Master’s Degree Programme in
Environmental Sciences
D.M. 270/2004

Final Thesis

A Bayesian network approach for multi-sectoral flood damage assessment and multi-scenario analysis

Supervisor
Ch. Prof. Andrea Critto

Co-tutor
Dr. Elisa Furlan

Graduand
Remi Harris
Matriculation Number 877147

Academic Year
2019/2020

Contents

LIST OF FIGURES

LIST OF TABLES

SUMMARY

Extreme weather events, from river flooding to droughts and tropical cyclones, are likely to become both more severe and more common in the coming years, resulting from drivers such as climate change and mismanagement of natural resources.

The damages caused by these extreme events will be felt across all sectors of society, particularly in the form of economic losses to productivity and physical assets. In the face of this threat, policy- and decision-makers are increasingly calling for new approaches and tools to support risk management and climate adaptation pathways that are able to capture the full extent of multi-sectoral damages.

Beginning with a review of the state-of-the art literature concerning Machine Learning (ML) methods, this Thesis builds a GIS-based Bayesian Network (BN) approach that is capable of capturing and modelling multi-sectoral flooding damages against future '*what-if*' scenarios (e.g. changes in land use/cover and flooding hazard patterns). In doing so the work addresses key knowledge gaps in the literature, while providing support and additional insight for decision making underpinning the definition of strategies and policies for Disaster Risk Reduction and Management.

To enable this approach, a risk-based conceptual framework was developed in line with the current IPCC definitions, that highlights pathways of interaction between hazard, exposure, and vulnerability indicators related to flooding events, and the subsequent damages to the agricultural, industrial, and residential sectors. Building on this framework, the BN model was then trained and validated by exploiting data collected from the 2014 Secchia River flooding event, as well as other variables selected for the case study area. Moreover, in this stage, a novel approach to defining the structure of the BN was performed, reconfiguring the model according to expert judgment and data-based validation.

During the validation process, the finalized BN model showed a good predictive capacity for flooding damages in the agricultural, industrial and residential sectors, predicting the severity of damages with a classification accuracy of about 60% for each of these assessment endpoints. Model accuracy would be improved with additional and more detailed observational data, as indicated in the validation stage showing as the range of uncertainty in the model prediction for each sector increases with decreasing quantities of data. However the model still performs successfully despite the constraints within the available dataset. This

gives a greater picture of multi-sectoral flooding damages than that found in the current BN-related literature, which focuses almost exclusively on the residential sector. In support of this, a sensitivity analysis was completed, examining the relative importance of the explanatory BN variables (e.g. area of reported damages, land use, and flooding hazard) towards the multi-sectoral damages, and highlighting those that should therefore be prioritized in a more detailed data collection, allowing for better capturing and modelling of multi-sectoral damages against future flooding events.

Finally, based on the finalized BN model, '*what-if*' scenario analysis was performed to understand the potential impacts of future changes in i) land use patterns, as well as ii) increasing flood depths resulting from more severe flood events. The output of the model showed a rising probability of experiencing large monetary damages under both scenarios, representing the increased damages likely to be faced in the future as extreme events become more common and as under the projected land development in the case study area.

In summary, this Thesis presents a full narrative for the analysis of multi-sectoral flooding damages through a BN approach, not only offering a predictive damage model, but also describing its construction, as well the potential applications of the methodology for risk assessment purposes. In spite of constraints within the case study dataset, the results of the appraisal show good promise, and together with the designed BN model itself represent a valuable support for Disaster Risk Management and Reduction against extreme river flooding events, enabling more effective and reliable decision making.

OBJECTIVES AND MOTIVATIONS

Extreme weather events, from river flooding to droughts and tropical cyclones, pose a significant threat to communities across the world, in terms of economic losses to production and damages to physical assets, as well as wider societal losses to people and the environment (UNFCCC, 2020). The risks presented by these events are likely to become both more severe and more common in the coming years as a result of climate change and anthropic over-exploitation of natural resources, further worsening the potential impacts (IPCC, 2018).

In the face of this rising threat, in order to reduce damages, policy- and decision-makers require new integrated approaches and tools to support risk management and climate adaptation pathways, particularly to understand the nuances of damages across different sectors and under multiple possible scenarios. In terms of flood damage, where stage-damage curve models have previously been commonly used (Amadio et al., 2016; Jongman et al., 2012; Thieken et al., 2009; Wing et al., 2020), they are increasingly being replaced by more sophisticated models, drawing from expert knowledge, increased variable input, and more powerful computational methods, including those exploiting functionalities offered by Machine Learning (ML). Among these, upcoming Bayesian Network (BN) approaches (Schröter et al., 2014; Wagenaar et al., 2019, Paprotny et al., 2020), represent useful methods able to integrate heterogeneous data sources, allowing for predictive hazard and impact assessments based on real-world training dataset.

Working in collaboration with the Fondazione Centro Euro-Mediterraneo sui Cambiamenti Climatici (CMCC)[1], this thesis will contribute to the goals of the European Commission-funded LODE (Loss Data Enhancement) project[2], which aims to optimize the analytical tools and procedures used for damage data collection and processing, in order to support the development of policies and strategies for Climate Change Adaptation and Disaster Risk Reduction.

Accordingly, this thesis aims to build on the state-of-the-art research in the field of ML, through the design and application of a BN approach able to capture and model multi-sectoral

---

[1] https://www.cmcc.it
[2] https://www.lodeproject.polimi.it

damages against multiple '*what-if'* scenarios, exploiting damage data collected against the 2014 Secchia river flooding event as training dataset of the BN model.

To satisfy the full motivations of the work, the practical implementation of this approach will involve the following operational phases:

- An initial critical analysis of state-of-the-art BN approaches (within the wider family of ML-based models) for disaster damage assessments and modelling will be performed, discussing and comparing the current approaches, with particular attention paid to studies dealing with river flood-related hazards.
- In turn, this preliminary review will pave the way for the design of a novel BN model allowing multi-sectoral damage assessment under possible future scenarios.
- The collection and pre-processing of relevant hazard, exposure, and vulnerability GIS-based data, used to effectively train, calibrate and validate the BN model for application, as well as the projected data (e.g. river flood and land-use/cover projections from EU data portals) necessary to shape potential future scenarios to be investigated through the BN.
- Optimal model design will be assessed through a two-tiered approach to model configuration and validation, in order to finalize the BN for application for multi-sectoral damage assessment in the investigated area. These applications will include a sensitivity analysis allowing for the identification of the most influential set of variables.
- BN-based future '*what-if*' scenario analysis, for better understanding the potential multi-sectorial damages against several flood scenarios under different return periods, as well as changes in the land use/cover.

The methodology as presented, and the designed BN model itself, will represent a valuable support for disaster risk management and reduction against extreme river flooding events. Particularly, it will provide a more complete picture on the multi-sectoral impact of flood-related events across the agricultural, residential, and industrial sectors. Moreover, the examination of a variety of '*what-if'* future scenarios will further support these aims in the analysis of how expected variations in the magnitude of flood-related events, as well as changes in the land use will translate into more severe future impacts.

THESIS STRUCTURE

This Thesis is structured in three main sections: **Section A** provides an overall picture of the theoretical background underpinning the design and implementation of BN approaches for damage assessment and modelling; **Section B** describes the BN model as designed for the specific case study, including the collection and processing of all necessary data for the model training; **Section C** presents the results of the application of the proposed BN to the case study area of the Secchia river basin (Italy), with the main findings from the scenario analysis.

More specifically;

Section A introduces the main concepts of BN approaches, firstly within the wider ML family, before focusing on their specific key functionalities and methodological phases for their design and implementation. A review of the main approaches currently seen leads into a systematic review of the key relevant papers in the area, highlighting their strengths and weaknesses. Clarification of the main knowledge gaps addressed in the research are further discussed.

Section B describes the case study area, beginning with the wider context of the Po river basin, before more specifically targeting the Secchia basin, with a detailed description of the flooding event that occurred in 2014. It also provides details on the dataset collected to inform the BN model, as well as the processing methods used to prepare the data for the BN training. Further, the risk-based conceptual framework is presented, systematically identifying pathways of interaction between climate-related drivers and damages on the exposed sectors. Finally, the approach and operative steps for the application of the BN for '*what*-if' scenario analysis are introduced.

Section C describes and critically analyses the results of the BN methodology applied in the case study of the Secchia Basin for multi-sectoral damage assessment, including a range of statistics summarizing key damage-related metrics useful to drive disaster risk management and adaptation pathways in the analyzed area. Results of the BN model construction and configurations' testing are detailed and discussed, alongside the analysis performed on parameter sensitivity and future scenarios.

The **Conclusions** provide a summary of the main findings of the Thesis. Discussion of the results give an overview of the key study outcomes, highlighting the strengths and weaknesses of the approach, and also identifying areas of the research that could potentially further built upon.

SECTION A: THEORETICAL BACKGROUND

1. Introduction to Machine Learning and Bayesian Network approaches

Machine Learning (ML) is a branch of Artificial Intelligence that uses computer algorithms to learn previously unknown or unattainable patterns from various data inputs, performing tasks that would otherwise be too difficult or slow to compute (Alpaydin, 2020). It is an up-and-coming technique within the world of Disaster Risk assessment and Management (DRM), seeing an increasing use in all phases of risk evaluation (Deparday et al., 2019; GFDRR, 2018; Wagenaar et al., 2019).

With rapidly improving technological advances, ML-based methods give the opportunity to not only quickly handle large amounts of data (also known as 'big data'), but also learn from them in real time, capturing hidden patterns in the training data and extracting the required information for statistical predictions (Al-Jarrah et al., 2015). Damage analysis, in particular, can exploit methods under the branch of supervised ML, where a real-life output (i.e. class labels) is available for the training of the model, and it is possible to approximately map the input variables to their corresponding outputs (Ayodele, 2010). The learnt mapping function can then be used for predictive purposes with new input data. In contrast, unsupervised learning methods have no available user-provided labelled data to learn from, and as such are driven by the statistical patterns inherent within the input data (Kotsiantis et al., 2007).

Two major groups of supervised Machine Learning can currently be identified: i) regression methods such as Random Forest, and ii) classification methods that include Bayesian Network approaches (Deparday et al., 2019). These differ in terms of the output variables given by the model. Specifically, classification techniques compute discrete outputs, and can be used, for example, for the detection of inundated areas after flooding events through the classification of satellite images (Lamovec et al., 2013; Lin et al., 2019); while regression techniques provide continuous outputs and find wider application in the prediction of damages (Mayfield et al., 2018; Menderes et al., 2015; Wagenaar et al., 2017).

With the main aim of identifying the potential contribution and support of ML methods to damage modelling and assessment and better understanding the wider context and trends of this field of research, a literature review was performed. The Scopus database[3] was searched

[3] https://www.scopus.com/home.uri

by using two search strings, specifically i) ["flood*" AND "damage*" AND "machine learning"]; and ii) ["flood*" AND "risk*" AND "machine learning"]. Several potential methodological approaches were prevalent, including Bayesian Networks, Decision Trees, Random Forest (e.g. application of Random Forest and Decision Tree approaches to damage prediction; Sadler et al., 2018; Spekkers et al., 2014; Wang et al., 2015) and Artificial Neural Networks (e.g. for flood forecasting; Campolo et al., 2003). Upon detailed analysis of the various searches performed, Bayesian Networks (BN) were identified as an increasingly common tool for state-of-the art damage assessments and modelling (e.g. Schröter et al., 2014; Sperotto et al., 2017; Wagenaar et al., 2017). The graph below shows the growing rate of scientific production in this area, and highlights a significant increase in output in the last ten years.

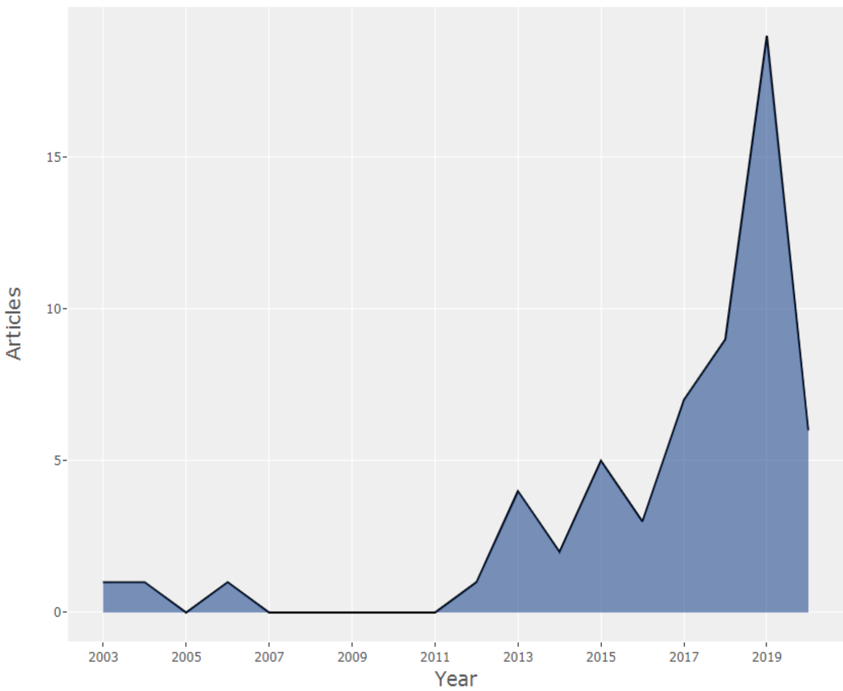


Figure 1: Annual scientific production referencing Bayesian methods within the context of flood assessment

BN are statistical approaches built in the form of qualitative structures known as directed acyclic graphs (DAGs) representing the variables of concern as nodes on the graph, with arcs to represent the probabilistic dependencies among variables within the system. Parameterization of the network then encodes marginal and conditional probabilities of the variables (Furlan et al., 2020; Sperotto et al., 2017).

BN have been noted for their ability to integrate heterogeneous data sources, that may include some inputs based on quantitative data, but also some that are classified qualitatively using expert judgement or by incorporating stakeholders' perspectives (Sperotto et al., 2017). These kind of methods can be designed to tackle complex environmental issues featured by non-linear behaviour and hampered by large uncertainties (Sperotto et al., 2017).

The construction of a BN model follows a stepwise progression (Deparday et al., 2019; Furlan et al., 2020; Sperotto et al., 2017). After the objectives of the model have been defined and the application contextualized, the model must then be **conceptualized** (i.e. BN conceptual model design) in order to ascertain the causal relationships between the system's components. This is often aided with the opinion of experts in the relevant fields of knowledge. From there, the model is then **parametrized** by assigning states to each of the variables, and the marginal and conditional probabilities of the system components are computed. Following this, various forms of model **validation** can be performed. Comparison with a validation dataset is used to check for accuracy; this can be from another subsection of the data, selected randomly for the original dataset, or potentially from another dataset extracted from a different location that can be used to assess the transferability of the model to other events (Wagenaar et al., 2017). Modification of the variables used or of the training set can be used to rerun the model and strengthen its learning capability, increasing its accuracy for prediction. **Sensitivity analysis** can then be performed to measure the sensitivity of the results to changes in the inputs or parameters (Furlan et al., 2020). Once training, calibration, and validation of the BN model have been performed, it is then possible to use the designed model for '*what-if*' **scenario analysis,** defining representative possible alternative scenarios with specific characteristics (e.g. the projected impacts of climate change under different Representative Concentration Pathways), assessing the relative changes in terms of posterior probabilities (for relevant variables) against the set changes in input.

The specific details of these developmental steps are presented in the Section 3.2 detailing the methodological approach underpinning the BN development for the case study of concern.

1.1 Review of BN application for flood risk and damage assessment

Building on the results of the previous literature search for BN approaches to flood damage assessment and modelling, a detailed review of selected '*key papers*' was undertaken, in order to better understand the latest state-of-the-art techniques and applications of BN for flood risk and damage appraisal.

Specifically, from the original search of ML methods, 59 papers were related to BN approaches, and those that were pertinent to the aims of the LODE project were identified, eliminating those not properly focusing on damage assessment or presenting datasets that were significantly different to those collected for the project. Specifically, as a relatively new field of research, few of the papers directly applied BN for damage assessment. Most of the research into similar applications in flood modelling for DRM has focused mainly on hazard modelling, identifying for example the geographical extent of a flooding event rather than investigating the resulting damages across sectors (D'Addabbo et al., 2016; Rosser et al., 2017).

Applications to damage modelling emerged as a young but growing branch of research, and of these, eight 'key papers' representing the current research trends within flood risk and damage assessments were selected. The majority of these papers considered the possibility of predicting residential damages (either monetarily or as a relative loss to the building), training the model using either a random subsection of the wider data for local prediction, or data from one flood event to build a model for prediction of events in other locations. Table 1 reports these '*key papers*', including an overview of their main objectives and the applied methodology, as well as the data used to train the constructed BN models.

Table 1 Summary of key papers identified for analysis

Key Paper ID	Reference	Objectives	Methodology	Damage data	Data Source
1	Schröter et al. (2014)	Predict relative damages to residential buildings; Spatial and temporal transfer to assess prediction of other flood-related events	BN models (expert driven, and data driven); Regression trees	Relative loss to residential buildings	Telephone survey
2	Notaro et al. (2014)	Retrospective assessment of mitigation options for increased resilience	BN Decision Models	Relative loss to residential buildings	Insurance companies
3	Balbi et al. (2016)	Vulnerability model to assess benefits of the implementation of an early-warning system	BN model	N/A	GIS –based data
4	Wagenaar et al. (2017)	Prediction of damages using a limited dataset enriched with additional variables	BN model; Random Forest; Regression trees	Household contents and structure	Damage claims to government
5	Wagenaar et al. (2018)	Regional and temporal transferability to assess prediction of other flood-related events; Comparison of multiple ML-based models;	BN model; Random forest	Relative loss to residential buildings; Household contents and structure	Telephone survey; Damage claims to government
6	Sairam, Schröter, Rözer, et al. (2019)	Improve spatio-temporal transferability over previous models to assess prediction of other flood-related events;	Hierarchical BN Model	Relative loss to residential buildings	Telephone survey
7	Sairam, Schröter, Lüdtke, et al. (2019)	Quantify vulnerability reduction resulting from differing levels of private precaution	BN model; Regression trees	Relative loss to residential buildings	Telephone survey
8	Paprotny, Kreibich, Morales-Nápoles, Wagenaar, et al. (2020)	Improved exposure and vulnerability estimation for residential and commercial buildings	BN model	Relative losses to commercial and residential assets	Telephone survey; Damage claims to government

The information necessary for appropriate computation of BN models, with a variety of heterogeneous data, is rarely collected for damage modelling purposes. Combined with privacy issues and the magnitude of datapoints necessary for data-driven ML techniques, data that is detailed enough for these analyses is currently hard to come by. Accordingly, two case studies where this data is available form the basis of the assessment for five out of the eight '*key papers*', either individually or through combination of the available data for both cases. The first case study is set in Germany, where residential damage information for a random sample of local residences was collected via telephone interviews, in relation to five flooding events in the Elbe and Danube catchment areas (Schröter et al., 2014). Data collected includes early warning and precautionary measures, socioeconomic factors, and building characteristics, incorporated with hazard models for the investigated flood-related events. Residential damage data, given in the form of a loss ratio, is the sole assessment endpoint of these studies and related BN model. Specifically, Schröter et al. (2014) examined the German case study to evaluate the possible spatial and temporal transferability of various damage models. These include a traditional depth damage model, as well as regression trees and four different BN models, either data-driven or constructed using expert knowledge, and each composed of either 10 or 28 input variables. Each model was trained using data from one flooding event that occurred in the Elbe region, then cross-validated using a different event located in the same region, and several others in the Danube region. Results showed that increasing complexity of the model through additional variables consistently increased the prediction capacity of the BN models, particularly in a spatial transfer setting. Moreover, among the tested models, BN were the most successful at predicting damages.

The second case study concerns the Netherlands focusing on a single flooding event: the 1993 Meuse flood (Wagenaar et al., 2017), in line with the case study of concern of this Thesis, focusing on the 2014 Secchia flood event. There are also parallels in how data pertaining to damage and exposure was collected, i.e. coming from damage claims submitted to the respective local governments. The available variables are therefore also similar to those for the Secchia river. However in the Dutch case, data was only available for residential losses (as in Germany), and accurate localization was limited as each damage claim could only be georeferenced to postal code, which may contain up to 20 buildings. Specifically, for the Dutch case study, Wagenaar et al. (2017) hoped to improve upon traditional stage-damage curve models (e.g. Jongman et al., 2012; Thieken et al., 2009) by using variables such as flood depth, velocity and duration, as well as household size and building type and age in order to construct BN and Regression Tree models. A subsection of the data available was used to train the models, and the rest for their validation. In this case, due to the limitations in the input dataset presenting an imprecise nature in the localization method, BN prove less successful in prediction.

In the most recent study by Wagenaar et al. (2018), the authors build on the models trained in each of the previous two studies (Schröter et al., 2014; Wagenaar et al., 2017), applying them to both the German and Dutch case studies, looking to improve spatial and temporal transferability, while comparing the relative strength of BN and Random Forest approaches. For BN, the variables were harmonized to be applicable to both datasets, and two models were constructed, one from the Dutch dataset and another using the German data. After that, they were validated using the opposite datasets. BN, in this case, improve the transferability, outperforming the other methods such as Random Forest. The results stress the need to collect heterogeneous data, with the quantity of data points being less important for higher prediction accuracy in damage evaluation.

Research into the German case study was continued by Sairam et al. (2019), who constructed a Hierarchical Bayesian Model. The study attempts to incorporate a generalized model that is capable of modelling various regions and events (e.g. coastal, pluvial and river floods), with a localized one, which could be applied to the specific dataset of another chosen event. In doing so they were able to create a model that more effectively captured spatio-temporal variability in damages.

Of the selected '*key papers*', two do not directly address a full prediction of flood damages in the same way, instead performing some form of scenario analysis using a simplified BN model. Among these, Notaro et al. (2014) used BN to investigate several possible mitigation scenarios for a previous flooding event that occurred in Palermo, by retroactively applying different water management options. Sairam, Schröter, Lüdtke, et al. (2019) attempted to capture the difference in vulnerability of households to flooding damage, relative to their level of private precaution (e.g. installation of flood barriers or sealing of exposed areas such as basements). Using the German case study, BN were found to be the most effective models for understanding this variable.

Balbi et al. (2016) applied a BN model to assess the difference in flood risk with different availability and reliability of early warning measures, with a focus on human risk rather than monetary damages. In contrast to the other studies which simultaneously involved all collected variables within the BN, they separately compartmentalized all exposure, hazard, and vulnerability variables to produce a risk-based framework for the network.

Among the ‘key papers’, the most sophisticated model, built from the German and Dutch datasets, comes from Paprotny et al. (2020). Specifically, within this study, a more detailed estimation of the exposure to buildings is undertaken, and vulnerability is modelled separately through two different BN models, one covering residential assets and the other considering commercial assets. This framework used both training data concerning pluvial and river flood events, with the intention to have a wide range of potential uses. Moreover, the application of the models to a further case study addressing a coastal flooding event in France, showed good performance, predicting damage losses in the area better than other damage models.

Overall, BN models show a notable added value in this field of research, regularly outperforming both traditional models and other ML-based approaches (e.g. Random Forest). This has been applied in the majority of cases for prediction of residential damages caused by flooding events, trained either on a local scale or to multiple case studies for checking BN performance for transferability.

Among these added values, the biggest factor may be the flexibility of the proposed BN models, as their construction can draw from a wide range of knowledge bases and expertise, and incorporate a wide variety of input data, also giving an insight into the relative importance of the relevant variables. In doing so these BN models can capture uncertainties well, which is an important factor when dealing with disaster risk management and climate change adaptation. Further, this allows for the modelling of multi-faceted aspects related to hazard, exposure and vulnerability patterns, giving a more complete picture of disaster risk and damages compared to traditional models.

1.2 Knowledge gaps

There are a few overarching themes that are, at present, rarely seen in the current literature dealing with BN methods for flood risk and damage assessments.

In terms of sectoral analysis, BN approaches have been employed so far almost exclusively in these contexts for predicting residential damages, with minimal efforts devoted to a wider sectoral analysis that could, for instance, incorporate infrastructure, agricultural or industrial concerns within their assessment endpoints. At most, there was some consideration of flood-related impacts to commercial activities (Balbi et al., 2016; Dominik Paprotny, Kreibich, Morales-Nápoles, Wagenaar, et al., 2020); However for both of these papers, the picture built on the total damages within a case study was incomplete.

Moreover, across the investigated '*key papers'* it was also noted that although some forms of scenario analysis have been undertaken in this context, these are not common. Compared to other applications, these analyses are relatively simplified in the frame of the envisioned alternative scenarios, which may be affecting only one of the input variables and not a combination of multiple. In the cases where scenario analysis was conducted, it was limited to a retrospective investigation into one of the contributing factors of flood damage, most clearly in the investigation of mitigation scenarios by Notaro et al. (2014). As such, a common theme of these evaluations is the lack of future climate and land-use/cover projections. Notably, none of the papers addressed any possible future scenarios where the studied region may face an increased occurrence or severity of flooding events, nor with exacerbated exposure and vulnerability patterns.

In conjunction, there is also limited discussion on the impact of climate change on the results of the reviewed research. Whilst most papers acknowledged that the studies were important for DRM in the context of climate change and increasing natural disasters, there was little acknowledgement of how these ML techniques could be used to specifically investigate these impacts directly.

Also missing from the examined literature were any conceptualized approaches for detailed sensitivity analysis. Some papers touched upon the relative importance of the variables contributing to the training of the models, however this was never a major feature of the research, and no frameworks were elaborated for future use. Neither were there any stepwise approaches for variable integration in the BN model, instead each author chose to introduce all variables together, thus limiting the frame of model analysis.

SECTION B: METHODOLOGICAL DEVELOPMENT

2. Description and characterization of the case study

2.1 Geographical context: the Po and Secchia river basins

The largest and most economically important river basin in Italy is the Po Valley (Tockner et al., 2009), stretching from the Alps in the North of the country to the Adriatic Sea on the East coast. It comprises the Po River basin, the eastern lowlands of Veneto and Friuli and the south-eastern basins of Emilia–Romagna. Nearly half of the national GDP is produced in the basin area from one third the country's industries, including a large agricultural output (Amadio et al., 2019). It is also significantly populated, with 17 million people (about 29% of the Italian population) living within its eight regions (ISTAT, 2011).

The area is traditionally flood-prone (Lombardi et al., 2018), and vulnerability to flooding-related disasters has further increased in recent years, due to the rapid subsidence of the sedimentary Po basin, resulting from both natural and anthropogenic factors such as water withdrawal (Carminati & Martinelli, 2002). In combination with increasing flood peaks and river discharges (Govi & Maraga, 2005), it is therefore important to ensure that all aspects of flood risk management in the area are as effective as possible.

The Secchia river, one of the main tributaries to the Po, flows through Emilia Romagna region, within the south eastern part of the Po basin. The Secchia basin in and of itself covers a catchment area of over 2000 km^2, with 172 km of river flowing from the Apennines. Emilia Romagna has been identified as the most flood prone area of the country (ISPRA, 2014), with a flood exposure under a return period of 20 - 50 years for 10% of the population and 25,000 km^2 of land, and under 100 – 200 years for a further 64% of the population and an area of 10,000 km^2 (Hasanzadeh Nafari et al., 2017). The exposed population of 2.7 million is not only the highest in relative terms, but also in absolute numbers nationwide (ISPRA, 2018).

Italian flood adaptation measures, in line with the EU Flood Directive (EC, 2007), are mandated by the Ministry of the Environment, Land and Sea. Hydrogeological protection laws state that for each hydrological basin, the relevant authorities must detect hazard-prone areas that could be flooded during an event and prevent further additional risk. Among other

adaptation strategies, the implementation of an integrated early warning system was also mandated.
In this setting, protection in the Po river basin is provided by the likes of embankments, hydraulic works, and levees (Zanchettin et al., 2008). Specifically, in the lower part of the Po River, flood-prone areas have been protected by a complex system of embankments and hydraulic works that are part of the wider flood defense system in the Po Valley, extending for almost 3000 km as a result of a tradition of river embanking lasting centuries (Govi & Maraga, 2005). However, continued development of these systems upstream, whilst increasing protection in Upper Po areas have, according to some models, further exacerbated the vulnerability of the lower basin regions by increasing the value of the flood peak at any given probability (Zanchettin et al., 2008). Protection measures such as these have also been linked to a low risk awareness and heightened false sense of security among residents (Amadio et al., 2019).

2.2 The 2014 Secchia flooding event

In January 2014, the Secchia river basin was hit by a long-lasting stratiform rainfall event, which has led to significant stress on the levees of the local rivers. Specifically, on January 19th, a major flooding event occurred when these conditions led to 1m of water breaching the artificial levee protecting the surrounding area from the Secchia river; a portion of this levee also collapsed, leading to additional inundation of the surrounding plain.
This breach occurred by the town of San Matteo, close to the city of Modena with an overall population of 184,000. The affected area included the municipalities of Bastiglia and Bomporto, as well as a smaller area of the municipality of Modena, at a total of 122 $km^{2,}$ bordered to the west by the Secchia river and to the east by the Panaro river (Figure 2). The vulnerability of the area is high as the land is mostly flat, with relief coming only from minor levees as well as road and railway embankments (Carisi et al., 2018).
A similar potential levee failure was later spotted on the Panaro and fixed, however the situation in Secchia could not be averted (Orlandini et al., 2015).

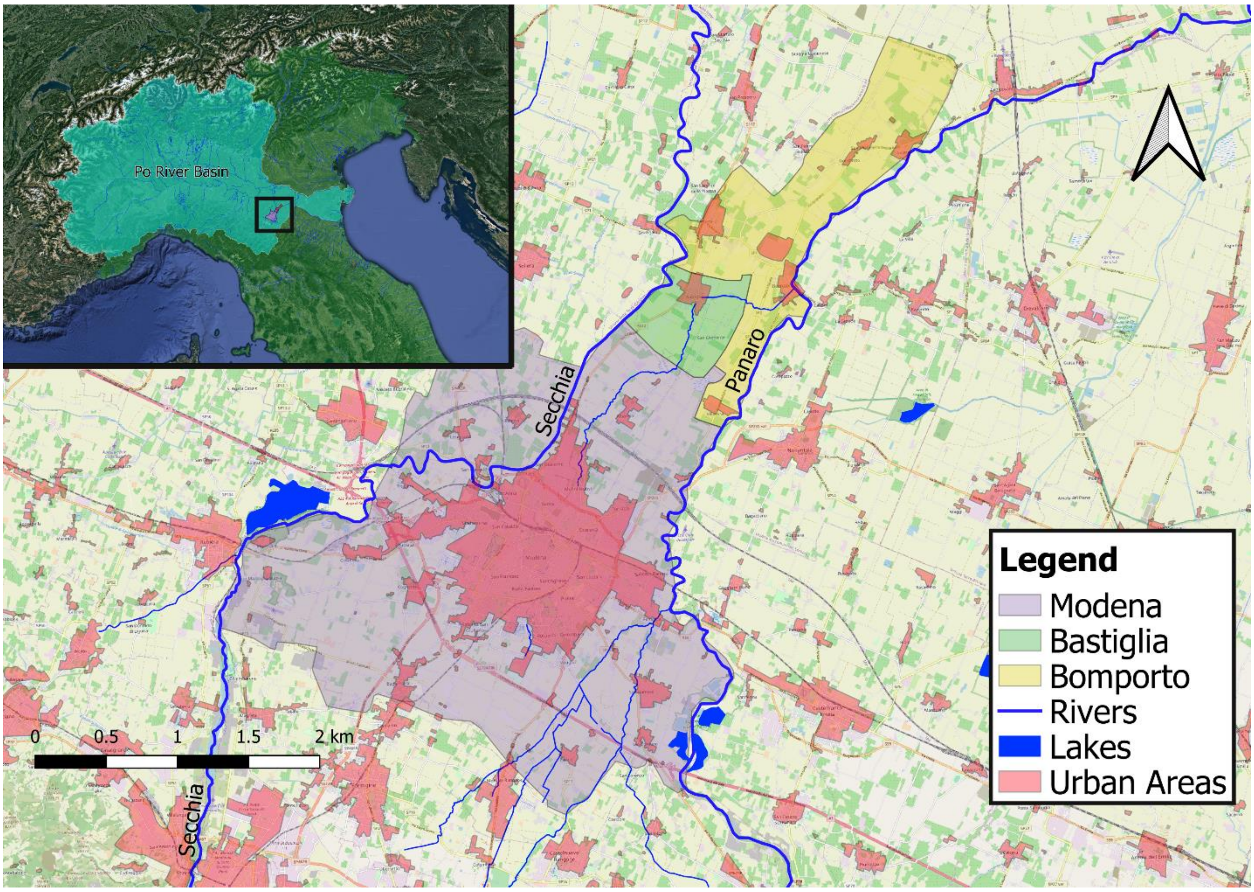


Figure 2: Map of the municipalities affected by the 2014 flood, and their location within the Po Basin

The impacts of the flood were widespread and devastating for many. Over 50km^2 was flooded at an average depth of 1m, displacing thousands and causing one fatality (Carisi et al., 2018). The towns of Bastiglia and Bomporto felt the impacts hardest, with the duration of flooding exceeding two full days and a volume of inundation of over 30 million cubic meters. The economic damage was particularly severe, estimated at least €500 million, including €36 million worth of damage to residential properties, according to damage declaration (Amadio et al., 2019; Orlandini et al., 2015). A significant proportion of the flood area was rural agricultural land, with crops including wheat, maize and forage, as well as some vineyards. Losses to the agricultural sector were fortunately minimized as most of the crops were in a vegetative state, however the losses have still been estimated at a value of 343 € / ha (Amadio et al., 2016).

Figure 3: Aerial image of the 2014 flooding event in the municipality of Bastiglia. 2014; Web reference: https://www.lapressa.it/notiziario/la_provincia/alluvione-bastiglia-ecco-lennesimo-ente-inutile

Figure 4: Image of the 2014 flooding event in the municipality of Bomporto 2014; Web reference: http://www.sassuolo2000.it/2014/02/07/alluvione-di-bomporto-una-tortellata-di-solidarieta-al-bocciodromo-di-casalgrande/

2.3 Case study data collection

A significant volume of data relating to the case study has been collected, both in terms of the event that occurred around the Secchia river in the 2014, and more notably in terms of the damage that was caused and recorded after the event. This includes detailed hazard modelling data and reported damage claims for a selection of affected residential, agricultural, and industrial properties. This is in line with the interpretation of economic loss and damages from climate-related events under the Warsaw International Mechanism (UNFCCC, 2020), comprising damages and losses to physical assets (infrastructure and property) and income (business operations, agricultural production, tourism).

Disaster risk, or the possibility of adverse effects from future disasters, derives from the vulnerabilities of any elements exposed to a physical hazard (IPCC, 2018). Along the framework of the IPCC's guidelines, hazard information should not be considered as a sole proxy for risk. Risk and damage assessments should instead involve the integration of suitable hazard, vulnerability, and exposure components, as well as recorded damage information where possible (IPCC, 2018).

As a result, a large quantity of data encompassing all analytical dimensions is required for informing risk assessment and modelling. For these purposes, there are several other sources of heterogeneous data available to be integrated within the dataset for analysis, particularly regarding the microscale exposure and vulnerability of the affected buildings, assets, and population.

The technical details (metadata) regarding the collected data for the case study of concern are summarized in the table below.

Table 2: Metadata of the dataset available for the implementation of BN approach in the Secchia river basin case study area

Variable	Unit	Data format	Spatial domain	Spatial resolution	Temporal reference	Reference
Hazard data						
Maximum water depth	m	Raster	Local	5m	19/1/14-22/1/14	Vacondio et al. (2016)
Maximum flow velocity	m s^{-1}	Raster	Local	5m	19/1/14-22/1/14	Vacondio et al. (2016)
Duration of inundation	h	Raster	Local	5m	19/1/14-22/1/14	Vacondio et al. (2016)
Exposure data						
Structure area	m^2	Tabular	Local	$1m^2$	2014	Damage claims
Population density	person/km^2	Raster	Europe	1km	2010	Aurambout & Lavalle (2016); https://data.jrc.ec.europa.eu/dataset/jrc-luisa-udp-popden-ref2016
Vulnerability data						
Number of storeys	-	Vector	National	Sub-municipal	2011	ISTAT (2011); https://www.istat.it/it/archivio/104317
Digital Elevation Model	m	Raster	Europe	10m	2011	Tarquini et al. (2007); http://tinitaly.pi.ingv.it/
Property market value	€ m^{-2}	Vector	National	Sub-municipal	2019	https://wwwt.agenziaentrate.gov.it/geopoi_omi/index.php
Conservation status	-	Vector	National	Sub-municipal	2011	ISTAT (2011); https://www.istat.it/it/archivio/104317
Building age	years	Vector	National	Sub-municipal	2011	ISTAT (2011); https://www.istat.it/it/archivio/104317
Land use	-	Raster	Europe	100m	2010	Lavalle (2014); https://data.jrc.ec.europa.eu/dataset/jrc-luisa-land-use-ref-2014
Population age	years	Vector	National	Sub-municipal	2011	ISTAT (2011); https://www.istat.it/it/archivio/104317
Damage data						
Residential damage	€	Tabular	Local	Building level	19/1/14-22/1/14	Damage claims
Agricultural damage	€	Tabular	Local	Building level	19/1/14-22/1/14	Damage claims
Industrial damage	€	Tabular	Local	Building level	19/1/14-22/1/14	Damage claims
Future scenario data						
Land use	-	Raster	Europe	1km	2020-2050	Lavalle (2014); https://data.jrc.ec.europa.eu/dataset/jrc-luisa-land-use-ref-2014
Flood depth at different return period	m	Raster	Europe	100m	10-500 years	Dottori et al. (2016); https://data.jrc.ec.europa.eu/collection/id-0054

2.3.1 Hazard Data

Hazard modelling for the 2014 Secchia flooding event provides information on the extent, depth and duration of the flood, as well as the flow velocity. This modelling has been performed through a combination of 2D hydraulic models and observational data to produce and validate flood hazard maps (Renato Vacondio et al., 2016). The hydraulic simulations were performed using a GPU-parallelized model for the solution of shallow-water equations, that combined topographic and bathymetric floodplain characteristics with numerical river discharge modelling (R Vacondio et al., 2014).

This dataset has been validated against field data and observations, including a high-resolution radar image acquired during the flood event. The output of this modelling as available for use are three 5m spatially resolved hazard metrics, depicting the flood velocity, depth, and duration of inundation respectively, pictured below.

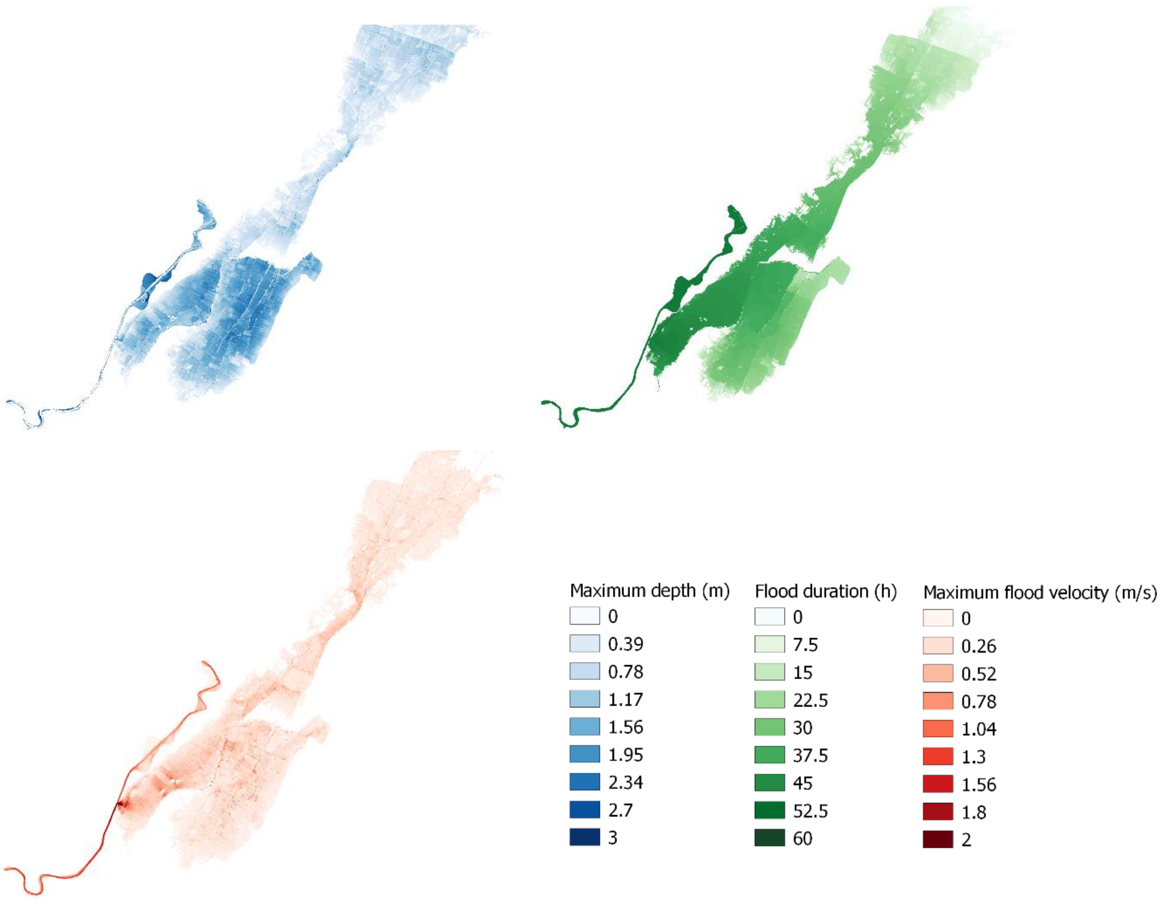


Figure 5: Hazard model outputs for the 2014 Secchia river flooding event

2.3.2 Exposure and Vulnerability Data

As seen in Section 1 additional data on the vulnerability and exposure of hazard-prone people and properties can be used to enhance the predictive capacity of BN models (Dominik Paprotny, Kreibich, Morales-Nápoles, Castellarin, et al., 2020). In the context of disaster risk assessment and management, exposed elements (generally considered as human beings, their livelihoods and assets (IPCC, 2018)) are those placed within an area and that could face some form of hazardous events. The level of exposure of an area where a hazardous event may occur, is thus a necessary component of risk computation.

However, this alone is not enough to fully understand the magnitude of risk. The vulnerability of the exposed elements, or their propensity to suffer adverse impacts from a particular hazardous event, will determine the observed losses and damages (IPCC, 2018). As such, a range of variables for exposed buildings, assets, and population for the residential, industrial, and agricultural sectors in the case study area has been collected, in order to better inform the BN model.

In this setting, the ISTAT 2011 building and population census data (ISTAT, 2011) contains information on building characteristics and population statistics, available at a sub municipal level (i.e. census area). Specifically, extracted for these purposes were building age (constructed before or after 1990), number of storeys, and conservation status (qualitatively categorized as good or bad condition), as well as the number of children under 5 and elderly over 65 per census area, characterizing the coping capacity of potentially affected people.

For each property affected, damage claims collected for the purpose of this study (and the connected LODE project), provide information on the area of the building structure.

Additional data were collected pertaining to the geographical characteristics of the case study area. Specifically, land use/cover data was taken from JRC[4] (Lavalle, 2014); while the Italian National Institute of Geophysics and Volcanology (INGV) provided detailed digital elevation modelling to provide topographical information.

[4] https://ec.europa.eu/jrc/en/luisa

2.3.3 Damage Data

Empirical damage-related information for residential buildings affected, as well as agricultural and industrial properties, has been collected by the local municipalities following the 2014 Secchia flooding event. This was organized in order to gain information pertaining to the restoration needs of public and private properties and goods (Carisi et al., 2018). A total of 2313 claims were collected, of which 85% related to residential damages, with 8% and 7% coming from the industrial and agricultural sectors respectively. In total the damage claims across all three sectors are equal to €71.5 million. The data is structured as represented in the Tables below:

Table 3: Structure of data concerning damage claims to residential buildings

	Building characteristics		Damage claimed (€)			
Address	**Area (m^2)**	**Use**	**Structure**	**Shared structure**	**Content**	**Vehicles**
Via X, 1, Bastiglia	154	Primary	19200	0	2500	1400

Table 4: Structure of the data concerning damage claims for the agricultural and industrial sectors

	Building characteristics		Damage claimed (€)			
Address	**Area (m^2)**	**Sector**	**Structure**	**Land**	**Machinery**	**Stock**
Via X, 1, Bomporto	385	Agriculture	1200	3000	500	2100

All data was collected from three local municipalities, these being the hardest hit municipalities of Bastiglia and Bomporto, as well as the largest city affected, Modena. As some of the claims were incomplete, a total of 1738 data points were used, of which 65% are from Bastiglia, 31% from Bomporto, and the remaining 4% from Modena.

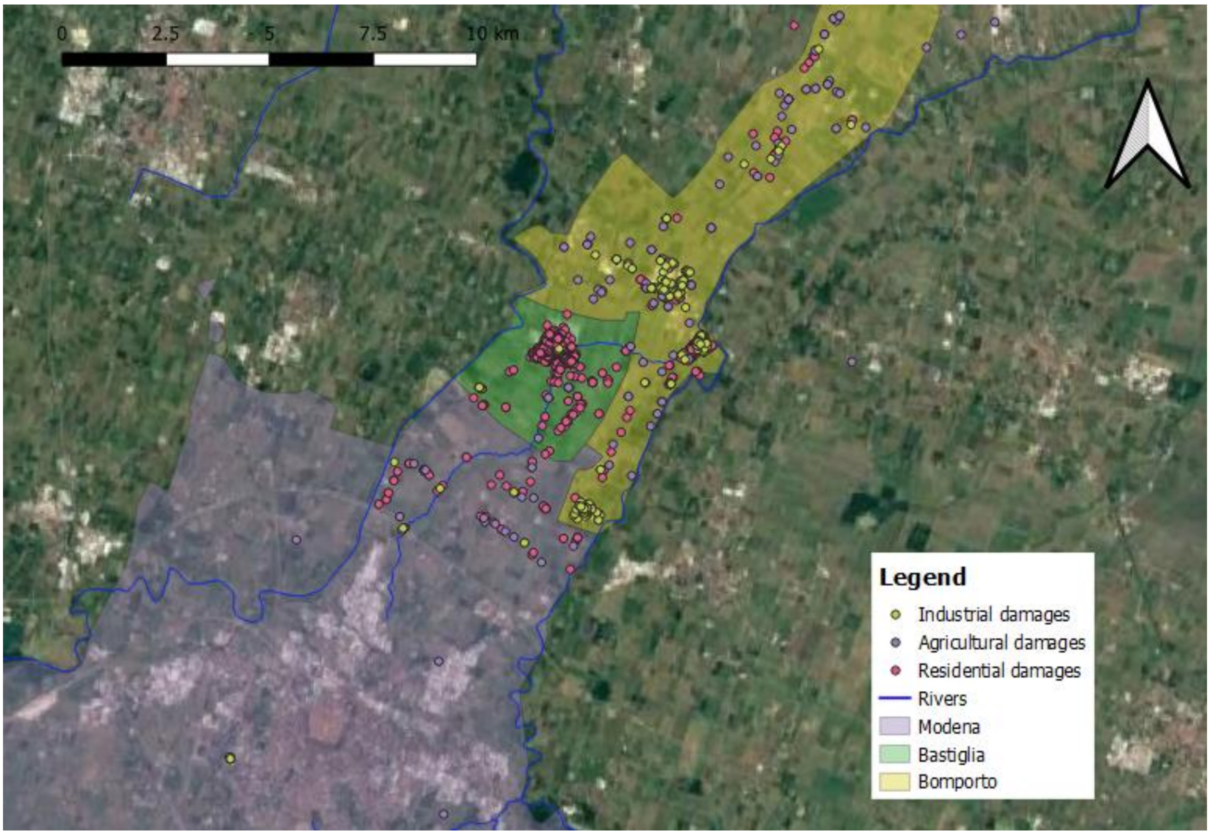


Figure 6: Visualization of multi-sectoral damage claims against the 2014 Secchia flooding event

Address data, though it must remain anonymized otherwise, is available for analysis within the context of this thesis and the LODE project itself. This allows for a microscale and precise geographical mapping by locating the addresses as points on the map that can be spatially matched to the other variables collected. For residential properties, these damage claims are recorded separately under four categories, individually detailing claims on building structures (both shared and private), household contents, and registered vehicles.

In fact, in terms of what damages could be claimed, building damage refers not only to structural parts such as roofs, foundations and supporting structures, but also non-structural parts including flooring, plastering and painting, as well as installations such as those for electricity, heating, and water.

The data collected for agriculture and industry is similarly formatted, categorized by the two sectors, with information available again regarding the address, building type, and area. In this case, damage claims are more geared towards business losses, capturing damages to structures and land used for agricultural purposes, as well as machinery and stock.

A disadvantage of the damage data available for all the three sectors, is presented by the nature of the data collection itself, as not all premises submitted damage claims. Specifically, depending on each individual claimant, this may be because there were negligible damages, or because of a deliberate choice not to submit a claim. This means that there will be some level of data gap across the investigated hazard-prone area.

2.3.4 Data for future scenarios analysis

In order to examine any '*what-if*' future scenarios through the BN model, it is necessary to have projected datasets that are consistent with the kind of data (variables) originally used to inform the training of the model under the reference scenario. For this reason, data has been collected at varying time ranges from the JRC LUISA platform[5], which models various dynamic functions pertaining to land, land use, and population across Europe, as well as from the JRC LISFLOOD[6] hydrological flood-simulation model.

Among their projections, flood depth simulations covering flood prone areas, such as the case study area, are available at return periods of 10, 20, 50, 100, 200, and 500 years (Dottori et al., 2016), reflecting the increasing intensity of potential future river flooding events.

Also available are projections of the land use cover for the years 2020, 2030, 2040, and 2050, which can be used in direct comparison to the 2010 data used for training and validation of the BN model (Aurambout & Lavalle, 2016; Lavalle, 2014). These indicators can then be used to further elaborate the BN-based scenario analysis for prediction of future changes in the probability of flood-related damages, in line with EU recognized projections.

[5] https://data.jrc.ec.europa.eu/collection/luisa

6 https://ec.europa.eu/jrc/en/publication/eur-scientific-and-technical-research-reports/lisflood-distributed-water-balance-and-flood-simulation-model-revised-user-manual-2013#:~:text=%C2%A9EU,JRC)%20of%20the%20European%20Commission.

2.4 Data processing methods

One advantage of BN models lies in their ability to incorporate a variety of variables, with different formats, units, and temporal or spatial resolutions. However, following the collection of the raw data for the training and validation of the BN model, it was necessary to pre-process the abovementioned datasets, transforming them to a consistent spatial reference system (i.e. WGS84/UTM zone 32N) for extracting the table of data then feeding the BN model.

Hazard data (i.e. flood depth, inundation time, and flood velocity) was available in raster format at 5m spatial resolution, showing the flooding extent and its severity. In order to best represent the maximum potential impact at each location, the maximum levels were captured for each of the variables of concern, these being flood depth, velocity, and duration.

Damage data points were automatically geolocated using the "Geocode" tool in "MMQGIS" plugin[7] in QGIS, by matching their address to the address as listed in Open Street Map[8] (OSM; www.openstreetmap.org). The identification from OSM was not able to geolocate all the points due to missing information in the geodatabase, and thus the missing points were manually localized using Google Earth[9], by matching the address of the reported damage and to the equivalent in Google Earth. The final outputs are three vector layers, one each for agricultural, residential, and industrial damages.

The rest of the vector and raster-based variables for exposure and vulnerability were then converted into the same spatial reference system, so that data for each variable considered could be correctly extracted at the exact location of the respective damaged property. In vector forms, shapefiles containing sub-municipal information were geocoded with the data for each variable being joint to the respective polygons.

Raster layers were taken at the highest spatial resolutions possible for data such as land use cover and population density. Moreover, the land use/cover layer was used as a reference point for the BN-based scenario analysis, with the collection of the equivalent projected data as far as 2050, allowing for a fair comparison for future scenarios. Similarly, projected flood maps under different return periods were geolocated as raster layers for future scenario

[7] https://github.com/michaelminn/mmqgis/
[8] https://www.openstreetmap.org/
[9] https://www.google.com/earth/

analysis, from the JRC LISFLOOD platform, with the increase in depth between the 10-year and 200-year return periods being calculated for analysis.

Finally, to provide the data in the suitable for to train the BN model, the values of these layers were extracted at the coordinates of the location of each damage report, using the "extract" function in the "raster" package[10] in R. The final output of this process is a matrix of 1738*19, where the number of rows indicate the number of reported damage points and the numbers of columns represent the numbers of considered variables. The list of variables and their notations are reported in Figure 8 (Section 3.1).

[10] https://www.rdocumentation.org/packages/raster/versions/3.3-13

3. Methodology for multi-sectoral flooding damage assessment and scenario analysis

The design of a BN model follows a stepwise approach as initially detailed in section 1.1. Here, these methods are described more specifically for application to this Thesis and the case study area, beginning with the definition of the BN conceptual model. It then follows through the necessary operative steps for the subsequent model parametrization, calibration, validation, and sensitivity analysis, as well as detailing the scenarios developed for 'what-if' analysis.

3.1 Risk-based conceptual framework

In the design of a BN model, the characterization of a conceptual framework is an important step in the formalization of the issue being studied (Furlan et al., 2020). As such, a conceptual framework is built, systematically identifying pathways of interaction between environmental, physical and socio-economic damages on the exposed sectors, and the drivers of those damages. Specifically, a properly constructed conceptual framework should give a comprehensive schematic representation of the cause and effect relationships within the system, encompassing all necessary sources of data, as well as their relative interactions (Defra, 2011). With the identification of these cause-effect relationships between the system variables, a 'roadmap' is laid for the training of the BN model from the observed data (Sperotto et al., 2017).

For this project, the IPCC's risk framework (IPCC, 2018), was selected as the basis for the construction of the conceptual framework, as depicted in Figure7. This framework has been applied for a wide range of risk assessments in the context of climate change adaptation and disaster risk management and reduction (Das et al., 2020; Sharma & Ravindranath, 2019; Tangney, 2019).

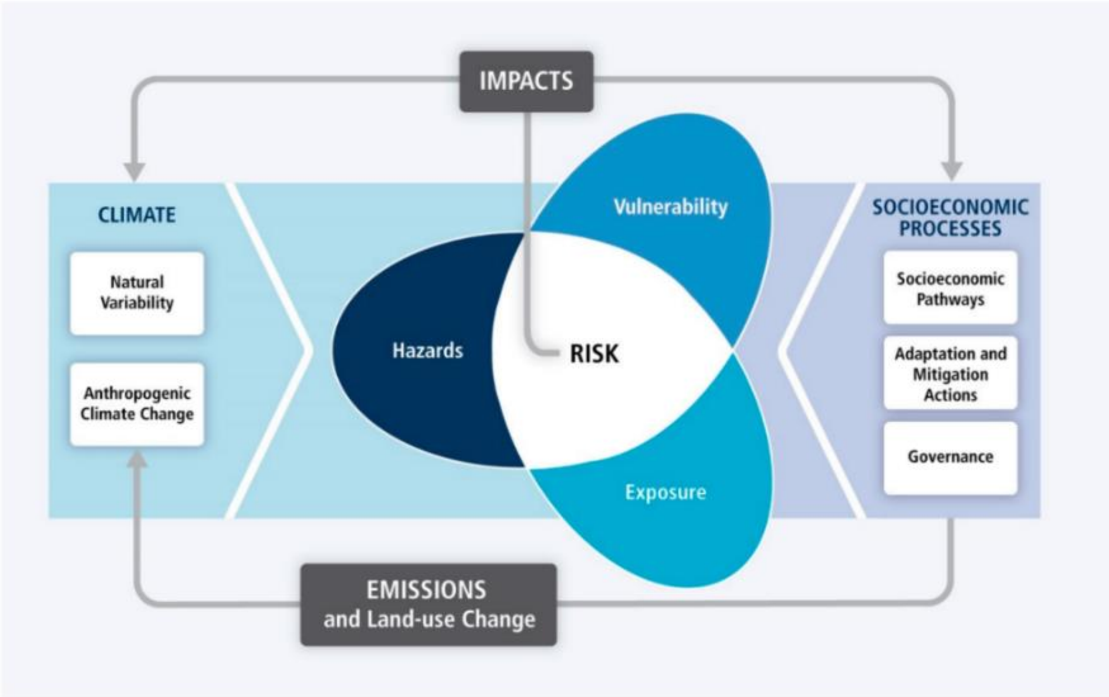


Figure 7: Framework for Disaster Risk (Field et al., 2014)

In this setting, it follows that when combined, the elements that contribute to hazard, exposure and vulnerability directly affect the disaster risk, and hence its impacts in terms of multi-sectoral damages for the case study of concern. In turn, changes in the hazard patterns and potentially exposed targets can be driven through anthropogenic and natural factors (Connelly et al., 2018).

The data collected for this project were chosen to best represent the factors that may contribute to the risk of flooding damages, under the separate categories of hazard, vulnerability and exposure.

An adapted version of the IPCC's framework that forms the basis of the conceptual model for training of the BN, designed specifically to consider the scope of the work, i.e. a multi-sectoral assessment of fluvial flooding damages. The data as identified and collected in section 2 will be used to evaluate multi-sectoral damages through the exposure of buildings and structures to river flood-related hazard across the residential, agricultural, and industrial sectors, together with the varying influence of different physical and socioeconomic vulnerability variables (e.g. the age and conservation status of the structures). In turn impacts against river-flood events is reported in terms of damage probabilities under the three investigated sectors.

This framework is translated into a box-and-arrow diagram (Figure 8) ready for BN structure definition, capturing all the variables collected for input into the BN model. Connections are made between the input variables and the endpoint values they are expected to have an impact upon (i.e. damages for each sector).

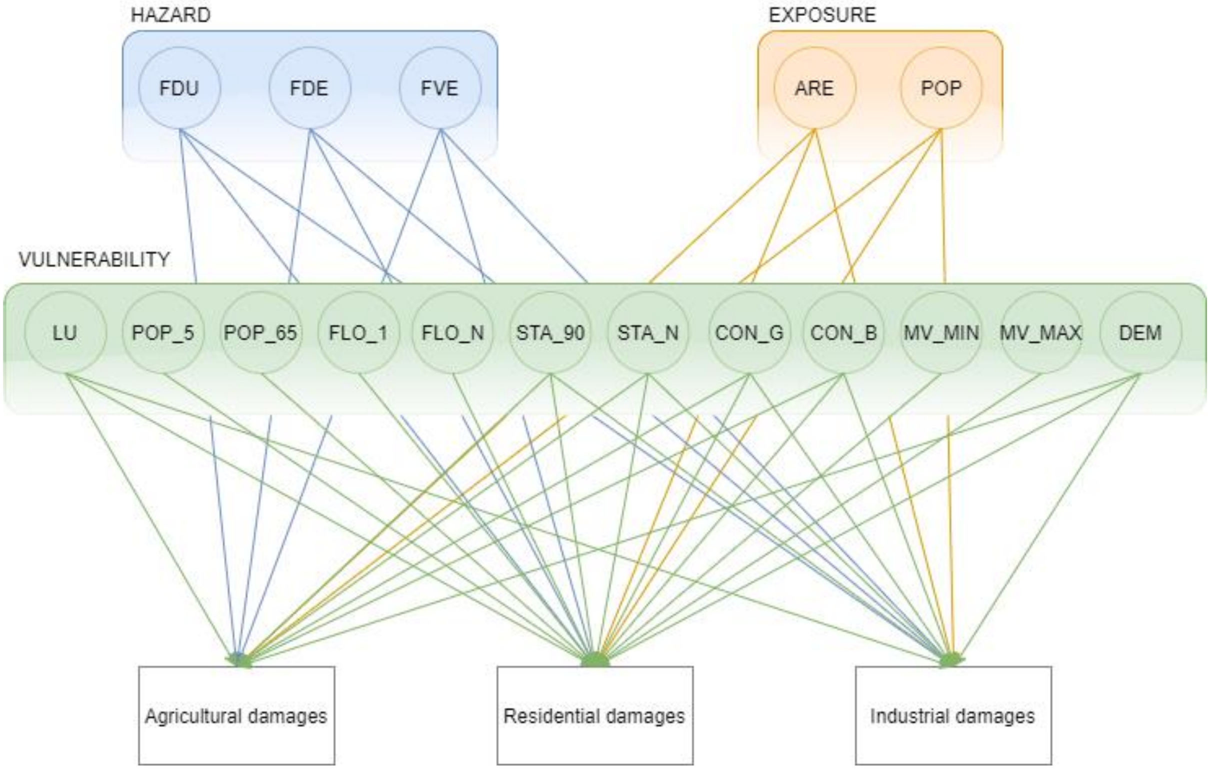


*Figure 8: Risk-based BN Conceptual Framework. List of acronyms used for **hazard variables**: FDU: Flood duration; FDE: Maximum flood depth; FVE: Maximum flood velocity; **Exposure variables**: ARE: Area of reported damage; POP: Population density; **Vulnerability variables**: LU: Land Use Cover; POP_5: Population under 5; POP_65: Population over 65; FLO_1: Number of houses with 1 storey; FLO_N: Number of houses with greater than 1 storey; STA_90: Number of houses constructed before 1990; STA_N: Number of houses constructed 1990 and later; CON_G: Number of houses with good conservation status; CON_B: Number of houses with bad conservation status; MV_MIN: Minimum residential market value; MV_MAX: Maximum residential market value; DEM: Digital Elevation Model*

3.2 BN development

3.2.1 Model design and configuration testing

The practical implementation of the BN requires a multi-stage approach, as initially described under the Review of the state-of-the-art BN application for damage assessment in section 1. This begins with the design of the BN model and the subsequent parametrization of necessary variables.

As seen in the risk-based conceptual framework, the chosen variables have been selected, through expert judgment and literature review (Merz et al., 2010; Dominik Paprotny, Kreibich, Morales-Nápoles, Wagenaar, et al., 2020), according to their potential influence on the overall flooding risk and as such their likely contribution to the multi-sectoral damages. This causal relationship has been depicted in Figure 8 with a box-and-arrow diagram that represents the relevant influential relationships; this graphical depiction can be used to define the BN model, incorporating all of the identified variables. The boxes of the diagram are equivalent to the nodes that represent the system variables, with unidirectional arrows between the boxes depicting the arcs that determine the causal relationships between variables in the model, eventually terminating at the assessment endpoints (Sperotto et al., 2019).

An alternative approach to understanding the optimal model performance is by analyzing various configurations of the model, as defined through expert judgement and pertinent literature. By setting these different configurations and observing the respective model outputs (also in terms of model prediction performance), it is possible to identify which models perform best in comparison to one another (Poelhekke et al., 2016).

Within this study a two-tiered approach was developed to define and test different BN model configurations. Specifically, the first stage involves starting with a simple model that initially only integrates the core variables for training, and then performing a stepwise integration of further variables one by one; in this way, it is possible to track improvements in performance of the model over time. This would be particularly useful in the case of a model constrained by limitations in the input data, where the introduction of too many variables may provide too many boundary conditions and restrict the performance of the model.

The second stage looks at reconfiguring the model (to improve its performance) by identifying, through expert judgment, new possible connections between the explanatory (parent) nodes and related child nodes, thus incorporating new layers of hierarchy within the BN structure.

Following the process of BN multiple configurations testing, for the training of the model, all the variables must each be assigned a state in the form of either a value or a condition. These variable states can be defined in three different ways, namely i) into qualitative categories such as high, moderate or low quality, ii) as true or false states (i.e. Boolean functions), or iii) quantitatively, as a range or in discrete intervals (De Santa Olalla et al., 2005).

After this input definition is complete, two computations are necessary as part of the parametrization process (Sperotto et al., 2019). Firstly, this involves the calculation of the associated prior probability of each state of the node, i.e. the relative likelihood of each possible state without any other knowledge of the variable relationships, based on the distribution of the input data. Secondly, the conditional probabilities of any child nodes must be calculated as dependent on all possible combinations of the associated parent nodes (Sperotto et al., 2017). Finally, a Conditional Probability Table (CPT) is developed to display the relative strengths of the causal relationship between all connected variables.

3.2.2 Model calibration and validation

Where the predicted probabilities of each variable and the strength of their relative relationship has been calculated, the next phase of the process is to thoroughly evaluate the output of the BN model, in order to fairly assess both the accuracy and the reliability of the results. This is important to understand the potential for use of the designed BN as a predictive model for new observations under scenario analysis (Furlan et al., 2020).

Validation of the various BN model configurations can take the form of a data-based evaluation, where errors in the model output are identified through the use of a statistical test, or in relation to a set of independent observational data. Alternatively, expert judgment can be utilized to form a qualitative evaluation of the results, or similarly through comparison of the model outputs to those of similar models found in the literature, however this is generally performed when there is insufficient data for statistical testing (Kragt, 2009).

For the estimation of the model predictive error, possible techniques range from Re-substitution and Hold-out methods, to the more complicated Bootstrap and Bolstered options (Furlan et al., 2020). One such data-based method for evaluating the accuracy of the model is k-fold cross validation (k-cv), where the data is split into k sets (or folds) of equal size and the model is trained on all but one of these folds, with the errors then calculated for the final set of observed data. This process is then repeated with all possible combinations of k-1 folds, and the average error of these different combinations is calculated to reflect the overall accuracy of the model (Yadav & Shukla, 2016).

3.2.3 Scenario analysis

The next phase of analysis of the BN approach concerns the scenario analysis, in which various potential scenarios are studied in order to predict their respective impacts. The conditions of these scenarios are simulated by 'setting' different evidence for one or more nodes within the BN model, and then propagating that information through the system, thus inferring the behaviour of the variables in order to observe the changes in posterior probability resulting from each scenario (Sperotto et al., 2017).

In order to infer this information, the direction of propagation must first be determined. A downward propagation of probability is known as prognostic inference, where the values of one or more input (or parent) nodes are set, and the impact to the posterior probabilities of the respective child nodes is observed, usually as far as the endpoints of the BN. Opposingly, the probability of a child node can be set to a fixed value as a form of diagnostic inference, where the change in probability is propagated upwards through the model towards the parent nodes (McNaught & Zagorecki, 2009).

Following the identified literature knowledge gaps (Section 1.2) the definition of potential future '*what-if'* scenarios was determined to be a key component in the context of this Thesis. To achieve this, variables have been identified to reflect possible future changes in flood risk. The specific simulation of these changes was defined by setting the values of the selected variables under the following scenarios:

Scenario 1, (SC_LU): Understanding the change in damages under changing land use patterns, as determined by comparison of the 2010 land cover training dataset from the JRC LUISA model, against the expected land cover changes up to 2050 from the same model projection. This involves classification of the land into three main categories, as shown in Figure 5. Specifically, the aggregated changes mostly concern a loss of agricultural land, replaced by urban areas. There are changes in some areas to and from industrial zoning within the case study area, however the overall share of industrial land remains roughly equal.

Table 5: Distribution of land use cover types within the case study area

Land use	% coverage in 2010	% coverage in 2050
Urban fabric	1	4
Industry/commercial/services	21	21
Agriculture	78	75

Scenario 2, (SC_FDE): Understanding the change in damages at different flood-related return periods. The flood depth data for the BN training was obtained from a high-resolution floods model, based on an explicit shock–capturing finite volume method for the solution of the 2D shallow-water equations, evaluated as a 5-year return period event (Vacondio et al., 2016). Then, to model the scenario SC_FDE, we compared the distribution of flood depth of a flood event under a 200-year return period, versus a 10-year return period flood event, comparable to that which occurred in the Secchia river in January 2014 (Shustikova et al., 2020).
This comparative analysis was used to find the changes in each of the flood-depth related classes, the relative changes then being used to set evidence for the scenario related to the flood depth. Figure 6 shows the comparative frequency of flood depths between the two return periods as parametrized into three classes for use in the BN model.

Table 6: Relative frequency of flooding depth within the case study area

Depth (m)	% frequency, 10-year return period	% frequency, 200-year return period
[0,0.87]	82	76
(0.87,1.21]	14	12
(1.21,3]	4	12

3.2.4 Sensitivity analysis

A further phase of analysis identified as mostly absent within the context of the recent literature was a detailed sensitivity analysis. This evaluation should provide information on the sensitivity of the assessment endpoints of the BN model (i.e. damages to the residential, agricultural, and industrial sectors), in relation to changes in their various explanatory nodes. This analysis can be completed through two phases (Kragt, 2009; Pollino et al., 2007). Firstly, the relative impacts of each parameter on the output are determined, thus allowing for the identification of the most influential set of variables. Further to this, the stepwise modification of individual input parameters can then be used to observe changes in the damage assessment endpoint probabilities. As such, it would also be possible to interpret how the various input nodes impact the model outcome, and understand their relative importance in determining the highest class of flood damages (Furlan et al., 2020).

SECTION C: APPLICATION TO THE CASE STUDY AREA

4. Application of the BN approach for multi-sectoral damage assessment.

4.1 Model design and configurations testing

As described in Section 3.1, an initial risk-based conceptual framework was established that attempted to highlight the cause-effect relationships between the various components of disaster risk (i.e. hazard, exposure, and vulnerability), and the resulting disaster damages.

According to the data available and the aims of the LODE project itself, this framework was then translated into an expert-based BN conceptual model, as seen in figure 8, that provided the basis of the practical BN model. This was composed of a multifaceted set of explanatory (parent) nodes representing the hazard, exposure, and vulnerability characteristics of the case study as collected, each connected to the child nodes, i.e. the multi-sectoral damages.

In order to introduce the data to the model, variables were assigned states, according to the characteristics of their values (i.e. continuous, discrete, or Boolean). Specifically, all continuous variables, e.g. structure area or flood depth, were classified using the discretization function in R into three intervals with similar frequencies of data. Other qualitative variables were categorized by the intrinsic characteristics of the data, such as land use cover into urban, agriculture, and infrastructure classes, and in these cases states were assigned to each category instead. The endpoint nodes were also classified into three separate continuous classes, representing the magnitude of monetary damages to the agricultural, industrial, and residential sectors.

The first generation of the model was built by following the risk-based conceptual framework (with the same structure and variables as Figure 8), in the R environment using the package "bnlearn" version 4.4.1 (Scutari, 2010). Unfortunately, the performance of this model, as shown in Figure 9, was hampered by limitations in the available data, together with the constraints set by the full set of input variables. The outcome represents the trade-off between the number of training datapoints and the number of variables and classes and, consequently, the exponential number of their conditional probabilities. Due to this poor data condition, especially for the agriculture sector, the model could not define the conditional distribution of the final assessment endpoints, leading the assignment of a normal distribution as a default mode of the model.

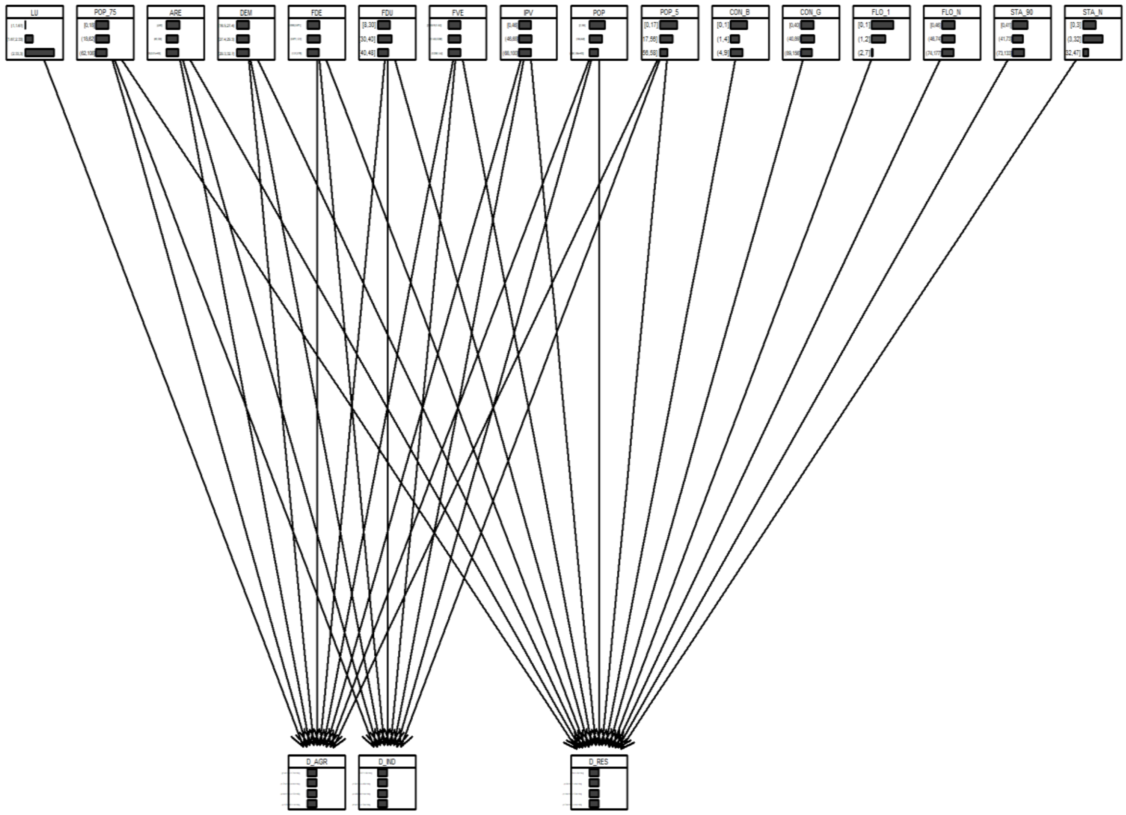


Figure 9: First generation conceptual risk-based BN model for multi-sectoral flooding damages assessment, with associated variable marginal distributions

As a result, a new two-tiered approach to the model configuration was conceptualized as discussed in section 3.2.2, in order to exploit the available training data while also incorporating a range of the collected variables to inform the model as best as possible. Beginning with a limited number of explanatory nodes, more variables were integrated step-by-step into the model until the stage at which the model performance was deemed to be unreliable. Specifically, the first iteration of this new model under the first tier of model reconfiguration utilized only the hazard variables (i.e. maximum flood depth (FDE), velocity (FVE), and duration (FDU)) as explanatory metrics of multi-sectoral damages (hereafter CONF_1A). Compared to the first-generation risk-based conceptual model, the impacts of the input variables were much more evident, showing a reliable distribution of conditional probabilities for each sector-based endpoint for flooding damages. This was most successful for the residential sector (D_RES), followed by the industrial (D_IND) and then the agricultural sectors (D_AGR), reflecting the relative volume of available observational data.

One by one, more explanatory nodes were added to further iterations of the model, with the aim of improving the model performance, i.e. reducing the classification errors of the final assessment endpoints, and terminating the process when no significant improvement was observed in the model's performance (as discussed in the next section 4.2). Firstly land use cover (LU) was chosen (hereafter CONF_1B), which also allowed for the planned future scenario analysis. To introduce an indicator of potential exposure, the area of reported damages (ARE) was the next selected variable (CONF_1C).

Several further iterations of the model were also tested, introducing other explanatory nodes e.g. the population density (CONF_1D). However, with the introduction of each new variable, the conditional probabilities of the utility nodes flattened, tending towards the output observed from the risk-based conceptual model (These configurations were also tested in the validation stage, with the results reported in Annex I). This evolution was particularly pronounced for the agricultural sector, again due to the data-poor condition. The decision was made to limit the model to five explanatory nodes (CONF 1C) so as to sufficiently balance the assessment of different variables with the performance of the model. The results of the continuous iterations of Configuration 1 of the BN model are shown in Figure 10.

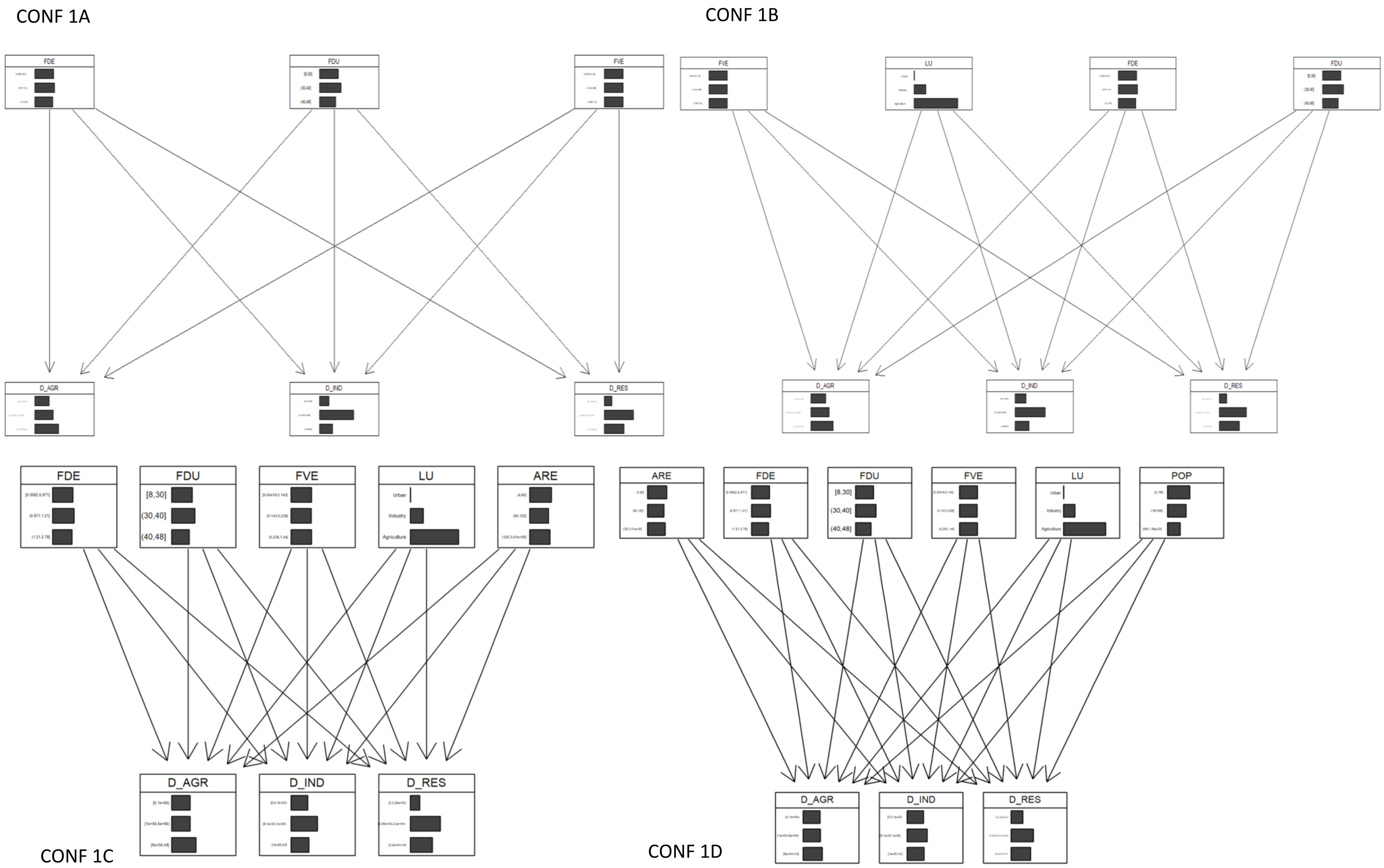


Figure 10: Outputs for the definition of the BN model from CONF 1A to CONF 1D. List of acronyms used for ***Explanatory nodes****: FDU: Flood duration; FDE: Maximum flood depth; FVE: Maximum flood velocity; ARE: Area of reported damage; POP: Population density; LU: Land Use Cover;* ***Damage nodes:*** *D_AGR: Agricultural; D_IND: Industrial; D_RES: Residential*

Then, the second tier of model configuration involved investigation of the structure of the model, by rearranging the input nodes to form a new layer of hierarchy within the model (CONF 2). In this instance, it was decided to treat the variable concerning the area of reported damages as not only an exposure-related variable, but specifically to define it as a receptor of the flooding hazard. As such, instead of being connected directly to the endpoint nodes, the hazard variables (FDE, FDU, FVE) are first connected to the area node, which then leads to the multi-sectoral damages (as also developed in D Paprotny, Kreibich, Morales-Nápoles, Terefenko, et al., 2020; Sayers et al., 2002). Figure 11 shows the new **CONF 2** as selected for analysis in the validation stage (Section 4.2). This parameter training process was implemented under the *bn.fit* function on the “bnlearn” package in R (Scutari, 2010).

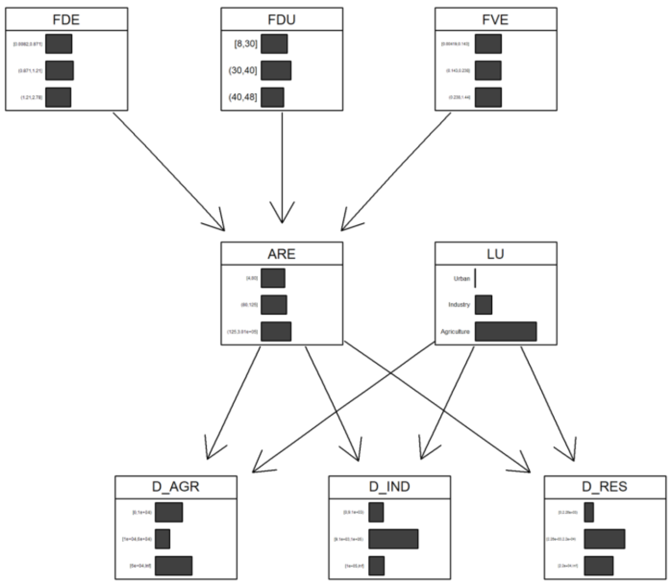


Figure 11: CONF 2 - BN model for multi-sectoral flooding damages assessment with associated variable marginal distributions

4.2 Model calibration and validation

Once the different configurations of the BN model have been set, with the appropriate parametrization of the chosen variables, the models could then be validated in order to give an estimation of their prediction error and then select the model the highest performance in the multi-sectoral damages estimation. As explained in Section 3.2.2, data-based evaluation was performed in order to determine the probability of observational data being misclassified in the three damage-related assessment endpoints.

K-fold validation, was applied for this analysis, and carried through the *bn.cv* function in R with the “k-fold” method (Scutari, 2010). Should one fold of the training data not contain enough datapoints for each sector, the prediction of damages for the missing sector would be impossible. As such, due to the limited amount of observational data available for the agricultural sector for training of the model, the number of folds used for the validation process was limited to five, in order to minimize the probability of this adverse outcome, with the results of this analysis for **CONF 1C** illustrated in Figure 12.

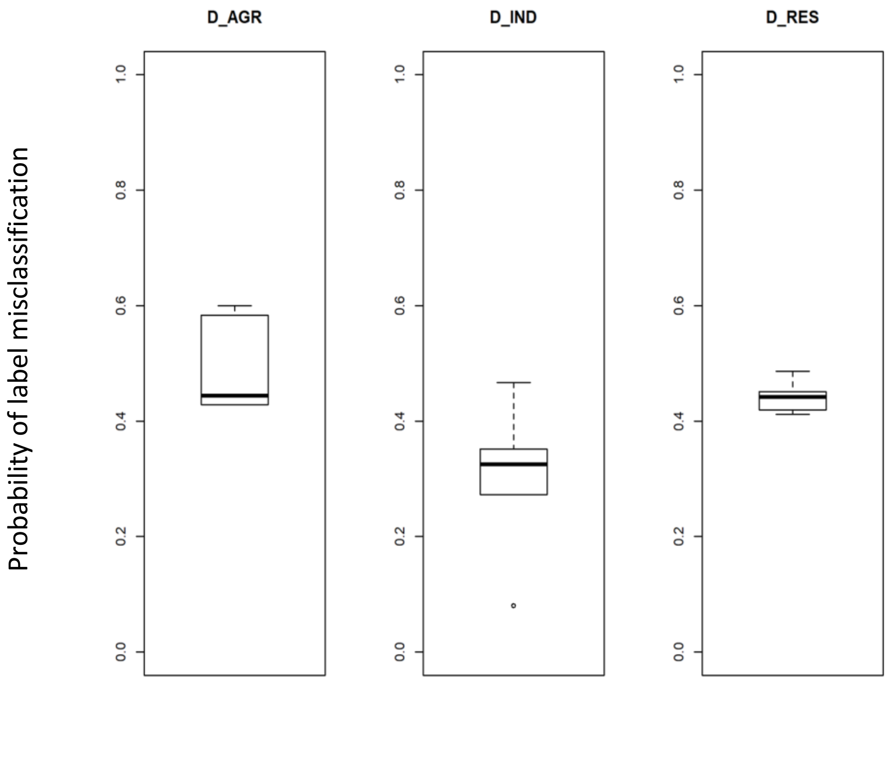


Figure 12: Boxplots representing the probability of misclassification of damages for the i) agricultural (D_AGR), ii) industrial (D_IND), and iii) residential (D_RES) sectors from CONF 1C of the BN model

The median classification errors were calculated as 45% for the agricultural sector, 34% for the industrial sector, and 44 % for the residential sector, showing a correctly classified output in the majority of cases. With increasing data points, the range of errors among folds narrowed, and with the most data available, error estimation showed the least uncertainty for the residential sector with less than 10% of variance.

The analysis was then repeated for **CONF 2** of the BN model in order to compare the respective predictive capability, with the results shown in Figure 13 below.

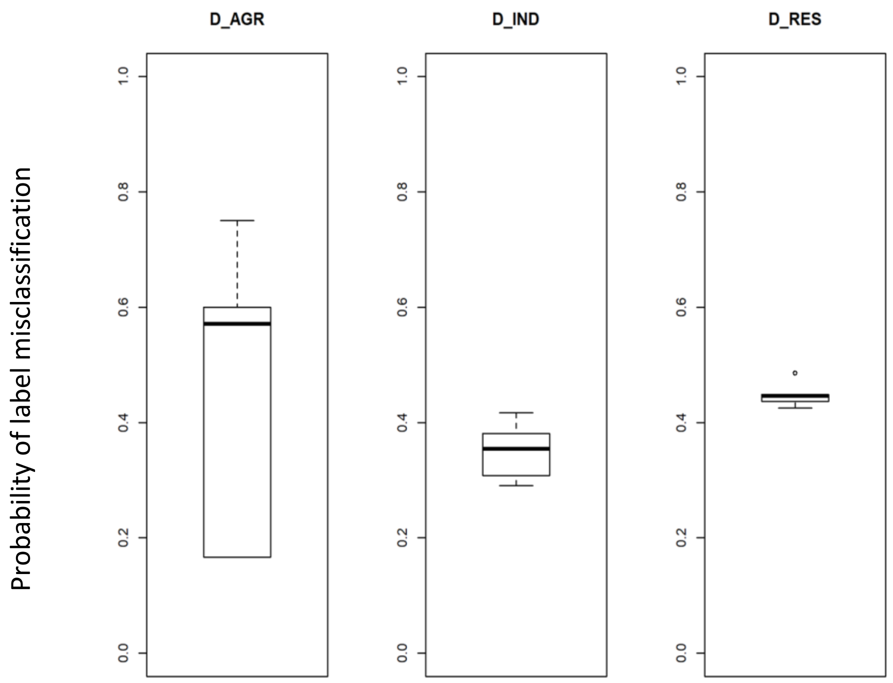


Figure 13: Boxplots representing the probability of misclassification of damages for the i) agricultural (D_AGR), ii) industrial (D_IND), and iii) residential (D_RES) sectors from CONF 2 of the BN model

As can be seen, for this new configuration the average performance of the model showed no significant improvement for any of the sectors in comparison to **CONF 1C**, and moreover the variance in the model output noticeably increased for the agricultural sector. As such, the decision was made to proceed with the **CONF 1C** as the final version of the BN model for application.

4.3 Scenario analysis

After validation of the model, it was then possible to apply it for inferential purposes, including the analysis of two future scenarios as identified in section 3.2.3. For each scenario, evidences were set based on the projected dataset for the variables of concern (i.e. land use/cover, flood depth), and the changes were propagated downward, with the output recorded for comparison with the original conditional probabilities

For **Scenario 1** (SC_LU), the model was trained with the land use/cover data for the 2010 timeframe, and then compared to the 2050 scenario exploiting the JRC LUISA dataset. For **Scenario 2** (SC_FDE), the comparative changes in flood depth classes between the JRC LISFLOOD model 10-year and 200-year return period flooding event projections were used to build the future scenario. The impacts of these evidences were then propagated through the BN to each damage-related assessment endpoint as a form of prognostic inference.

The resulting model outputs for the agricultural, industrial and residential sectors are shown in Figure 14 with sub-index (i), (ii), and (iii), respectively. Particularly, the coloured boxes represent the frequency of labels classed under the three classifications of monetary damage under the different scenarios, with the value of damages increasing from blue to red.

i) ii) iii)

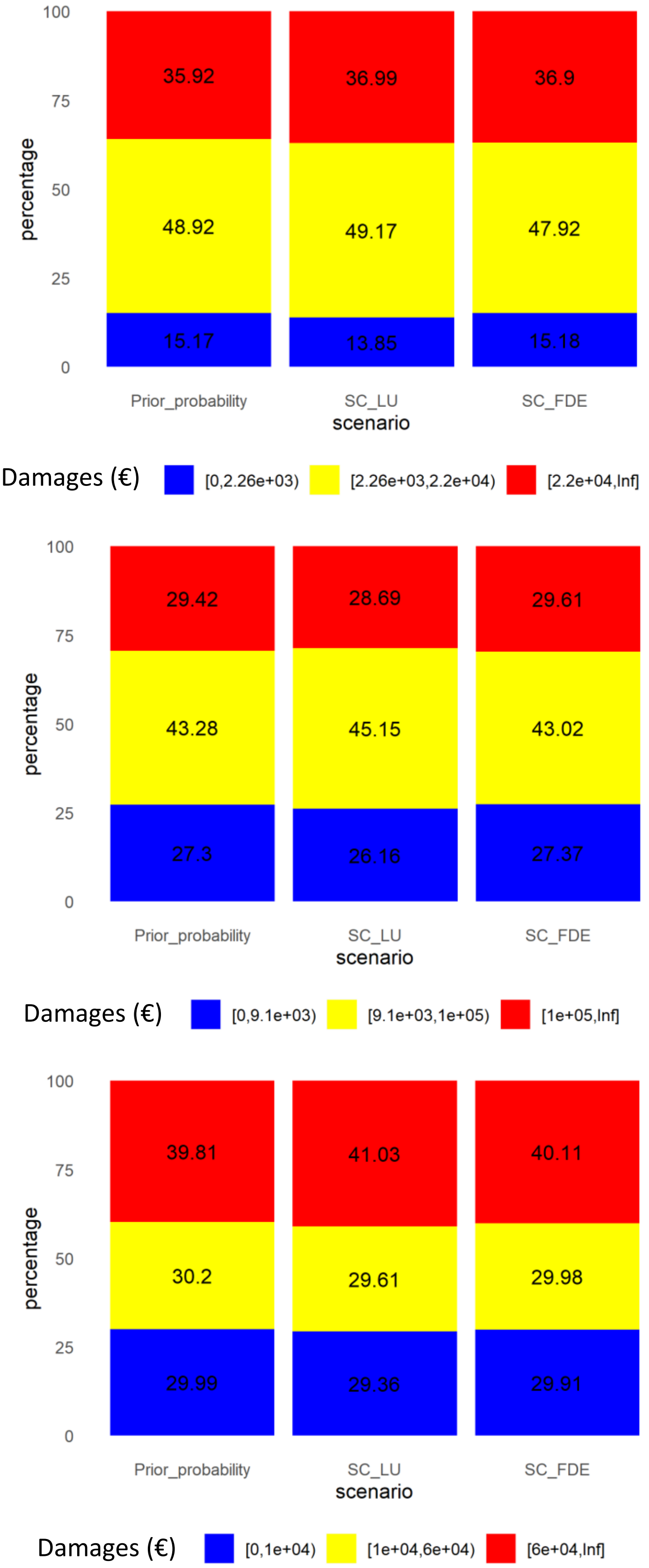


Figure 14: Outputs of the BN model, comparing the prior probability of the BN model against the two simulated scenarios (SC_LU) and (SC_FDE), for the evaluation of potential damages in the i) agricultural, ii) industrial, and iii) residential sectors

Scenario 1: Changing land use (SC_LU)

In this scenario, in comparison to the prior probability, there is a decrease in the probability of damages within the lowest class (indicated by the blue segments) for all sectors, with consequently a noticeable increase in the highest damage class for the agricultural and residential sectors. For these two sectors in particular, these results indicate that there is an expected increase in damages to be seen over the next few decades under the projected changes in the land use. The industrial sector does not conform to these patterns in the same way, mainly due to the very small change in total industrial area, as evidenced in Table 5 (section 3.2.3).

However, while the results do show signs of changes in future damages, they are of limited magnitude; this is a result of the magnitude of change in the land cover over the 40 year period, where approximately 80% of the case study area does not change land use classification, of which the majority remains agricultural. In fact, as seen in Table 5, the aggregated changes show a slight increase in residential land over agricultural. Should there be a period of more intense development, it could be expected that the effect on flooding damages would be much more severe.

Scenario 2: Changing flood depth (SC_FDE)

Similarly to the land use case, this scenario shows a limited but consistent increase in flooding damages across the multiple sectors, showing the expected impacts of more severe future river flooding events. While this effect is most pronounced for the agricultural sector, the changes are smaller for the residential and industrial damages.

These limited changes are again likely a result of the input training dataset, where the flood depth does not increase too significantly between the 10-year and 200-year return periods. Specifically, there is an approximately 8% increase in the probability of the highest flood depth (specifically over 1.21m), as evidenced in Table 6 (Section 3.2.3). Further, while the flood depth at different return periods has been projected, equivalent datasets for the other components of flooding hazard (i.e. duration and velocity) are not made available, limiting the scope of the analysis to solely considering the flooding depth. As such, while flooding depth is the key variable used within damage modelling and assessment, it is more difficult to capture the changes in hazard characteristics.

4.4 Sensitivity analysis

As the last stage of the process, a sensitivity analysis was performed in order to provide information on the sensitivity of the assessment endpoints of the BN model (i.e. damages to the residential, agricultural, and industrial sectors), in relation to changes in their various explanatory nodes, based on the methodological approach detailed in Section 3.2.4.

This was performed individually for each of the explanatory nodes in the model, by setting a 100% probability of the highest state (e.g. highest flood depth, largest area of reported damages), while keeping all other nodes constant. In doing so, it is possible to see the relative impact of each variable in context with the other explanatory nodes.

The results of this analysis are shown in the rose charts reported in Figure 15 below, with the relative probability of each damage class given for the simulation that changed each of the five explanatory nodes (FDU, FDE, FVE, LU, ARE), with the prior probability (PrP) shown for comparison. As with the scenario analysis, the red, yellow and blue sections represent the highest, moderate, and lowest class of damages respectively.

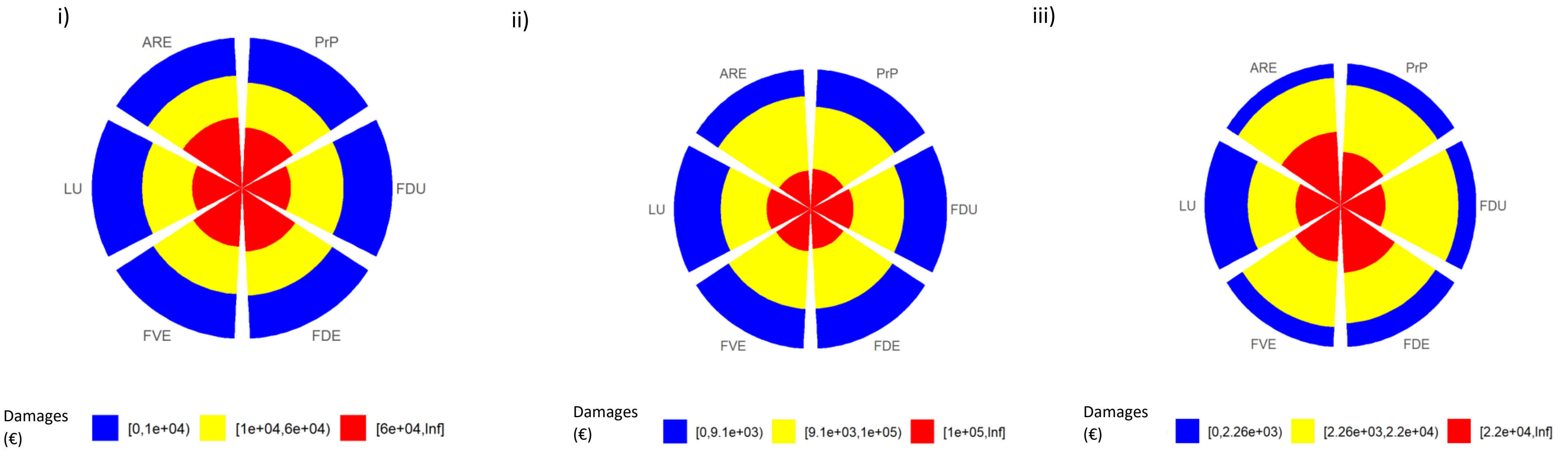


Figure 15: Sensitivity analysis for the explanatory nodes of the constructed BN model for the i) agricultural, ii) industrial, and iii) residential sectors.

The results indicate that the importance of each variable in terms of contributing to the potential damages varies by sector. Specifically, changes in the posterior probability for damages compared to the prior probability indicate that for the agricultural and residential sectors, the damages are particularly sensitive to changes in the variables concerning the area of reported damages (ARE) and flood depth (FDE), in line with the results found in the corresponding key literature (Kreibich et al., 2009; Merz et al., 2010). The probability of an output in the highest damage classification increases significantly for these sectors with increasing area or flood depth, although similar results are not seen for the industrial sector. Instead these damages are more susceptible to land use changes, as well as flood duration and velocity, which may go some way to explaining the unexpected response of the industrial sector to the identified future scenarios.

CONCLUSIONS

This thesis presented a GIS-based Bayesian Network (BN) approach capable of capturing and modelling multi-sectoral flooding damages, by exploiting damage data collected from the 2014 Secchia river flooding event. With the aim of providing support for Disaster Risk Management and Reduction against extreme river flooding events, the developed BN-based methodology represents a novel approach for better understanding flooding damages for the agricultural, residential, and industrial sectors, and the prediction of future damages under possible changes in hazard and exposure patterns.

Building on the state-of-the-art research in the field of Machine Learning, the work aimed to expand upon the current literature by addressing several identified knowledge gaps. Specifically, the methodology as presented offers a more complete picture on multi-sectoral damages by including not only residential damage prediction as an assessment endpoint, but also the industrial and agricultural sectors. It also provides an analysis of two '*what-if*' scenarios for the examination of potential future damages. Further, various approaches for the design and configuration of the BN model are deeply explained, alongside a sensitivity analysis, providing greater insight into the optimal design of BN models for multi-sectoral damage assessment. Specifically, the final model, as constructed, showed good capability of damage prediction for all three sectors studied within the case study area and under two different scenarios (i.e. changing patterns in land use and flood depth).

The analysis of two future scenarios envisioning on one side land use/cover change in the case study area, and on the other greater flood depths resulting from more severe river flood events, showed good promise in the capacity of the BN model to better understand possible future damages.

Moreover, the stepwise model configuration also provided insight on how to optimise these results through both expert- and data-driven procedures, while the sensitivity analysis highlighted the relative value of the integration in the BN model of each chosen explanatory variable.

These results, in combination with the inherent flexibility of the proposed BN model, allows for the potential integration of diverse heterogeneous datasets. Thus, it is possible to assimilate as much information and expert knowledge as is available during the training of the model and testing of other '*what-if*' scenarios. As such, the model also has the ability to be further improved through targeted data collection, where increasing the quality and quantity of the data will provide additional insight into the results, also improving the BN model reliability.

However, while the obtained results showed high capability for damage prediction, they also highlighted some potential limitations in the methodological approach and its application, particularly where the necessary data is limited. In fact, limitations in terms of the availability of a large amount of data for the case study caused constraints on the capability of the model performance. Resultingly, the integration of many of the variables collected for training the model was not possible, and as such it was more difficult to gain insight into the full extent of the various contributing factors of flooding damages. These results stress the need for the collection of sufficient damage data post-flooding events, in order to best enable a successful model training and then scenarios analysis, particularly for the agricultural and industrial sectors where model prediction accuracy showed higher uncertainty.

As a cascading effect, the scenario analysis identified the potential impacts of expected future changes in flood depth and land use cover in the case study area, however the magnitude of these impacts were lower than might have been expected, as a result of limitations linked to the input training datasets. As highlighted under the sensitivity analysis, the difficulty in incorporating variables such as those related to the flooding hazard may have a large impact on the assessment of flooding damages, particularly under the consideration of their different multi-sectoral impacts.

The construction of the BN model does however allow for many possible future developments, building on strong results to either elaborate or improve upon the resulting outputs. The consideration of other possible '*what-if*' scenarios would allow for a better understanding of the likely damages of the increasing frequency and severity of flooding events, and how expected changes in hazard, exposure and vulnerability patterns will likely play into these impacts.

To this aim, the integration of data concerning other flooding events occurring in the same area at different times would provide greater heterogeneity in the training dataset of the model, and thus improve the overall understanding (and then modelling under scenario analysis) of potential damages.

Moreover, a more ambitious development could involve the spatialisation of the output of the model, building on the GIS-based structure of the training dataset, in order to capture damages should the flood extent increase in future events. This would allow for a larger picture on potential damages and provide further support to the management of disaster risk under changing hazard patterns, by widening the scope of potential damages that can be captured.

Overall, despite limitations inherent in the data available for the construction of the presented BN model, the results that were achieved show high promise in the prediction of multi-sectoral flooding damages, and insight into their contributing factors. Building on the previous literature, this work provides a novel approach that improves the understanding of multi-sectoral flooding damages, and in doing so, will add to the state-of-the-art knowledge in the fields of Disaster Risk Management and Climate Change Adaptation. This will provide valuable support for policy- and decision-makers who can use the results of this study to prioritize efficient collection, organization and application of post-disaster damage data, and more efficiently plan ahead for the management of potential future flooding events.

BIBLIOGRAPHY

Al-Jarrah, O. Y., Yoo, P. D., Muhaidat, S., Karagiannidis, G. K., & Taha, K. (2015). Efficient Machine Learning for Big Data: A Review. In *Big Data Research* (Vol. 2, Issue 3, pp. 87–93). Elsevier Inc. https://doi.org/10.1016/j.bdr.2015.04.001

Alpaydin, E. (2020). *Introduction to machine learning*. MIT press.

Amadio, M., Mysiak, J., Carrera, • Lorenzo, & Koks, E. (2016). Improving flood damage assessment models in Italy. *Natural Hazards*, *82*. https://doi.org/10.1007/s11069-016-2286-0

Amadio, M., Scorzini, A. R., Carisi, F., Essenfelder, A. H., Domeneghetti, A., Mysiak, J., & Castellarin, A. (2019). Testing empirical and synthetic flood damage models: the case of Italy. *Hazards Earth Syst. Sci*, *19*, 661–678. https://doi.org/10.5194/nhess-19-661-2019

Aurambout, J.-P., & Lavalle, C. (2016). *UDP - Population density, 2010 - 2050 (JRC LUISA Reference Scenario 2016)*. European Commission, Joint Research Centre (JRC). http://data.europa.eu/89h/jrc-luisa-udp-popden-ref2016

Ayodele, T. O. (2010). *New advances in machine learning 3* (Y. Zhang (ed.)). BoD–Books on Demand.

Balbi, S., Villa, F., Mojtahed, V., Hegetschweiler, K. T., & Giupponi, C. (2016). A spatial Bayesian network model to assess the benefits of early warning for urban flood risk to people. *Hazards Earth Syst. Sci*, *16*, 1323–1337. https://doi.org/10.5194/nhess-16-1323-2016

Campolo, M., Soldati, A., & Andreussi, P. (2003). Artificial neural network approach to flood forecasting in the River Arno. *Hydrological Sciences Journal*, *48*(3), 381–398. https://doi.org/10.1623/hysj.48.3.381.45286

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

Carminati, E., & Martinelli, G. (2002). Subsidence rates in the Po Plain, northern Italy: The relative impact of natural and anthropogenic causation. *Engineering Geology*, *66*(3–4), 241–255. https://doi.org/10.1016/S0013-7952(02)00031-5

Connelly, A., Carter, J., Handley, J., & Hincks, S. (2018). Enhancing the Practical Utility of Risk Assessments in Climate Change Adaptation. *Sustainability*, *10*(5), 1399. https://doi.org/10.3390/su10051399

D'Addabbo, A., Refice, A., Pasquariello, G., Lovergine, F. P., Capolongo, D., & Manfreda, S. (2016). A Bayesian Network for Flood Detection Combining SAR Imagery and Ancillary Data. *IEEE Transactions on Geoscience and Remote Sensing*, *54*(6), 3612–3625. https://doi.org/10.1109/TGRS.2016.2520487

Das, S., Ghosh, A., Hazra, S., Ghosh, T., de Campos, R. S., & Samanta, S. (2020). Linking IPCC AR4 & AR5 frameworks for assessing vulnerability and risk to climate change in the Indian Bengal Delta. *Progress in Disaster Science*, *7*, 100110. https://doi.org/10.1016/j.pdisas.2020.100110

De Santa Olalla, F. J. M., Domínguez, A., Artigao, A., Fabeiro, C., & Ortega, J. F. (2005). Integrated water resources management of the Hydrogeological Unit "Eastern Mancha" using Bayesian Belief Networks. *Agricultural Water Management*, *77*(1–3), 21–36. https://doi.org/10.1016/j.agwat.2004.09.029

Defra. (2011). *Guidelines for Environmental Risk Assessment and Management - Green Leaves III*.

www.defra.gov.uk

Deparday, V., Gevaert, C. M., Molinario, G. M., Soden, R. J., & Balog-Way, S. A. B. (2019). *Machine Learning for Disaster Risk Management*.

Dottori, F., Alfieri, L., Salamon, P., Bianchi, A., Feyen, L., & Lorini, V. (2016). *Flood hazard map for Europe, 10-year return period*. European Commission, Joint Research Centre (JRC). http://data.europa.eu/89h/jrc-floods-floodmapeu_rp10y-tif

EU Floods Directive 2007/60/EC, (2007) (testimony of European Commission).

Furlan, E., Slanzi, D., Torresan, S., Critto, A., & Marcomini, A. (2020). Multi-scenario analysis in the Adriatic Sea: A GIS-based Bayesian network to support maritime spatial planning. *Science of the Total Environment*, *703*, 134972. https://doi.org/10.1016/j.scitotenv.2019.134972

GFDRR. (2018). *Machine Learning for Disaster Risk Management*.

Govi, M., & Maraga, F. (2005). Inundation on the Po Plain caused by levee breaches. *Giornale Di Geologia Applicata*, *1*, 167–176. https://doi.org/10.1474/GGA.2005-01.0-17.0017

Hasanzadeh Nafari, R., Amadio, M., Ngo, T., & Mysiak, J. (2017). Flood loss modelling with FLF-IT: A new flood loss function for Italian residential structures. *Natural Hazards and Earth System Sciences*, *17*(7), 1047–1059. https://doi.org/10.5194/nhess-17-1047-2017

IPCC. (2018). *Global warming of 1.5°C. An IPCC Special Report on the impacts of global warming of 1.5°C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change,*.

ISPRA. (2014). *Mappe di pericolosita idraulica e popolazione es- ` posta a rischio alluvioni in Italia*.

ISPRA. (2018). *Hydrogeological instability in Italy: hazard and risk indicators*.

ISTAT. (2011). *15◦ censimento della populazione e delle abitazioni, Istituto nazionale di statistica*.

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. (2012). Comparative flood damage model assessment: Towards a European approach. *Natural Hazards and Earth System Science*, *12*(12), 3733–3752. https://doi.org/10.5194/nhess-12-3733-2012

Kotsiantis, S. B., Zaharakis, I., & Pintelas, P. (2007). Supervised machine learning: A review of classification techniques. *Emerging Artificial Intelligence Applications in Computer Engineering*, *160*, 3–24.

Kragt, M. E. (2009). A beginners guide to Bayesian network modelling for integrated catchment management. *Landscape Logic*.

Kreibich, H., Seifert, K., Kunert, I., Schwarz, U., Merz, J., Thieken, B., Piroth, K., Seifert, I., Maiwald, H., Kunert, U., Schwarz, J., Merz, B., & Thieken, A. H. (2009). Is flow velocity a significant parameter in flood damage modelling. In *Nat. Hazards Earth Syst. Sci* (Vol. 9, Issue 5). www.nat-hazards-earth-syst-sci.net/9/1679/2009/

Lamovec, P., Veljanovski, T., Mikoš, M., & Oštir, K. (2013). Detecting flooded areas with machine learning techniques: case study of the Selška Sora river flash flood in September 2007. *Journal of Applied Remote Sensing*, *7*(1), 073564. https://doi.org/10.1117/1.jrs.7.073564

Lavalle, C. (2014). *OUTPUT - Land-use/cover maps (LUISA Platform REF2014)*. European Commission, Joint Research Centre (JRC). http://data.europa.eu/89h/jrc-luisa-land-use-ref-2014

Lin, Yun, Bhardwaj, & Hill. (2019). Urban Flood Detection with Sentinel-1 Multi-Temporal Synthetic

Aperture Radar (SAR) Observations in a Bayesian Framework: A Case Study for Hurricane Matthew. *Remote Sensing*, *11*(15), 1778. https://doi.org/10.3390/rs11151778

Lombardi, A., Tomassetti, B., Colaiuda, V., & Verdecchia, M. (2018). Increasing of severe hydrological events in the Po basin under global warming. In *Geophysical Research Abstracts* (Vol. 20). https://ui.adsabs.harvard.edu/abs/2018EGUGA..20.6354L/abstract

Mayfield, H. J., Smith, C. S., Lowry, J. H., Watson, C. H., Baker, M. G., Kama, M., Nilles, E. J., & Lau, C. L. (2018). Predictive risk mapping of an environmentally-driven infectious disease using spatial Bayesian networks: A case study of leptospirosis in Fiji. *PLOS Neglected Tropical Diseases*, *12*(10), e0006857. https://doi.org/10.1371/journal.pntd.0006857

McNaught, K. R., & Zagorecki, A. (2009). Using dynamic Bayesian networks for prognostic modelling to inform maintenance decision making. *IEEM 2009 - IEEE International Conference on Industrial Engineering and Engineering Management*, 1155–1159. https://doi.org/10.1109/IEEM.2009.5372973

Menderes, A., Erener, A., & Sarp, G. (2015). Automatic Detection of Damaged Buildings after Earthquake Hazard by Using Remote Sensing and Information Technologies. *Procedia Earth and Planetary Science*, *15*, 257–262. https://doi.org/10.1016/j.proeps.2015.08.063

Merz, B., Kreibich, H., Schwarze, R., & Thieken, A. (2010). Natural Hazards and Earth System Sciences "Assessment of economic flood damage." *Hazards Earth Syst. Sci*, *10*, 1697–1724. https://doi.org/10.5194/nhess-10-1697-2010

Notaro, V., Fontanazza, C. M., La Loggia, G., & Freni, G. (2014). Identification of the best flood retrofitting scenario in an urban watershed by means of a Bayesian Decision Network. *WIT Transactions on the Built Environment*, *139*, 341–352. https://doi.org/10.2495/UW140291

Orlandini, S., Moretti, G., & Albertson, J. D. (2015). Evidence of an emerging levee failure mechanism causing disastrous floods in <scp>I</scp> taly. *Water Resources Research*, *51*(10), 7995–8011. https://doi.org/10.1002/2015WR017426

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), 323–343. https://doi.org/10.5194/nhess-20-323-2020

Paprotny, Dominik, Kreibich, H., Morales-Nápoles, O., Castellarin, A., Carisi, F., & Schröter, K. (2020). Exposure and vulnerability estimation for modelling flood losses to commercial assets in Europe. *Science of the Total Environment*, *737*, 140011. https://doi.org/10.1016/j.scitotenv.2020.140011

Paprotny, Dominik, Kreibich, H., Morales-Nápoles, O., Wagenaar, D., Castellarin, A., Carisi, F., Bertin, X., Terefenko, P., Merz, B., & Schröter, K. (2020). *Flood exposure and vulnerability estimation methods for residential and commercial assets in Europe*. https://doi.org/10.5194/egusphere-egu2020-4753

Poelhekke, L., Jäger, W. S., van Dongeren, A., Plomaritis, T. A., McCall, R., & Ferreira, Ó. (2016). Predicting coastal hazards for sandy coasts with a Bayesian Network. *Coastal Engineering*, *118*, 21–34. https://doi.org/10.1016/j.coastaleng.2016.08.011

Pollino, C. A., Woodberry, O., Nicholson, A., Korb, K., & Hart, B. T. (2007). Parameterisation and evaluation of a Bayesian network for use in an ecological risk assessment. *Environmental Modelling and Software*, *22*(8), 1140–1152. https://doi.org/10.1016/j.envsoft.2006.03.006

Rosser, J. F., Leibovici, D. G., & Jackson, M. J. (2017). Rapid flood inundation mapping using social media, remote sensing and topographic data. *Natural Hazards*, *87*(1), 103–120.

https://doi.org/10.1007/s11069-017-2755-0

Sadler, J. M., Goodall, J. L., Morsy, M. M., & Spencer, K. (2018). Modeling urban coastal flood severity from crowd-sourced flood reports using Poisson regression and Random Forest. *Journal of Hydrology*, *559*, 43–55. https://doi.org/10.1016/j.jhydrol.2018.01.044

Sairam, N., Schröter, K., Lüdtke, S., Merz, B., & Kreibich, H. (2019). Quantifying Flood Vulnerability Reduction via Private Precaution. *Earth's Future*, *7*(3), 235–249. https://doi.org/10.1029/2018EF000994

Sairam, N., Schröter, K., Rözer, V., Merz, B., & Kreibich, H. (2019). Hierarchical Bayesian Approach for Modeling Spatiotemporal Variability in Flood Damage Processes. *Water Resources Research*, *55*(10), 8223–8237. https://doi.org/10.1029/2019WR025068

Sayers, P. B., Hall, J. W., & Meadowcroft, I. C. (2002). Towards risk-based flood hazard management in the UK. *Proceedings of the Institution of Civil Engineers: Civil Engineering*, *150*(1 SPECIAL ISSUE), 36–42. https://doi.org/10.1680/cien.2002.150.5.36

Schröter, K., Kreibich, H., Vogel, K., Riggelsen, C., Scherbaum, F., & Merz, B. (2014). How useful are complex flood damage models? *Water Resources Research*, *50*(4), 3378–3395. https://doi.org/10.1002/2013WR014396

Scutari, M. (2010). Learning Bayesian Networks with the bnlearn R Package. *Journal of Statistical Software*, *35*(3), 1–22.

Sharma, J., & Ravindranath, N. H. (2019). Applying IPCC 2014 framework for hazard-specific vulnerability assessment under climate change. *Environmental Research Communications*, *1*(5), 051004. https://doi.org/10.1088/2515-7620/ab24ed

Shustikova, I., Neal, J. C., Domeneghetti, A., Bates, P. D., Vorogushyn, S., & Castellarin, A. (2020). Levee breaching: A new extension to the LISFLOOD-FP model. *Water (Switzerland)*, *12*(4), 942. https://doi.org/10.3390/W12040942

Spekkers, M. H., Kok, M., Clemens, F. H. L. R., & ten Veldhuis, J. A. E. (2014). Decision-tree analysis of factors influencing rainfall-related building structure and content damage. *Natural Hazards and Earth System Sciences*, *14*(9), 2531–2547. https://doi.org/10.5194/nhess-14-2531-2014

Sperotto, A., Molina, J. L., Torresan, S., Critto, A., & Marcomini, A. (2017). Reviewing Bayesian Networks potentials for climate change impacts assessment and management: A multi-risk perspective. In *Journal of Environmental Management* (Vol. 202, pp. 320–331). Academic Press. https://doi.org/10.1016/j.jenvman.2017.07.044

Sperotto, A., Molina, J. L., Torresan, S., Critto, A., Pulido-Velazquez, M., & Marcomini, A. (2019). A Bayesian Networks approach for the assessment of climate change impacts on nutrients loading. *Environmental Science and Policy*, *100*, 21–36. https://doi.org/10.1016/j.envsci.2019.06.004

Tangney, P. (2019). Understanding climate change as risk: a review of IPCC guidance for decision-making. *Journal of Risk Research*. https://doi.org/10.1080/13669877.2019.1673801

Tarquini, S., Isola, I., Favalli, M., & Battistini, A. (2007). *TINITALY, a digital elevation model of Italy with a 10 m-cell size (Version 1.0)*. Istituto Nazionale di Geofisica e Vulcanologia (INGV). https://doi.org/https://doi.org/10.13127/TINITALY/1.0.

Thieken, A. H., Ackermann, V., Elmer, F., Kreibich, H., Kuhlmann, B., Kunert, U., Maiwald, H., Merz, B., Müller, M., Piroth, K., Schwarz, J., Schwarze, R., Seifert, I., & Seifert, J. (2009). *4 th International Symposium on Flood Defence: Managing Flood Risk, Reliability and Vulnerability*

METHODS FOR THE EVALUATION OF DIRECT AND INDIRECT FLOOD LOSSES. Deutsches GeoForschungsZentrum GFZ.

Tockner, K., Uehlinger, U., & Robinson, C. (2009). *Rivers of Europe*. Academic Press. https://doi.org/https://doi.org/10.1016/B978-0-12-369449-2.X0001-X

UNFCCC. (2020). *Online Guide to Loss and Damage*. https://unfccc.int/sites/default/files/resource/Online_Guide_feb_2020.pdf

Vacondio, R, Dal Palù, A., & Mignosa, P. (2014). GPU-enhanced finite volume shallow water solver for fast flood simulations. *Environmental Modelling and Software*, *57*, 60–75. https://doi.org/10.1016/j.envsoft.2014.02.003

Vacondio, Renato, Aureli, F., Ferrari, A., Mignosa, P., & Dal Palù, A. (2016). Simulation of the January 2014 flood on the Secchia River using a fast and high-resolution 2D parallel shallow-water numerical scheme. *Natural Hazards*, *80*(1), 103–125. https://doi.org/10.1007/s11069-015-1959-4

Wagenaar, D., Curran, A., Balbi, M., Bhardwaj, A., Soden, R., Hartato, E., Mestav Sarica, G., Ruangpan, L., Molinario, G., & Lallemant, D. (2019). Invited perspectives: How machine learning will change flood risk and impact assessment. *Natural Hazards and Earth System Sciences Discussions*, 1–23. https://doi.org/10.5194/nhess-2019-341

Wagenaar, D., De Jong, J., & Bouwer, L. M. (2017). Multi-variable flood damage modelling with limited data using supervised learning approaches. *Hazards Earth Syst. Sci*, *17*, 1683–1696. https://doi.org/10.5194/nhess-17-1683-2017

Wagenaar, D., Lüdtke, S., Schröter, K., Bouwer, L. M., & Kreibich, H. (2018). Regional and Temporal Transferability of Multivariable Flood Damage Models. *Water Resources Research*, *54*(5), 3688–3703. https://doi.org/10.1029/2017WR022233

Wang, Z., Lai, C., Chen, X., Yang, B., Zhao, S., & Bai, X. (2015). Flood hazard risk assessment model based on random forest. *Journal of Hydrology*, *527*, 1130–1141. https://doi.org/10.1016/j.jhydrol.2015.06.008

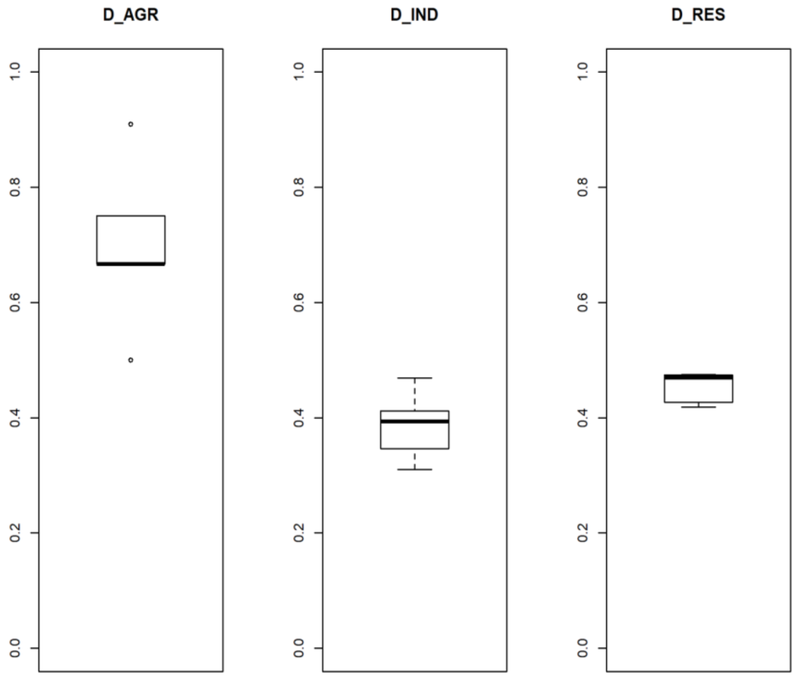
Wing, O. E. J., Pinter, N., Bates, P. D., & Kousky, C. (2020). New insights into US flood vulnerability revealed from flood insurance big data. *Nature Communications*, *11*(1), 1–10. https://doi.org/10.1038/s41467-020-15264-2

Yadav, S., & Shukla, S. (2016). Analysis of k-Fold Cross-Validation over Hold-Out Validation on Colossal Datasets for Quality Classification. *Proceedings - 6th International Advanced Computing Conference, IACC 2016*, 78–83. https://doi.org/10.1109/IACC.2016.25

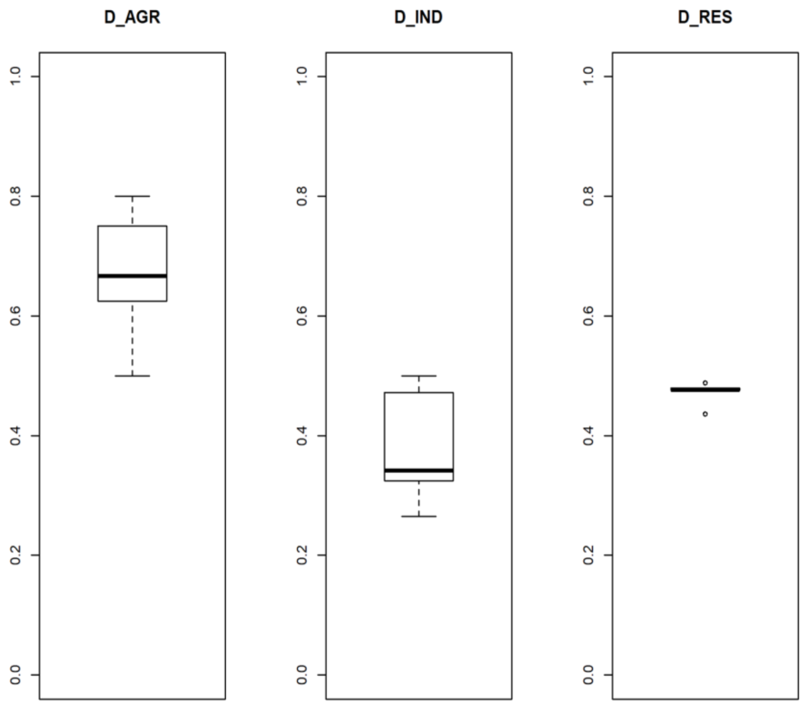
Zanchettin, D., Traverso, P., & Tomasino, M. (2008). Po River discharges: A preliminary analysis of a 200-year time series. *Climatic Change*, *89*(3–4), 411–433. https://doi.org/10.1007/s10584-008-9395-z

Annex I – Validation of BN model configurations 1A-1D

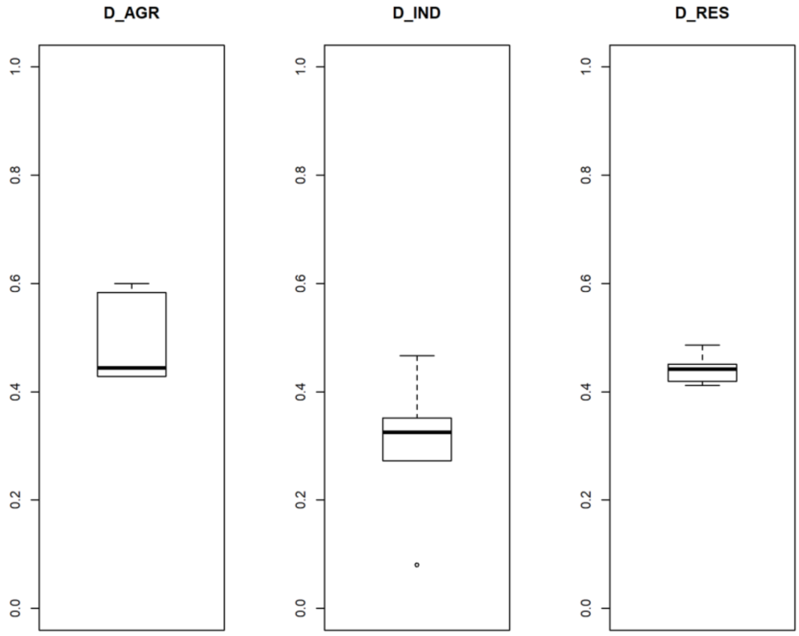
CONF 1A



CONF 1B



CONF 1C



CONF 1D