peak discharges for two different gauge stations*

Lobith gauge station - Netherlands			
Discharge (m^3/s)	**Min (yrs)**	**Max (yrs)**	**Range (yrs)**
5000	1.27	1.44	0.17
5650	1.59	2.03	0.44
6250	1.76	2.24	0.49
7500	2.81	5.09	2.28
8750	5.08	16.79	11.71
10000	9.58	101.40	91.82
11250	19.76	4797.78	4778.02
Cologne gauge station - Germany			
6250	1.67	3.17	1.50
7500	3.44	4.48	1.04
8750	6.65	9.30	2.65
10000	25.79	35.70	9.91
11250	37.68	121.99	84.31

Appendix C – Supporting Information for Chapter 4

C1 JBA Flood Risk Model

In this thesis I used the flood risk model provided by JBA Risk Management Ltd. It consists of various datasets which may have their own physical or statistical modelling. In the next paragraphs we briefly describe each model dataset/component.

The Stochastic Event Set (Event Set) is a database of plausible event-driving conditions. In our case it is a catalogue of millions of plausible flood events (observed and simulated) in a 10,000 years period. Their severity is expressed in terms of return periods. The modelling chain to generate these events starts with the simulation of precipitation time series (including tropical and non-tropical cyclone related extremes) using spatio-temporal extreme value models. Simulated precipitation is one of the primary inputs that feed rainfall-runoff models to produce river discharge estimates. The other primary inputs are temperature and observed river discharge in locations with available measurements. At ungauged locations JBA uses a conceptual rainfall-runoff model and river discharge estimates are produced through regionalisation (using interpolation methods and the properties of nearby catchments). The simulated daily precipitation and river discharge data (10,000 years) are then transformed into return period estimates based on extreme value theory. Events are then extracted using a multivariate declustering scheme.

Hazard Maps are used to estimate the damage driving characteristics (flood depths and extents) over the area where assets are located. As mentioned in Section 4.2.2 Hazard Maps an undefended view of flood depths and extents for six different return periods (in this case 20, 50, 100, 200, 500 and 1,500 years) To generate these maps two modelling phases took place:

- Hydrological modelling which was performed to derive return period flows across the river network. The hydrological modelling stage starts with the gauge data collection and cleaning and the creation of a network of inflow points. Then, it continued with the delineation of geographically similar hydrological modelling regions and the development of regional flood estimation models. The final step is the derivation of flood hydrographs that will be used in the hydraulic modelling.
- Hydraulic modelling which was performed to apply of those flows onto the digital terrain model (DTM) using JFlow® software to create flood hazard maps.

The *exposure portfolio* contains all the relevant information regarding the exposed assets, e.g., location (coordinates, postcode, country etc) line of business (residential, commercial, industrial etc), economic/insured value etc. JBA model accepts portfolios compatible with Open Exposure Data format. More information can be found in https://oasislmf.org/open-data-standards .

The vulnerability component translates the impact of flood hazard to damages normally using a set of vulnerability functions. JBA model use step functions that vary based on the line of business, coverage and geographic region. In this project I used the curve for residential line of business for Europe.

C2 Definition of dominant input factor

I calculated sensitivity indices to rank the input uncertainties with respect to their relative contribution to the output uncertainty. An intuitive approach to define the dominant input uncertainty is to look at the mean of the bootstrap resamples and the one with the higher mean is considered the dominant. But, this approach may not work well when:

- The means of the bootstrap resamples do not differ significantly.
- There is significant overlapping in the bootstrap resamples' sensitivity indices distributions among the input uncertainties.
- The bootstrap resamples sensitivity indices distributions of the input uncertainties are not normally distributed but skewed.

In such cases, it is not enough to look only at the mean of the bootstrap resamples. An example is shown in Figure 4.4. 4 boxplots are shown and the dots within them represent the calculated sensitivity indices for different bootstrap resamples. Comparison based on the means indicates the Event Set as the dominant input uncertainty for this area. However, if one compare the calculated sensitivity indices in each bootstrap resample separately then there might be cases where other factors could be dominant, like the case I have highlighted in red colour. If this happens frequently, a factor other than the one with the highest mean could be dominant.

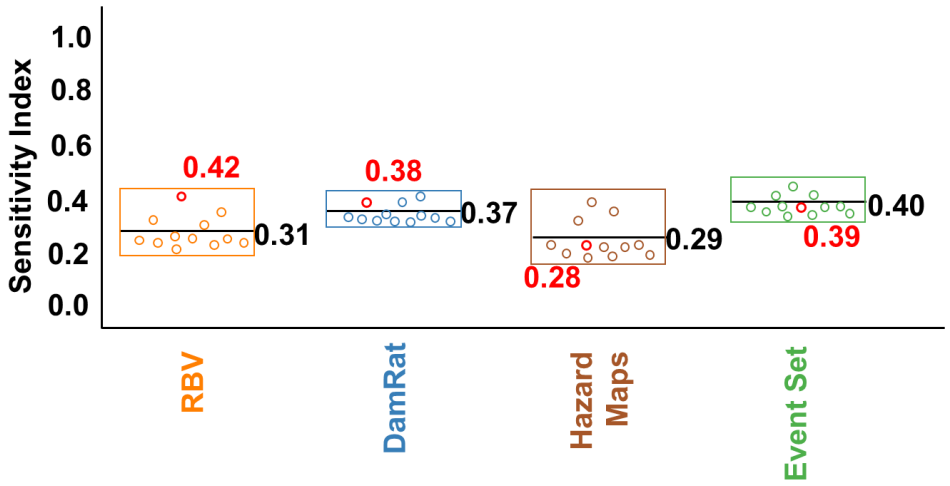


Figure C. 1 *Showcase of the two different approaches for the definition of the dominant input uncertainty.*

.

I proposed a new definition for the dominant input factors that considers the full distribution of the bootstrap resamples and which I described in Chapter 4. In Figure S3.2 I show the effect of the two approaches with results for dominant input uncertainties for Switzerland.

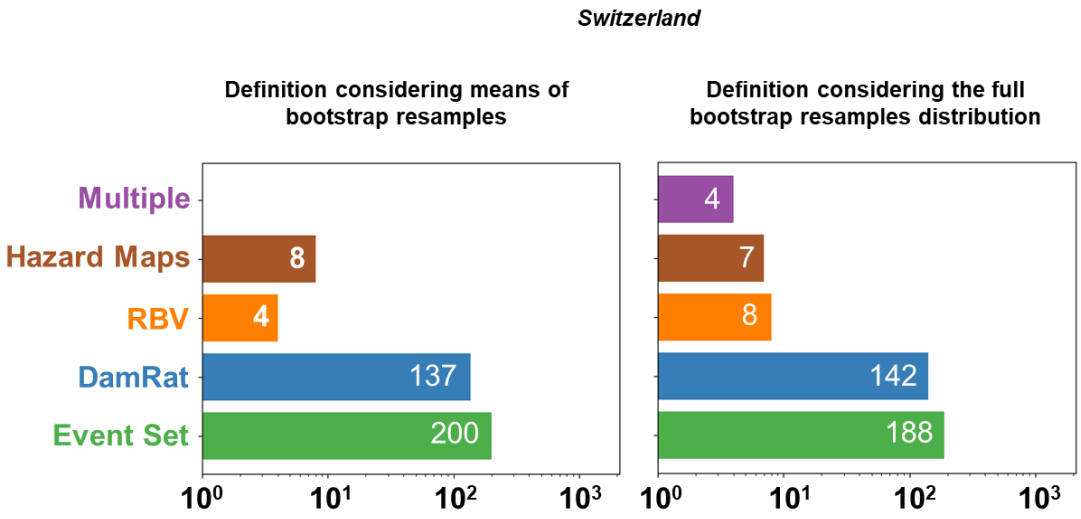


Figure C. 2 *Comparison of two different definitions for the dominant input uncertainties with results from Switzerland.*

C3 Dataset of hydrological, climatic, land cover, socio-economic variables and the associated dominant input uncertainties.

At each polygon the dominant input uncertainty is known and additionally, I calculated the values of the properties shown in Table 4.1 (Chapter 4). The resulting dataset consists of 2,102 of rows and 18 columns. Each row represents a different High Resolution CRESTA

zone (CRESTA, 2022). Each column represents a different property, except the last one which represents the dominant input uncertainty and will be treated as the classes by the DT.

In the following table I show part of the dataset I used to link the dominant input uncertainties to hydrological, climatic, land cover and socio-economic properties of the places.

Table C. 1 *The table shows as an example part of the dataset used to link system characteristics with spatially varying dominant input uncertainties.*

CRESTA ID	Riv Len	Av Disch	Max StOr	Maj FD	Mean FD	Max FD	FD3	Hyd Bas	Clim Class	ARTF	AGR	FRST	WETLN	WB	MeanVal	Pop	Area	Classes
AUT_6700	23	26	2	0.03	0.66	5	23	2	26	19	11	70	0	0	4175478	419	32	Event Set
AUT_6708	17	2	1	0.1	0.22	4	3	2	26	4	3	93	0	0	8147274	179	40	Event Set
AUT_6710	41	38	1	0.02	0.67	12	27	2	26	4	4	92	0	0	2240500	288	100	Hazard Maps
AUT_6712	11	6	3	0.02	0.67	5	2	2	26	24	33	43	0	0	1222091	73	6	Event Set
AUT_6714	13	27	3	0.19	1.04	4	28	2	26	13	10	77	0	0	2036818	181	22	Event Set
AUT_6764	5	0	1	12	1.04	12	27	2	27	3	3	94	0	0	1120250	203	90	Damage Ratio
…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…

C4 Derivation of 15 Statistically optimal DT

I derived 15 statistically optimal (SOT) DT by trying different number of leaf nodes (varied from 10 to 30). For each case, I chose the SOT based on 10-fold Cross Validation strategy and recorded their classification accuracy (on the train and test sets) and their interpretability (number of leaf nodes and tree depth). Then, I computed mean metrics of classification and interpretability performance. These are summarized in the tables on the top right of Figure C.3.

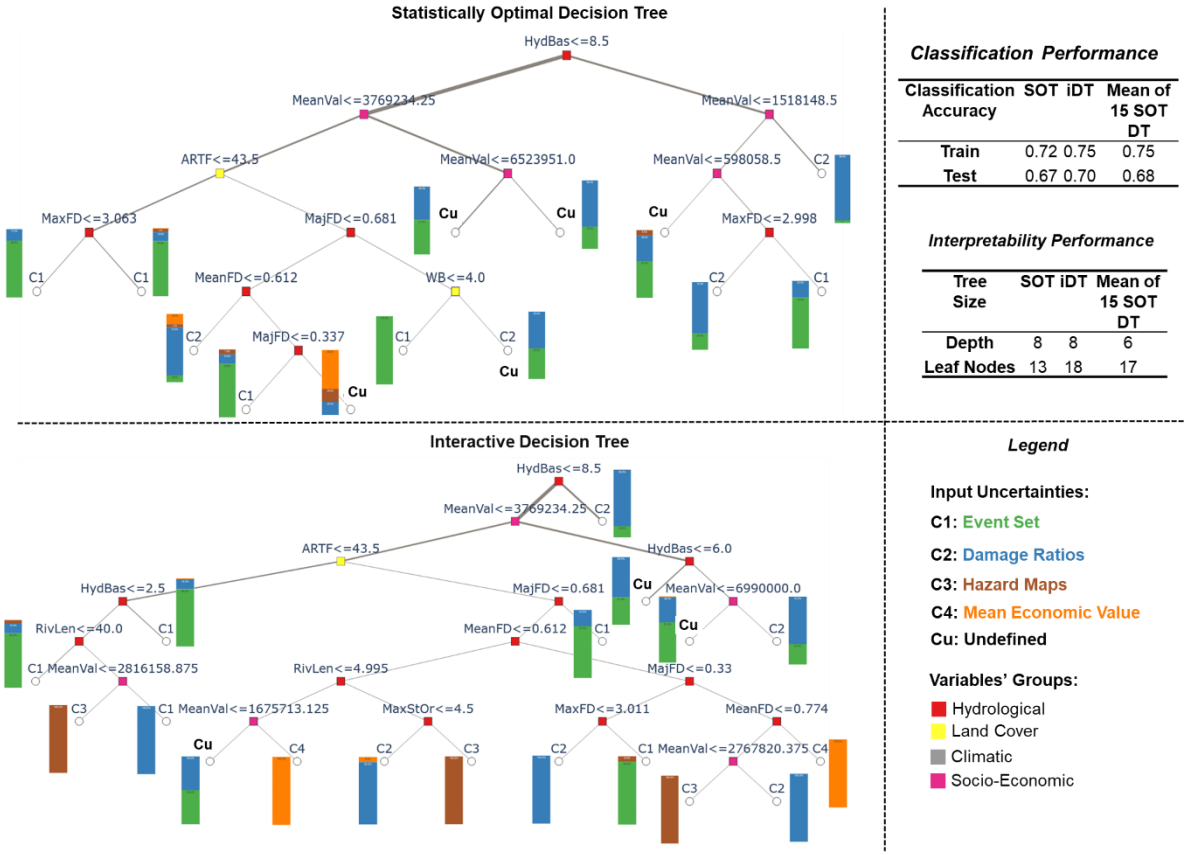


Figure C. 3 *Statistically optimal and Interactive Decision Trees and their classification and interpretability performances.*

Appendix D – Curriculum Vitae

Georgios Sarailidis
g.sarailidis@bristol.ac.uk

Higher Education	
2018-Present	**PhD Water Informatics Science & Engineering (WISE CDT)** University of Bristol, United Kingdom *Thesis: Uncertainty quantification and attribution in flood risk modelling.* *Supervisors: Francesca Pianosi, Thorsten Wagener*
2015-2017	**MSc in Management of Hydrometeorological Hazards – Hydrohazards** Universite Grenoble Alpes, France & University of Thessaly, Greece *Master Thesis: Analysis of Drought Variability in Sicily using SPI and SPEI drought indexes* *Supervisor: Antonino Cancelliere*
2009-2015	**Meng in Civil Engineering** University of Thessaly, Greece *Master Thesis: The quantification of the threshold level method on low flow studies* *Supervisors: Athanasios Loukas, Lampros Vasiliades*
2017	**Observer Meteorologist Diploma** Hellenic National Meteorological Service, Ministry of National Defense, Greece

Work Experience	
2023-Present:	**Catastrophe Modelling KTP Associate** University of Bristol & JBA Risk Management Ltd *Fixed term contract: Embed uncertainty and sensitivity analysis functionalities into JBA Risk Management's Global Flood Modelling capabilities*
2018-2022	**Teaching Support Assistant** University of Bristol *Water Resources Project, Numerical Analysis Using Matlab, Research Project Showcase*
2017-2018	**Service at the Greek Regional Meteorological Centre** Hellenic Air Force, Greece *Fixed term Project: Statistical frequency analysis of extreme rainfall in Greece*

Research Visits & Collaborations	
2018-2023	**Industry Collaboration** JBA Trust, JBA Risk Management Ltd, Skipton, United Kingdom

	Short research visits & Regular meetings (presenting, reviewing and discussing the PhD work with experts within the catastrophe modelling team)
2022	**Research Visit to GFZ Potsdam** *Short research visit (Presenting, reviewing and discussing my PhD work with experts in the Hydrology department of GFZ, German Research Center for Geosciences)*

Distinctions and Awards	
2018	Awarded an EPSRC Doctoral Training Partnership Studentship for the Water Informatics: Science and Engineering (WISE) CDT Program
2017	Graduated with honors on the MSc program “Management of Hydrometeorological hazards-Hydrohazards”.
2014	First place at the business and innovative ideas contest “Startup Weekend of Volos”

Skills	
Languages	Greek (Mother tongue), English (fluent), French (good), Russian (good)
Software	Python, R, Matlab, ArcGIS, QGIS
Leadership and Organization Skills	**Organizing Seminar and Workshops:** Organizer of bi-weekly research seminars of the Water and Environmental Engineering research group at the University of Bristol (2021-22) Co-organizer of 1-day Workshop on “Uncertainty in Mathematical Modelling” for PhD students in the Civil Engineering Dept. at University of Bristol, 15th July 2022 Co-organizer of Workshop on Equality in Engineering (University of Bristol) Organizer of Workshop on the “Physics, chemistry and aerosols of the atmosphere” (Universite Grenoble Alpes) **Leading and representing students:** President and Secretary of the Student Union of the Civil Engineering Dept. at the University of Thessaly Student Representative at the General Assembly of the Civil Engineering Dept. at the University of Thessaly

Publications	
Journals	Sarailidis G., Pianosi, F., Wagener, T., (in preparation) Linking the relative importance of input uncertainties of a flood risk model to River Rhine spatial characteristics. To be submitted to *Natural Hazards and Earth System Sciences* Sarailidis G., Wagener, T., Pianosi,F. (2022) Integrating scientific knowledge into machine learning using interactive decision trees. Computers &

Geosciences, 170, 105248, https://doi.org/10.1016/j.cageo.2022.105248

Sarailidis, G., Vasiliades L., Loukas, A. (2018) Analysis of streamflow droughts using fixed and variable thresholds, Hydrological Processes, 33 (3), 414-431,

Presentations at scientific conferences	
Oral	Sarailidis, G., Pianosi, F., Wagener, T., Styles, K., Lamb, R., Hutchings S., What controls uncertainty in flood risk estimates? An analysis across the Rhine River basin, EGU General Assembly 2023, https://doi.org/10.5194/egusphere-egu23-3836.
	Sarailidis, G., Pianosi, F., Wagener, T., Lamb, R., Styles, K., Hutchings S., Linking the relative importance of input uncertainties of a flood risk model with basin characteristics, EGU General Assembly 2022, https://doi.org/10.5194/egusphere-egu22-3122
	Sarailidis, G., Pianosi, F., Wagener, T., Styles, K., Hutchings S., Lamb, R., Uncertainty quantification and attribution in flood risk assessment using Global Flood Models: an application to the river Rhine basin, EGU General Assembly 2021, https://doi.org/10.5194/egusphere-egu21-3488
Virtual	Sarailidis, G., Wagener, T., Pianosi, F., Integrating Human Knowledge and Data Mining through interactive Classification and Regression Trees (CART), 10th International Congress on Environmental Modelling and Software 2020
	Sarailidis, G., Wagener, T., Pianosi, F., Integrating Human Knowledge and Data Mining through interactive Classification and Regression Trees (CART), AGU Fall Meeting 2020

References

Addor, N., Nearing, G., Prieto, C., Newman, A. J., le Vine, N., & Clark, M. P. (2018). A Ranking of Hydrological Signatures Based on Their Predictability in Space. *Water Resources Research*, *54*(11), 8792–8812. https://doi.org/10.1029/2018WR022606

Alfieri, L., Burek, P., Dutra, E., Krzeminski, B., Muraro, D., Thielen, J., & Pappenberger, F. (2013). GloFAS-global ensemble streamflow forecasting and flood early warning. *Hydrology and Earth System Sciences*, *17*(3), 1161–1175. https://doi.org/10.5194/hess-17-1161-2013

Almeida, S., Ann Holcombe, E., Pianosi, F., & Wagener, T. (2017). Dealing with deep uncertainties in landslide modelling for disaster risk reduction under climate change. *Natural Hazards and Earth System Sciences*, *17*(2), 225–241. https://doi.org/10.5194/nhess-17-225-2017

Altaf, M. (2022). *Pakistan: Flood Damages and Economic Losses Over USD 30 billion and Reconstruction Needs Over USD 16 billion - New Assessment*. https://www.worldbank.org/en/news/press-release/2022/10/28/pakistan-flood-damages-and-economic-losses-over-usd-30-billion-and-reconstruction-needs-over-usd-16-billion-new-assessme

Ankerst, M., Ester, M., & Kriegel, H. P. (2000). Towards an effective cooperation of the user and the computer for classification. *Proceeding of the Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 179–188. https://doi.org/10.1145/347090.347124

Apel, H., Aronica, G. T., Kreibich, H., & Thieken, A. H. (2009). Flood risk analyses - How detailed do we need to be? *Natural Hazards*, *49*(1). https://doi.org/10.1007/s11069-008-9277-8

Apel, H., Merz, B., & Thieken, A. H. (2008). Quantification of uncertainties in flood risk assessments. *International Journal of River Basin Management*, *6*(2). https://doi.org/10.1080/15715124.2008.9635344

Apel, H., Thieken, A. H., Merz, B., & Blöschl, G. (2004). Flood risk assessment and associated uncertainty. *Natural Hazards and Earth System Science*, *4*(2). https://doi.org/10.5194/nhess-4-295-2004

Apel, H., Vorogushyn, S., & Merz, B. (2022). Brief communication - Impact Forecasting Could Substantially Improve the Emergency Management of Deadly Floods: Case

Study July 2021 floods in Germany. *Natural Hazards and Earth System Sciences Discussions*, *February*.

Archer, G. E. B., Saltelli, A., & Sobol, I. M. (1997). Sensitivity measures, anova-like techniques and the use of bootstrap. *Journal of Statistical Computation and Simulation*, *58*(2). https://doi.org/10.1080/00949659708811825

AXA Future Risks Report. (2022). *AXA Future Risks Report 2022*. https://www.axa.com/en/news/2022-future-risks-report

Baldassarre, G. di, Kreibich, H., Vorogushyn, S., Aerts, J., Arnbjerg-Nielsen, K., Barendrecht, M., Bates, P., Borga, M., Botzen, W., Bubeck, P., de Marchi, B., Llasat, C., Mazzoleni, M., Molinari, D., Mondino, E., Märd, J., Petrucci, O., Scolobig, A., Viglione, A., & Ward, P. J. (2018). Hess opinions: An interdisciplinary research agenda to explore the unintended consequences of structural flood protection. *Hydrology and Earth System Sciences*, *22*(11). https://doi.org/10.5194/hess-22-5629-2018

Beck, H. E., Zimmermann, N. E., McVicar, T. R., Vergopolan, N., Berg, A., & Wood, E. F. (2018). Present and future köppen-geiger climate classification maps at 1-km resolution. *Scientific Data*, *5*. https://doi.org/10.1038/sdata.2018.214

Bentivoglio, R., Isufi, E., Jonkman, S. N., & Taormina, R. (2022). Deep learning methods for flood mapping: a review of existing applications and future research directions. In *Hydrology and Earth System Sciences* (Vol. 26, Issue 16). https://doi.org/10.5194/hess-26-4345-2022

Bergen, K. J., Johnson, P. A., de Hoop, M. v., & Beroza, G. C. (2019). Machine learning for data-driven discovery in solid Earth geoscience. *Science*, *363*(6433). https://doi.org/10.1126/science.aau0323

Bernhofen, M. v., Cooper, S., Trigg, M., Mdee, A., Carr, A., Bhave, A., Solano-Correa, Y. T., Pencue-Fierro, E. L., Teferi, E., Haile, A. T., Yusop, Z., Alias, N. E., Sa'adi, Z., bin Ramzan, M. A., Dhanya, C. T., & Shukla, P. (2022). The Role of Global Data Sets for Riverine Flood Risk Management at National Scales. *Water Resources Research*, *58*(4). https://doi.org/10.1029/2021WR031555

Beven, K. J., Almeida, S., Aspinall, W. P., Bates, P. D., Blazkova, S., Borgomeo, E., Freer, J., Goda, K., Hall, J. W., Phillips, J. C., Simpson, M., Smith, P. J., Stephenson, D. B., Wagener, T., Watson, M., & Wilkins, K. L. (2018). Epistemic uncertainties and natural hazard risk assessment - Part 1: A review of different natural hazard areas. *Natural*

Hazards and Earth System Sciences, *18*(10), 2741–2768. https://doi.org/10.5194/nhess-18-2741-2018

Blöschl, G., Hall, J., Parajka, J., Perdigão, R. A. P., Merz, B., Arheimer, B., Aronica, G. T., Bilibashi, A., Bonacci, O., Borga, M., Čanjevac, I., Castellarin, A., Chirico, G. B., Claps, P., Fiala, K., Frolova, N., Gorbachova, L., Gül, A., Hannaford, J., … Živković, N. (2017). Changing climate shifts timing of European floods. *Science*, *357*(6351). https://doi.org/10.1126/science.aan2506

Bomers, A., Schielen, R. M. J., & Hulscher, S. J. M. H. (2019). Decreasing uncertainty in flood frequency analyses by including historic flood events in an efficient bootstrap approach. *Natural Hazards and Earth System Sciences*, *19*(8). https://doi.org/10.5194/nhess-19-1895-2019

Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). Classification and regression trees. In *Classification and Regression Trees*. CRC Press. https://doi.org/10.1201/9781315139470

Brewer, J. F., Bishop, M., Kelp, M., Keller, C. A., Ravishankara, A. R., & Fischer, E. v. (2017). A sensitivity analysis of key natural factors in the modeled global acetone budget. *Journal of Geophysical Research*, *122*(3). https://doi.org/10.1002/2016JD025935

Bubeck, P., & de Moel, H. (2010). *Sensitivity analysis of flood damage calculations for the river Rhine*.

Butler, D. (2007). Earth monitoring: The planetary panopticon. *Nature*, *450*(7171), 778–781. https://doi.org/10.1038/450778a

Bzdok, D., Krzywinski, M., & Altman, N. (2017). Machine learning: a primer. *Nature Methods*, *14*(12), 1119–1120. https://doi.org/10.1038/nmeth.4526

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). https://doi.org/10.5194/nhess-13-3063-2013

Chawla, N. v., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, *16*, 321–357. https://doi.org/10.1613/jair.953

Chen, J., Huang, G., & Chen, W. (2021). Towards better flood risk management: Assessing flood risk and investigating the potential mechanism based on machine learning

models. *Journal of Environmental Management*, *293*. https://doi.org/10.1016/j.jenvman.2021.112810

Choo, J., & Liu, S. (2018). Visual Analytics for Explainable Deep Learning. *IEEE Computer Graphics and Applications*, *38*(4). https://doi.org/10.1109/MCG.2018.042731661

Cooke, R. M. (1991). Experts in Uncertainty: Opinion and Subjective Probability in Science (Environmental Ethics and Science Policy). In *Oxford University Press* (Vol. 44, Issue 2).

Copernicus. (2018). *CORINE Land Cover*. https://land.copernicus.eu/pan-european/corine-land-cover

Coxon, G., Freer, J., Westerberg, I. K., Wagener, T., Woods, R., & Smith, P. J. (2015). A novel framework for discharge uncertainty quantification applied to 500 UK gauging stations. *Water Resources Research*, *51*(7). https://doi.org/10.1002/2014WR016532

CRESTA. (2022). *CRESTA zones and maps*. https://about.cresta.org/

de Moel, H., & Aerts, J. C. J. H. (2011). Effect of uncertainty in land use, damage models and inundation depth on flood damage estimates. *Natural Hazards*, *58*(1). https://doi.org/10.1007/s11069-010-9675-6

de Moel, H., Bouwer, L. M., & Aerts, J. C. J. H. (2014). Uncertainty and sensitivity of flood risk calculations for a dike ring in the south of the Netherlands. *Science of the Total Environment*, *473–474*. https://doi.org/10.1016/j.scitotenv.2013.12.015

Deroche, M. S. (2023). Invited perspectives: An insurer's perspective on the knowns and unknowns in natural hazard risk modelling. Natural Hazards and Earth System Sciences, 23, 251–259. https://doi.org/https://doi.org/10.5194/nhess-23-251-2023

Devitt, L., Neal, J., Wagener, T., & Coxon, G. (2021). Uncertainty in the extreme flood magnitude estimates of large-scale flood hazard models. *Environmental Research Letters*, *16*(6). https://doi.org/10.1088/1748-9326/abfac4

Doshi-Velez, F., & Been, K. (2017). *Towards A Rigorous Science of Interpretable Machine Learning* (pp. 1–13).

Duha Metin, A., Viet Dung, N., Schröter, K., Guse, B., Apel, H., Kreibich, H., Vorogushyn, S., & Merz, B. (2018). How do changes along the risk chain affect flood risk? *Natural Hazards and Earth System Sciences*, *18*(11). https://doi.org/10.5194/nhess-18-3089-2018

Efron, B., & Tibshirani, R. J. (1994). An Introduction to the Bootstrap. In *An Introduction to the Bootstrap*. https://doi.org/10.1201/9780429246593

Egorova, R., van Noortwijk, J. M., & Holterman, S. R. (2008). Uncertainty in flood damage estimation Regina. *International Journal of River Basin Management*, 139–148. https://doi.org/https://doi.org/10.1080/15715124.2008.9635343

Elia, M., Gajek, C., Schiendorfer, A., & Wolfgang, R. (2021). An interactive web application for decision tree learning. *Proceedings of the European Conference on Machine Learning, Teaching Machine Learning Workshop*.

Elmer, F., Thieken, A. H., Pech, I., & Kreibich, H. (2010). Influence of flood frequency on residential building losses. *Natural Hazards and Earth System Science*, *10*(10). https://doi.org/10.5194/nhess-10-2145-2010

Estivill-Castro, V., Gilmore, E., & Hexel, R. (2020). Human-In-The-Loop Construction of Decision Tree Classifiers with Parallel Coordinates. *Conference Proceedings - IEEE International Conference on Systems, Man and Cybernetics*, *2020-October*. https://doi.org/10.1109/SMC42975.2020.9283240

European Commission. (2007). *Directive 2007/60/EC of the European Parliament and of the Council of 23 October 2007 on the assessment and management of flood risks*. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32007L0060

Eurostat. (2013). Handbook on Residential Property Prices (RPPIs). In *Handbook on Residential Property Prices (RPPIs)*. https://doi.org/10.5089/9789279259845.069

Eurostat. (2022a). *Eurostat Building Area*. https://ec.europa.eu/eurostat/databrowser/view/STS_COBP_A__custom_4851173/default/table?lang=en

Eurostat. (2022b). *Eurostat Building Construction Costs index*. https://ec.europa.eu/eurostat/databrowser/view/STS_COPI_A__custom_4853382/default/table?lang=en

Eurostat. (2022c). *Eurostat House Price*. https://ec.europa.eu/eurostat/databrowser/view/prc_hpi_a/default/table?lang=en

Faghmous, J. H., & Kumar, V. (2014). A Big Data Guide to Understanding Climate Change: The Case for Theory-Guided Data Science. *Big Data*, *2*(3), 155–163. https://doi.org/10.1089/big.2014.0026

Fails, J. A., & Olsen, D. R. (2003). Interactive machine learning. *International Conference on Intelligent User Interfaces, Proceedings IUI*. https://doi.org/10.1145/604045.604056

Flach, P. (2012). *Machine Learning The Art and Science of Algorithms that Make Sense of Data*. Cambridge University Press. https://doi.org/10.1017/CBO9780511973000

Franco, G., Becker, J. F., & Arguimbau, N. (2020). Evaluation methods of flood risk models in the (re)insurance industry. *Water Security*, *11*. https://doi.org/10.1016/j.wasec.2020.100069

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). https://doi.org/10.2166/wst.2010.177

García, S., & Herrera, F. (2009). Evolutionary undersampling for classification with imbalanced datasets: Proposals and taxonomy. *Evolutionary Computation*, *17*(3), 275–306. https://doi.org/10.1162/evco.2009.17.3.275

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). https://doi.org/10.1371/journal.pone.0159791

Gibert, K., García-Alonso, C., & Salvador-Carulla, L. (2010). Integrating clinicians, knowledge and data: Expert-based cooperative analysis in healthcare decision support. In *Health Research Policy and Systems* (Vol. 8). https://doi.org/10.1186/1478-4505-8-28

Gil, Y., David, C. H., Demir, I., Essawy, B. T., Fulweiler, R. W., Goodall, J. L., Karlstrom, L., Lee, H., Mills, H. J., Oh, J. H., Pierce, S. A., Pope, A., Tzeng, M. W., Villamizar, S. R., & Yu, X. (2016). Toward the Geoscience Paper of the Future: Best practices for documenting and sharing research from data to software to provenance. In *Earth and Space Science* (Vol. 3, Issue 10). https://doi.org/10.1002/2015EA000136

Gislason, P. O., Benediktsson, J. A., & Sveinsson, J. R. (2006). Random forests for land cover classification. *Pattern Recognition Letters*, *27*(4), 294–300. https://doi.org/10.1016/j.patrec.2005.08.011

Grimm, R., Behrens, T., Märker, M., & Elsenbeer, H. (2008). Soil organic carbon concentrations and stocks on Barro Colorado Island - Digital soil mapping using Random Forests analysis. *Geoderma*, *146*(1–2), 102–113. https://doi.org/10.1016/j.geoderma.2008.05.008

Grossi Partricia, & Kunreuther Howard. (2005). Catastrophe Modeling: A New Approach to Managing Risk. In *Catastrophe Modeling: A New Approach to Managing Risk*. Kluwer Academic Publishers. https://doi.org/10.1007/b100669

Hall, J. W. (2014). Editorial: Steps towards global flood risk modelling. In *Journal of Flood Risk Management* (Vol. 7, Issue 3). https://doi.org/10.1111/jfr3.12119

Hall, J. W., Sayers, P. B., & Dawson, R. J. (2005). National-scale Assessment of Current and Future Flood Risk in England and Wales. *Natural Hazards*, *36*, 147–164. https://doi.org/https://doi.org/10.1007/s11069-004-4546-7

Han, J., & Cercone, N. (2001). Interactive Construction of Decision Trees. *5th Pacific-Asia Conference on Knowledge Discovery and Data Mining*. https://doi.org/10.1007/3-540-45357-1_61

Haraguchi, M., & Lall, U. (2015). Flood risks and impacts: A case study of Thailand's floods in 2011 and research questions for supply chain decision making. *International Journal of Disaster Risk Reduction*, *14*. https://doi.org/10.1016/j.ijdrr.2014.09.005

Hart, J. K., & Martinez, K. (2006). Environmental Sensor Networks: A revolution in the earth system science? *Earth-Science Reviews*, *78*(3–4), 177–191. https://doi.org/10.1016/j.earscirev.2006.05.001

Heffernan, J. E., & Tawn, J. A. (2004). A conditional approach for multivariate extreme values. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, *66*(3). https://doi.org/10.1111/j.1467-9868.2004.02050.x

Hengl, T., de Jesus, J. M., Heuvelink, G. B. M., Gonzalez, M. R., Kilibarda, M., Blagotić, A., Shangguan, W., Wright, M. N., Geng, X., Bauer-Marschallinger, B., Guevara, M. A., Vargas, R., MacMillan, R. A., Batjes, N. H., Leenaars, J. G. B., Ribeiro, E., Wheeler, I., Mantel, S., & Kempen, B. (2017). SoilGrids250m: Global gridded soil information based on machine learning. *PLoS ONE*, *12*(2). https://doi.org/10.1371/journal.pone.0169748

Holdridge, L. R. (1947). Determination of world plant formations from simple climatic data. *Science*, *105*(2727). https://doi.org/10.1126/science.105.2727.367

Holzinger, A. (2016). Interactive machine learning for health informatics: when do we need the human-in-the-loop? *Brain Informatics*, *3*(2). https://doi.org/10.1007/s40708-016-0042-6

Huizinga, J., de Moel, H., & Szewczyk, W. (2017). Global flood depth-damage functions. Methodology and the database with guidelines. In *Joint Research Centre (JRC)*.

Hutton, C., Wagener, T., Freer, J., Han, D., Duffy, C., & Arheimer, B. (2016). Most computational hydrology is not reproducible, so is it really science? In *Water Resources Research* (Vol. 52, Issue 10, pp. 7548–7555). https://doi.org/10.1002/2016WR019285

IBM. (2020, July 15). *Machine Learning*. https://www.ibm.com/cloud/learn/machine-learning

Iorgulescu, I., & Beven, K. J. (2004). Nonparametric direct mapping of rainfall-runoff relationships: An alternative approach to data analysis and modeling? *Water Resources Research*, *40*(8). https://doi.org/10.1029/2004WR003094

IPCC Working Group II. (2022). IPCC AR6 Working Group II: Summary for policymakers: Climate Change 2022, Impacts, Adaptation anf Vulnerability. *Implementing a US Carbon Tax: Challenges and Debates*.

Jongman, B., Koks, E. E., Husby, T. G., & Ward, P. J. (2014). Increasing flood exposure in the Netherlands: Implications for risk financing. *Natural Hazards and Earth System Sciences*, *14*(5). https://doi.org/10.5194/nhess-14-1245-2014

Jongman, B., Kreibich, H., Apel, H., Barredo, J. I., Bates, P. D., Feyen, L., Gericke, A., Neal, J., Aerts, J. C. J. H., & Ward, P. J. (2012a). Comparative flood damage model assessment: Towards a European approach. *Natural Hazards and Earth System Science*, *12*(12). https://doi.org/10.5194/nhess-12-3733-2012

Jongman, B., Ward, P. J., & Aerts, J. C. J. H. (2012b). Global exposure to river and coastal flooding: Long term trends and changes. *Global Environmental Change*, *22*(4). https://doi.org/10.1016/j.gloenvcha.2012.07.004

Jongman, B., Winsemius, H. C., Aerts, J. C. J. H., Coughlan de Perez, E., 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. https://doi.org/10.1073/pnas.1414439112

Jonkman, S. N., Kok, M., & Vrijling, J. K. (2008). Flood risk assessment in the Netherlands: A case study for dike ring South Holland. *Risk Analysis*, *28*(5). https://doi.org/10.1111/j.1539-6924.2008.01103.x

Judith Ellison. (2023). *Regulators build momentum on Climate Change and ESG Reporting*. Regulators Build Momentum on Climate Change and ESG Reporting. https://www.jbarisk.com/news-blogs/regulators-build-momentum-on-climate-change-and-esg-reporting/

Kaczmarska, J., Jewson, S., & Bellone, E. (2018). Quantifying the sources of simulation uncertainty in natural catastrophe models. *Stochastic Environmental Research and Risk Assessment*, *32*(3). https://doi.org/10.1007/s00477-017-1393-0

Karpatne, A., Atluri, G., Faghmous, J. H., Steinbach, M., Banerjee, A., Ganguly, A., Shekhar, S., Samatova, N., & Kumar, V. (2017). Theory-guided data science: A new paradigm for scientific discovery from data. *IEEE Transactions on Knowledge and Data Engineering*, *29*(10), 2318–2331. https://doi.org/10.1109/TKDE.2017.2720168

Karpatne, A., Ebert-Uphoff, I., Ravela, S., Babaie, H. A., & Kumar, V. (2019). Machine Learning for the Geosciences: Challenges and Opportunities. *IEEE Transactions on Knowledge and Data Engineering*, *31*(8), 1544–1554. https://doi.org/10.1109/TKDE.2018.2861006

Keef, C., Tawn, J. A., & Lamb, R. (2013). Estimating the probability of widespread flood events. *Environmetrics*, *24*(1). https://doi.org/10.1002/env.2190

Kirchner, J. W., Berghuijs, W. R., Allen, S. T., Hrachowitz, M., Hut, R., & Rizzo, D. M. (2020). Streamflow response to forest management. *Nature*, *578*(7794), E12–E15. https://doi.org/10.1038/s41586-020-1940-6

Kolmogorov, A. (1933). Sulla determinazione empirica di una legge didistribuzione. *Giornale Dell'Instituto Italiano Degli Attuari* , *4*, 83–91.

Kreibich, H., di Baldassarre, G., Vorogushyn, S., Aerts, J. C. J. H., Apel, H., Aronica, G. T., Arnbjerg-Nielsen, K., Bouwer, L. M., Bubeck, P., Caloiero, T., Chinh, D. T., Cortès, M., Gain, A. K., Giampá, V., Kuhlicke, C., Kundzewicz, Z. W., Llasat, M. C., Mård, J., Matczak, P., … Merz, B. (2017). Adaptation to flood risk: Results of international paired flood event studies. *Earth's Future*, *5*(10). https://doi.org/10.1002/2017EF000606

Krueger, T., Page, T., Hubacek, K., Smith, L., & Hiscock, K. (2012). The role of expert opinion in environmental modelling. *Environmental Modelling and Software*, *36*. https://doi.org/10.1016/j.envsoft.2012.01.011

Kuentz, A., Arheimer, B., Hundecha, Y., & Wagener, T. (2017). Understanding hydrologic variability across Europe through catchment classification. *Hydrology and Earth System Sciences*, *21*(6), 2863–2879. https://doi.org/10.5194/hess-21-2863-2017

Kundzewicz, Z. W., Hegger, D. L. T., Matczak, P., & Driessen, P. P. J. (2018). Flood-risk reduction: Structural measures and diverse strategies. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(49). https://doi.org/10.1073/pnas.1818227115

Lamb, R., Keef, C., Tawn, J., Laeger, S., Meadowcroft, I., Surendran, S., Dunning, P., & Batstone, C. (2010). A new method to assess the risk of local and widespread flooding on rivers and coasts. *Journal of Flood Risk Management*, *3*(4). https://doi.org/10.1111/j.1753-318X.2010.01081.x

le Cozannet, G., Rohmer, J., Cazenave, A., Idier, D., van de Wal, R., de Winter, R., Pedreros, R., Balouin, Y., Vinchon, C., & Oliveros, C. (2015). Evaluating uncertainties of future marine flooding occurrence as sea-level rises. *Environmental Modelling and Software*, *73*. https://doi.org/10.1016/j.envsoft.2015.07.021

Lehner, B., & Grill, G. (2013). Global river hydrography and network routing: Baseline data and new approaches to study the world's large river systems. *Hydrological Processes*, *27*(15). https://doi.org/10.1002/hyp.9740

Lighthill Risk Network. (2019). *Flood Research Needs of the (Re)insurance sector. Collaborating to improve risk understanding and management*.

Lindersson, S., Brandimarte, L., Märd, J., & di Baldassarre, G. (2021). Global riverine flood risk - How do hydrogeomorphic floodplain maps compare to flood hazard maps? *Natural Hazards and Earth System Sciences*, *21*(10). https://doi.org/10.5194/nhess-21-2921-2021

Lipton, Z. C. (2018). The mythos of model interpretability. *Communications of the ACM*, *61*(10), 36–43. https://doi.org/10.1145/3233231

Loh, W. Y. (2014). Fifty years of classification and regression trees. *International Statistical Review*, *82*(3), 329–348. https://doi.org/10.1111/insr.12016

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*(12). https://doi.org/10.1029/2019WR026213

Mallapaty, S. (2022). Why are Pakistan's floods so extreme this year? *Nature*. https://www.nature.com/articles/d41586-022-02813-6

McGrath, H., Abo El Ezz, A., & Nastev, M. (2019). Probabilistic depth–damage curves for assessment of flood-induced building losses. *Natural Hazards*, *97*(1). https://doi.org/10.1007/s11069-019-03622-3

Merwade, V., Olivera, F., Arabi, M., & Edleman, S. (2008). Uncertainty in Flood Inundation Mapping: Current Issues and Future Directions. *Journal of Hydrologic Engineering*, *13*(7). https://doi.org/10.1061/(asce)1084-0699(2008)13:7(608)

Merz, B., Kreibich, H., Schwarze, R., & Thieken, A. (2010). Review article “assessment of economic flood damage.” In *Natural Hazards and Earth System Science* (Vol. 10, Issue 8, pp. 1697–1724). https://doi.org/10.5194/nhess-10-1697-2010

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). https://doi.org/10.5194/nhess-4-153-2004

Merz, B., & Thieken, A. H. (2005). Separating natural and epistemic uncertainty in flood frequency analysis. *Journal of Hydrology*, *309*(1–4). https://doi.org/10.1016/j.jhydrol.2004.11.015

Merz, B., & Thieken, A. H. (2009). Flood risk curves and uncertainty bounds. *Natural Hazards*, *51*(3). https://doi.org/10.1007/s11069-009-9452-6

Messner, F., Penning-rowsell, E., Green, C., Tunstall, S., Veen, A. van der, Tapsell, S., Wilson, T., Krywkow, J., Logtmeijer, C., Fernández-bilbao, A., Geurts, P., & Haase, D. (2007). Evaluating flood damages: guidance and recommendations on principles and methods. *Flood Risk Management: Hazards, Vulnerability and Mitigation Measures*.

Mickens, J., Szummer, M., & Narayanan, D. (2007). Snitch: Interactive decision trees for troubleshooting misconfigurations. *2nd Workshop on Tackling Computer Systems Problems with Machine Learning Techniques, SysML 2007, Co-Located with NSDI 2007*.

Mitchell-Wallace, K. (2022). *Thematic Review: Catastrophe Modelling & Climate Change*. https://www.lloyds.com/news-and-insights/risk-reports/library/catastrophe-modelling-and-climate-change-report

Mitchell-Wallace, K., Jones, M., Hillier, J., & Foote, M. (2017). Natural catastrophe risk management and modelling: A Practitioner’s Guide. In *Wiley-Blackwell*.

Molnar, C. (2019). Interpretable Machine Learning. A Guide for Making Black Box Models Explainable. *Book*.

Moret, S., Codina Gironès, V., Bierlaire, M., & Maréchal, F. (2017). Characterization of input uncertainties in strategic energy planning models. *Applied Energy*, *202*. https://doi.org/10.1016/j.apenergy.2017.05.106

Morris, D. E., Oakley, J. E., & Crowe, J. A. (2014). A web-based tool for eliciting probability distributions from experts. *Environmental Modelling and Software*, *52*. https://doi.org/10.1016/j.envsoft.2013.10.010

Nativi, S., Mazzetti, P., Santoro, M., Papeschi, F., Craglia, M., & Ochiai, O. (2015). Big Data challenges in building the Global Earth Observation System of Systems. *Environmental Modelling and Software*, *68*. https://doi.org/10.1016/j.envsoft.2015.01.017

O'Hagan, A., Buck, C. E., Daneshkhah, A., Eiser, J. R., Garthwaite, P. H., Jenkinson, D. J., Oakley, J. E., & Rakow, T. (2006). Uncertain judgements: Eliciting experts' probabilities. In *Uncertain Judgements: Eliciting Experts' Probabilities*. https://doi.org/10.1002/0470033312

Oommen, T., Baise, L. G., & Vogel, R. M. (2011). Sampling Bias and Class Imbalance in Maximum-likelihood Logistic Regression. *Mathematical Geosciences*, *43*(1). https://doi.org/10.1007/s11004-010-9311-8

Pal, M., & Mather, P. M. (2003). An assessment of the effectiveness of decision tree methods for land cover classification. *Remote Sensing of Environment*, *86*(4), 554–565. https://doi.org/10.1016/S0034-4257(03)00132-9

Paleari, L., & Confalonieri, R. (2016). Sensitivity analysis of a sensitivity analysis: We are likely overlooking the impact of distributional assumptions. *Ecological Modelling*, *340*. https://doi.org/10.1016/j.ecolmodel.2016.09.008

Pappenberger, F., Beven, K. J., Ratto, M., & Matgen, P. (2008). Multi-method global sensitivity analysis of flood inundation models. *Advances in Water Resources*, *31*(1), 1–14. https://doi.org/10.1016/j.advwatres.2007.04.009

Paprotny, D., Kreibich, H., Morales-Nápoles, O., Terefenko, P., & Schröter, K. (2020). Estimating exposure of residential assets to natural hazards in Europe using open data. *Natural Hazards and Earth System Sciences*, *20*(1). https://doi.org/10.5194/nhess-20-323-2020

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). https://doi.org/10.1038/s41467-018-04253-1

Park, S. J., & Lee, D. K. (2020). Prediction of coastal flooding risk under climate change impacts in South Korea using machine learning algorithms. *Environmental Research Letters*, *15*(9). https://doi.org/10.1088/1748-9326/aba5b3

Paulik, R., Wild, A., Zorn, C., & Wotherspoon, L. (2022). Residential building flood damage: Insights on processes and implications for risk assessments. *Journal of Flood Risk Management*, *15*(4).

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*(85), 2825–2830.

Perol, T., Gharbi, M., & Denolle, M. (2018). Convolutional neural network for earthquake detection and location. *Science Advances*, *4*(2). https://doi.org/10.1126/sciadv.1700578

Pianosi, F., Beven, K., Freer, J., Hall, J. W., Rougier, J., Stephenson, D. B., & Wagener, T. (2016). Sensitivity analysis of environmental models: A systematic review with practical workflow. In *Environmental Modelling and Software* (Vol. 79, pp. 214–232). Elsevier Ltd. https://doi.org/10.1016/j.envsoft.2016.02.008

Pianosi, F., & Wagener, T. (2015). A simple and efficient method for global sensitivity analysis based oncumulative distribution functions. *Environmental Modelling and Software*, *67*. https://doi.org/10.1016/j.envsoft.2015.01.004

Pianosi, F., & Wagener, T. (2018). Distribution-based sensitivity analysis from a generic input-output sample. *Environmental Modelling and Software*, *108*. https://doi.org/10.1016/j.envsoft.2018.07.019

Piketty, T., & Zucman, G. (2014). Capital is back: Wealth-income ratios in rich countries 1700-2010. *Quarterly Journal of Economics*, *129*(3). https://doi.org/10.1093/qje/qju018

Pittore, M., Wieland, M., & Fleming, K. (2017). Perspectives on global dynamic exposure modelling for geo-risk assessment. In *Natural Hazards* (Vol. 86). https://doi.org/10.1007/s11069-016-2437-3

Preusser, F. (2008). Characterisation and evolution of the River Rhine system. *Geologie En Mijnbouw/Netherlands Journal of Geosciences*, *87*(1). https://doi.org/10.1017/S0016774600024008

Read, J. S., Jia, X., Willard, J., Appling, A. P., Zwart, J. A., Oliver, S. K., Karpatne, A., Hansen, G. J. A., Hanson, P. C., Watkins, W., Steinbach, M., & Kumar, V. (2019). Process-Guided Deep Learning Predictions of Lake Water Temperature. *Water Resources Research*, *55*(11), 9173–9190. https://doi.org/10.1029/2019WR024922

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., & Prabhat. (2019). Deep learning and process understanding for data-driven Earth system science. *Nature*, *566*(7743), 195–204. https://doi.org/10.1038/s41586-019-0912-1

Ritchie, H., Rosado, P., & Roser, M. (2023). *Our World In Data*.

Rogger, M., Agnoletti, M., Alaoui, A., Bathurst, J. C., Bodner, G., Borga, M., Chaplot, V., Gallart, F., Glatzel, G., Hall, J., Holden, J., Holko, L., Horn, R., Kiss, A., Kohnová, S., Leitinger, G., Lennartz, B., Parajka, J., Perdigão, R., … Blöschl, G. (2017). Land use change impacts on floods at the catchment scale: Challenges and opportunities for future research. In *Water Resources Research* (Vol. 53, Issue 7). https://doi.org/10.1002/2017WR020723

Roscher, R., Bohn, B., Duarte, M. F., & Garcke, J. (2020). Explainable Machine Learning for Scientific Insights and Discoveries. *IEEE Access*, *8*, 42200–42216. https://doi.org/10.1109/ACCESS.2020.2976199

Saint-Geours, N., Grelot, F., Bailly, J. S., & Lavergne, C. (2015). Ranking sources of uncertainty in flood damage modelling: A case study on the cost-benefit analysis of a flood mitigation project in the Orb Delta, France. *Journal of Flood Risk Management*. https://doi.org/10.1111/jfr3.12068

Sairam, N., Brill, F., Sieg, T., Farrag, M., Kellermann, P., Nguyen, V. D., Lüdtke, S., Merz, B., Schröter, K., Vorogushyn, S., & Kreibich, H. (2021). Process-Based Flood Risk Assessment for Germany. *Earth's Future*, *9*(10). https://doi.org/10.1029/2021EF002259

Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M., & Tarantola, S. (2008). Global sensitivity analysis: The primer. In *Global Sensitivity Analysis: The Primer*. https://doi.org/10.1002/9780470725184

Sampson, C. C., Fewtrell, T. J., O'Loughlin, F., Pappenberger, F., Bates, P. B., Freer, J. E., & Cloke, H. L. (2014). The impact of uncertain precipitation data on insurance loss estimates using a flood catastrophe model. *Hydrology and Earth System Sciences*, *18*(6), 2305–2324. https://doi.org/10.5194/hess-18-2305-2014

Samuel, A. L. (1959). Some studies in machine learning using the game of checkers. *IBM Journal of Research and Development*, *3*(3), 210–229. https://doi.org/10.1147/rd.33.0210

Sarazin, F. (2018). *Understanding the sensitivity of karst groundwater recharge to climate and land cover changes at a large-scale*. University of Bristol.

Sarrazin, F. (2018). *Understanding the sensitivity of karst groundwater recharge to climate and land cover changes at a large-scale*. University of Bristol.

Sawicz, K. A., Kelleher, C., Wagener, T., Troch, P., Sivapalan, M., & Carrillo, G. (2014). Characterizing hydrologic change through catchment classification. *Hydrology and Earth System Sciences*, *18*(1), 273–285. https://doi.org/10.5194/hess-18-273-2014

Schanze, J. (2012). Dealing with future change in flood risk management. In *Journal of Flood Risk Management* (Vol. 5, Issue 1). https://doi.org/10.1111/j.1753-318X.2011.01129.x

Schendel, T., & Thongwichian, R. (2017). Considering historical flood events in flood frequency analysis: Is it worth the effort? *Advances in Water Resources*, *105*. https://doi.org/10.1016/j.advwatres.2017.05.002

Schumann, G. J. P., Bates, P. D., Neal, J. C., & Andreadis, K. M. (2014). Technology: Fight floods on a global scale. In Nature (Vol. 507, Issue 7491). https://doi.org/10.1038/507169e

Shoji, D., Noguchi, R., Otsuki, S., & Hino, H. (2018). Classification of volcanic ash particles using a convolutional neural network and probability. *Scientific Reports*, *8*(1). https://doi.org/10.1038/s41598-018-26200-2

Shortridge, J. E., Guikema, S. D., & Zaitchik, B. F. (2016). Machine learning methods for empirical streamflow simulation: A comparison of model accuracy, interpretability, and uncertainty in seasonal watersheds. *Hydrology and Earth System Sciences*, *20*(7). https://doi.org/10.5194/hess-20-2611-2016

Sieg, T., & Thieken, A. H. (2022). Improving flood impact estimations. *Environmental Research Letters*, *17*(6).

Singh, R., Wagener, T., Crane, R., Mann, M. E., & Ning, L. (2014). A vulnerability driven approach to identify adverse climate and land use change combinations for critical hydrologic indicator thresholds: Application to a watershed in Pennsylvania, USA. *Water Resources Research*, *50*(4). https://doi.org/10.1002/2013WR014988

Sinha, A. P., & Zhao, H. (2008). Incorporating domain knowledge into data mining classifiers: An application in indirect lending. *Decision Support Systems*, *46*(1). https://doi.org/10.1016/j.dss.2008.06.013

Sit, M., Demiray, B. Z., Xiang, Z., Ewing, G. J., Sermet, Y., & Demir, I. (2020). A comprehensive review of deep learning applications in hydrology and water resources. *Water Science and Technology*, *82*(12). https://doi.org/10.2166/wst.2020.369

Slater, L. J., & Villarini, G. (2016). Recent trends in U.S. flood risk. *Geophysical Research Letters*, *43*(24). https://doi.org/10.1002/2016GL071199

Smirnov, N. V. (1939). On the estimation of the discrepancy between empirical curves of distribution for two independent samples. *Bull. Math. Univ. Moscou*, *2*(16).

Smith, A., Bates, P. D., Wing, O., Sampson, C., Quinn, N., & Neal, J. (2019). New estimates of flood exposure in developing countries using high-resolution population data. *Nature Communications*, *10*(1), 1814. https://doi.org/10.1038/s41467-019-09282-y

Solomatine, D. P., & Ostfeld, A. (2008). Data-driven modelling: Some past experiences and new approaches. *Journal of Hydroinformatics*, *10*(1). https://doi.org/10.2166/hydro.2008.015

SOLOMATINE, D. P., & SIEK, M. B. L. A. (2004). FLEXIBLE AND OPTIMAL M5 MODEL TREES WITH APPLICATIONS TO FLOW PREDICTIONS. In *Hydroinformatics*. https://doi.org/10.1142/9789812702838_0212

Stein, L., Pianosi, F., & Woods, R. (2020). Event-based classification for global study of river flood generating processes. *Hydrological Processes*, *34*(7). https://doi.org/10.1002/hyp.13678

Sun, Z., Sandoval, L., Crystal-Ornelas, R., Mousavi, S. M., Wang, J., Lin, C., Cristea, N., Tong, D., Hawley Carande, W., Ma, X., Rao, Y., A. Bednar, J., Tan, A., Wang, J., Purushotham, S., E. Gill, T., Chastang, J., Howard, D., Holt, B., … John, A. (2022). A review of Earth Artificial Intelligence. *Computers and Geosciences*, *159*. https://doi.org/https://doi.org/10.1016/j.cageo.2022.105034

Tanoue, M., Hirabayashi, Y., & Ikeuchi, H. (2016). Global-scale river flood vulnerability in the last 50 years. *Scientific Reports*, *6*. https://doi.org/10.1038/srep36021

Teoh, S. T., & Ma, K. L. (2003). PaintingClass: Interactive construction, visualization and exploration of decision trees. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 667–672. https://doi.org/10.1145/956750.956837

Thanh-Nghi Do. (2006). Towards simple, easy-to-understand, an interactive decision tree algorithm. *9th National Conference in Computer Science*.

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). https://doi.org/10.1029/2005WR004177

Thomas Steven Savage, J., Pianosi, F., Bates, P., Freer, J., & Wagener, T. (2016). Quantifying the importance of spatial resolution and other factors through global sensitivity analysis of a flood inundation model. *Water Resources Research*, *52*(11), 9146–9163. https://doi.org/10.1002/2015WR018198

Tollefson, J. (2022). Climate change is hitting the planet faster than scientists originally thought. *Nature*. https://doi.org/10.1038/d41586-022-00585-7

Toonen, W. H. J. (2015). Flood frequency analysis and discussion of non-stationarity of the Lower Rhine flooding regime (AD 1350-2011): Using discharge data, water level measurements, and historical records. *Journal of Hydrology*, *528*. https://doi.org/10.1016/j.jhydrol.2015.06.014

Trigg, M. A., Birch, C. E., Neal, J. C., Bates, P. D., Smith, A., Sampson, C. C., Yamazaki, D., Hirabayashi, Y., Pappenberger, F., Dutra, E., Ward, P. J., Winsemius, H. C., Salamon, P., Dottori, F., Rudari, R., Kappes, M. S., Simpson, A. L., Hadzilacos, G., & Fewtrell, T. J. (2016). The credibility challenge for global fluvial flood risk analysis. *Environmental Research Letters*, *11*(9). https://doi.org/10.1088/1748-9326/11/9/094014

United Nations. (2005). Hyogo Framework for Action (HFA) 2005-2015 Building the Resilience of Nations and Communities to Disasters. *United Nations*, *January*. https://doi.org/10.1017/CBO9781107415324.004

United Nations. (2015). *Sendai Framework for Disaster Risk Reduction 2015 - 2030*. https://www.undrr.org/publication/sendai-framework-disaster-risk-reduction-2015-2030

van den Elzen, S., & van Wijk, J. J. (2011). BaobabView: Interactive construction and analysis of decision trees. *VAST 2011 - IEEE Conference on Visual Analytics Science and Technology 2011, Proceedings*, 151–160. https://doi.org/10.1109/VAST.2011.6102453

van Werkhoven, K., Wagener, T., Reed, P., & Tang, Y. (2008). Characterization of watershed model behavior across a hydroclimatic gradient. *Water Resources Research*, *44*(1). https://doi.org/10.1029/2007WR006271

Vorogushyn, S., Bates, P. D., Debruijn, K., Castellarin, A., Kreibich, H., Priest, S., Schroter, K., Bagli, S., Bloschl, G., Domeneghetti, A., Gouldby, B., Klijn, F., Lammersen, R., Neal, J. C., Ridder, N., Terink, W., Viavattene, C., Viglione, A., Zanardo, S., & Merz, B. (2018). Evolutionary leap in large-scale flood risk assessment needed. *Wiley Interdisciplinary Reviews: Water*, *5*(2). https://doi.org/10.1002/WAT2.1266

Wagenaar, D. J., de Bruijn, K. M., Bouwer, L. M., & de Moel, H. (2016). Uncertainty in flood damage estimates and its potential effect on investment decisions. *Natural Hazards and Earth System Sciences*, *16*(1). https://doi.org/10.5194/nhess-16-1-2016

Wagener, T., & Pianosi, F. (2019). What has Global Sensitivity Analysis ever done for us? A systematic review to support scientific advancement and to inform policy-making in earth system modelling. In *Earth-Science Reviews* (Vol. 194). https://doi.org/10.1016/j.earscirev.2019.04.006

Wagener, T., Reinecke, R., & Pianosi, F. (2022). On the evaluation of climate change impact models. *Wiley Interdisciplinary Reviews: Climate Change*, *13*(3). https://doi.org/10.1002/wcc.772

Ward, P. J., Jongman, B., Salamon, P., Simpson, A., Bates, P., de Groeve, T., Muis, S., de Perez, E. C., Rudari, R., Trigg, M. A., & Winsemius, H. C. (2015). Usefulness and limitations of global flood risk models. In *Nature Climate Change* (Vol. 5, Issue 8). https://doi.org/10.1038/nclimate2742

Washington, W. M., Buja, L., & Craig, A. (2009). The computational future for climate and Earth system models: On the path to petaflop and beyond. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *367*(1890), 833–846. https://doi.org/10.1098/rsta.2008.0219

Wing, O. E. J., Pinter, N., Bates, P. D., & Kousky, C. (2020). New insights into US flood vulnerability revealed from flood insurance big data. *Nature Communications*, *11*(1). https://doi.org/10.1038/s41467-020-15264-2

Winter, B., Schneeberger, K., Huttenlau, M., & Stötter, J. (2018). Sources of uncertainty in a probabilistic flood risk model. *Natural Hazards*, *91*(2). https://doi.org/10.1007/s11069-017-3135-5

World Pop. (2020). *World Pop Hub*. https://www.worldpop.org/

Wünsch, A., Herrmann, U., Kreibich, H., & Thieken, A. H. (2009). The role of disaggregation of asset values in flood loss estimation: A comparison of different modeling approaches at the mulde river, Germany. *Environmental Management*, *44*(3). https://doi.org/10.1007/s00267-009-9335-3

Zanardo, S., & Salinas, J. L. (2022). An introduction to flood modeling for catastrophe risk management. *Wiley Interdisciplinary Reviews: Water*, *9*(1). https://doi.org/10.1002/wat2.1568

Zhou, L., Pan, S., Wang, J., & Vasilakos, A. v. (2017). Machine learning on big data: Opportunities and challenges. *Neurocomputing*, *237*, 350–361. https://doi.org/10.1016/j.neucom.2017.01.026

Zounemat-Kermani, M., Batelaan, O., Fadaee, M., & Hinkelmann, R. (2021). Ensemble machine learning paradigms in hydrology: A review. In *Journal of Hydrology* (Vol. 598). https://doi.org/10.1016/j.jhydrol.2021.126266