Climate Risk Management 35 (2022) 100410

Contents lists available at ScienceDirect

Climate Risk Management

journal homepage: www.elsevier.com/locate/crm

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

Remi Harris [a,b], Elisa Furlan [a,b], Hung Vuong Pham [a,b], Silvia Torresan [a,b], Jaroslav Mysiak [a,b], Andrea Critto [a,b,*]

[a] *Department of Environmental Sciences, Informatics and Statistics, University Ca' Foscari Venice, I-30170 Venice, Italy*
[b] *Fondazione Centro-Euro-Mediterraneo sui Cambiamenti Climatici, I-73100 Lecce, Italy*

ARTICLE INFO

Keywords:
Machine Learning
Flood risk assessment
Climate adaptation
Sensitivity analysis
Secchia river

ABSTRACT

Extreme weather and climate related events, from river flooding to droughts and tropical cyclones, are likely to become both more severe and more frequent in the coming decades, and the damages caused by these events will be felt across all sectors of society. 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 can capture the full extent of the impacts. In this frame, a GIS-based Bayesian Network (BN) approach is presented for the capturing and modelling of multi-sectoral flooding damages against future 'what-if' scenarios. Building on a risk-based conceptual framework, the BN model was trained and validated by exploiting data collected from the 2014 Secchia River flooding event, as well as other contextual variables. Moreover, a novel approach to defining the structure of the BN was performed, reconfiguring the model according to expert judgment and data-based validation. The model showed a good predictive capacity for 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. 'What-if' scenario analysis was performed to understand the potential impacts of future changes in i) land use patterns and ii) increasing flood depths resulting from more severe flood events. The output of the model showed a rising probability of experiencing high monetary damages under both scenarios. 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 actions against extreme river flooding events, enabling better informed decision making.

1. Introduction

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).

* Corresponding author at: Informatics and Statistics, University Ca' Foscari Venice, Via Torino 155, 30170 Venezia-Mestre, Italy.
E-mail address: critto@unive.it (A. Critto).

https://doi.org/10.1016/j.crm.2022.100410
Received 20 April 2021; Received in revised form 13 October 2021; Accepted 30 January 2022
Available online 3 February 2022

In the face of this rising threat, 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 future scenarios. In terms of flood damage, sophisticated models are being introduced that can draw from expert knowledge, increased variable input, and more powerful computational methods, including those exploiting functionalities offered by Machine Learning (ML) (Amadio et al., 2016; Jongman et al., 2012; Thieken et al., 2009; Wing et al., 2020). Among these, upcoming Bayesian Network (BN) approaches (Paprotny et al., 2020a; Schröter et al., 2014; Wagenaar et al., 2019), represent useful methods able to provide predictive hazard and impact assessment capabilities based on real-world training datasets.

BN are statistical approaches built in the form of qualitative structures, known as directed acyclic graphs, representing the variables of concern as nodes on the graph, with arcs to characterize the probabilistic dependencies among variables at stake in 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 (Li et al., 2016; Sperotto et al., 2017). These methods can be designed to tackle complex environmental issues featured by non-linear behavior and hampered by large uncertainties (Sperotto et al., 2017). This flexibility also 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 depth-damage models (Schröter et al., 2014). Overall, BN models show a notable added value in flood damage modelling, regularly out-performing 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 (Schröter et al., 2014; Wagenaar et al., 2017), trained either on a local scale or to multiple case studies for checking BN performance in a transfer setting (Sairam et al., 2019; Wagenaar et al., 2018). Where others have focused solely on the residential sector, Paprotny et al. (2020b) introduced a new flood damage model for the commercial sector, this representing the first of its kind in the field. However, the evaluation of the overall multi-sectoral damages within any case study still remains unclear, with minimal efforts devoted to a wider sectoral analysis that could, for instance, incorporate infrastructure, agricultural or industrial concerns as assessment endpoints.

Furthermore, where scenario analysis has been conducted, it was limited to a retrospective investigation into one of the factors that influence the experienced flood damages, such as mitigation options or early warning systems (Balbi et al., 2016; Notaro et al., 2014), and as such these models have not been used for the projection of future changes in damages.

Accordingly, this work aims to build on the state-of-the-art research in the field of ML, through the design and application of an advanced GIS-based BN approach able to capture and model multi-sectoral damages against multiple '*what-if*' scenarios, exploiting damage data collected against the 2014 Secchia river flooding event, as training dataset of the BN model.

In the following sections, the investigated case study and its associated dataset are described (Section 2), followed by a detailed description of the methodological approach underpinning BN design and implementation for '*what-if*' scenario analysis (Section 3). The results of the model output and the main conclusions drawn are discussed in Sections 4 and 5.

2. Case study and data collection

2.1. Geographical context and the 2014 Secchia river flooding event

The largest and most economically important river basin in Italy is the Po Valley (Tockner et al., 2009), which produces nearly half of the national GDP from one third the country's industries, including a large agricultural output (Amadio et al., 2019). It is also significantly populated: in 2011 its eight regions were home to 17 million people (about 29% of the Italian population) (ISTAT, 2011).

The area is traditionally flood-prone (Lombardi et al., 2018), and flooding frequency in low-lying areas 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 winter river discharges (Coppola et al., 2014; Montanari, 2012).

Flood 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, together with the increasing protection in Upper Po areas have, according to some models, increased the value of the flood peak at any given probability, and further exacerbated the risk to the lower basin regions (Zanchettin et al., 2008).

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).

In January 2014, the Secchia river basin was hit by a long-lasting stratiform rainfall event, which 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 1 m 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. The affected area (Fig. 1) 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. A similar potential levee failure was later spotted on the Panaro and fixed, however the situation around the Secchia river could not be averted (Orlandini et al., 2015).

The impacts of the flood were widespread and devastating for many. Over 50 km^2 was flooded at an average depth of 1 m, 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. The majority of the damage was to the industrial sector, with residential properties experiencing €36 million worth of damage, 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 damaged. 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).

2.2. Data collection and Pre-processing

A large quantity of data encompassing all analytical dimensions (i.e. those relating to the hazards and the exposed targets, across temporal and spatial scales) is required for informing risk assessment and modelling. The IPCC's risk framework (IPCC, 2018), which 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), was chosen as an appropriate means for identifying key data encompassing the different interacting components of disaster risk, i.e. hazard, exposure, and vulnerability.

For these purposes, data relating to the case study has been collected, both in terms of the event that occurred around the Secchia river in 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. Several other sources of heterogeneous data were retrieved for integration within the dataset for analysis, particularly regarding the microscale exposure and vulnerability of the affected buildings, assets, and population (metadata are reported in Table 1). The model parameters were selected through a multi-stage approach, beginning with an in-depth review of the literature for similar applications that focused on single sector models to gain an understanding on the parameters that have previously been effectively used to evaluate flooding damages through a Machine Learning approach.

The next stage concerned the review of the available indicator databases, paying attention to available datasets such as open-source modelling and socio-economic data. Where data applicable to the case study area was available, the possible parameters for application were then narrowed down to those that contained a suitably high temporal and spatial resolution.

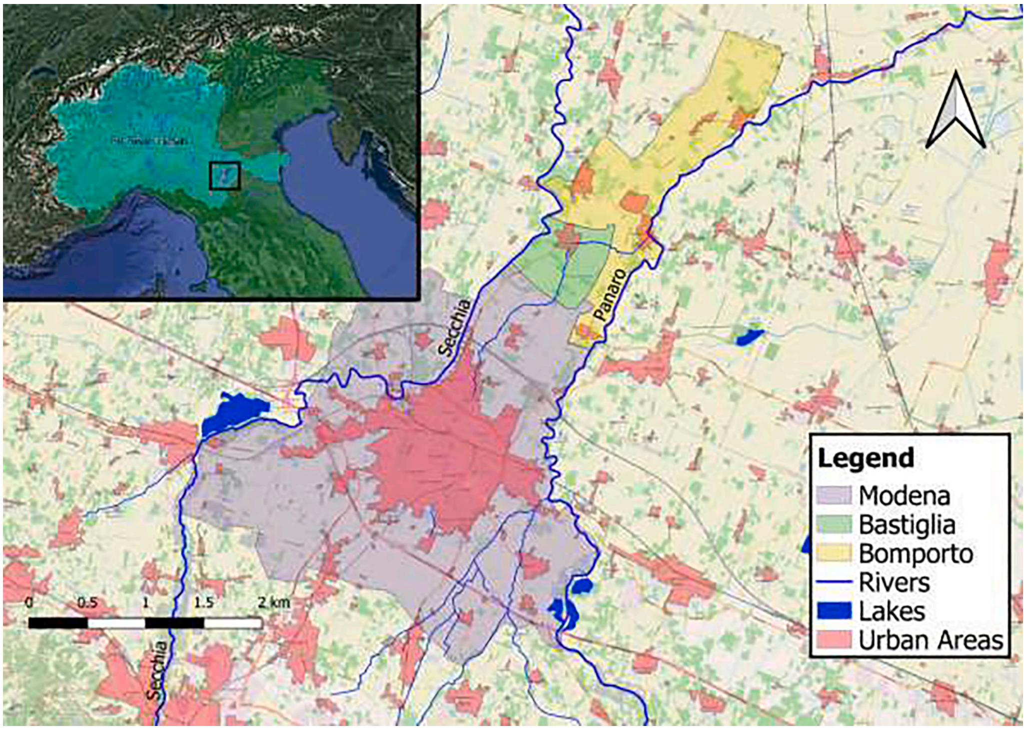

Fig. 1. Map of the municipalities affected by the 2014 flood, and their location within the Po Basin.

Table 1
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	5 m	19/1/14–22/1/14	Vacondio et al. (2016)
Maximum flow velocity	$m\ s^{-1}$	Raster	Local	5 m	19/1/14–22/1/14	Vacondio et al. (2016)
Duration of inundation	h	Raster	Local	5 m	19/1/14–22/1/14	Vacondio et al. (2016)
Exposure data						
Structure area	m^2	Tabular	Local	1 m^2	2014	Damage claims
Population density	person/km^2	Raster	Europe	1 km	2010	Aurambout & Lavalle (2016); https://data.jrc.ec.europa.eu/dataset/jrc-luisa-udp-popden-ref2016
Vulnerability data						
Number of stories	–	Vector	National	Sub-municipal	2011	ISTAT (2011); https://www.istat.it/it/archivio/104317
Digital Elevation Model	m	Raster	Europe	10 m	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	100 *m*	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 LUISA	–	Raster	Europe	1 km	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	100 m	10–500 years	Dottori et al. (2016); https://data.jrc.ec.europa.eu/collection/id-0054

2.2.1. Hazard

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, allowing together to produce and validate flood hazard maps for the area and event of concern (Vacondio et al., 2016). Specifically, 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 (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 outputs of this modelling, as available for use, are three 5 m spatially resolved hazard metrics, depicting the maximum velocity, depth, and duration of inundation respectively.

2.2.2. Exposure and vulnerability

Additional data on the vulnerability and exposure of hazard-prone people and properties can be used to enhance the predictive capacity of BN models (Paprotny et al., 2020a). 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, as reported in Table 1, a range of variables (e.g., area of reported damages, building age, and land use) 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. Among these, the area of reported damages corresponded to the overall footprint of the damaged building as self-reported by the individual claimants (against the 2014 Secchia flood event), and does not refer to the specific area that was flooded.

2.2.3. Damages

Empirical damage-related information for residential buildings affected, as well as agricultural and industrial assets, has been collected by the local municipalities following the 2014 Secchia flooding event. This information represents 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, damage claims across all three sectors are equal to €71.5 million (in 2014 Euro value).

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 did not contain information on the structure area, a total of 1738 complete data points were used, of which 65% are from Bastiglia, 31% from Bomporto, and the remaining 4% from Modena.

Damage data points were automatically geolocated using the “Geocode” function in “MMQGIS” plugin of QGIS (https://github.com/michaelminn/mmqgis/), by matching their address to the address as listed in Open Street Map (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 (https://www.google.com/earth/), by matching the address of the reported damage to the equivalent address in Google Earth. 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 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 for electricity, heating, and water.

The data collected for agriculture and industry again contained information 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.

3. Methodological development

The construction of a BN model follows a stepwise progression (Deparday et al., 2019; Furlan et al., 2020; Sperotto et al., 2017), from model conceptualization to parametrization and validation, followed by the use of the designed model for scenario and sensitivity analysis. Following a brief discussion of the theoretical aspects behind this approach, the specific details of these developmental steps are presented in the following sections, detailing the methodological approach underpinning the BN development for the case study of concern.

3.1. Bayesian belief networks

As previously discussed, BN are statistical approaches built in the form of directed acyclic graphs, that represent the variables of concern as nodes on the graph, with arcs to characterize the probabilistic dependencies among variables at stake in the system (Landuyt et al., 2013). These probabilities are calculated on the basis of Bayes’ theorem, which relates the posterior condition of a system to its prior position through probabilistic inference. This is detailed in the below equation (Joyce, 2021):

$$\mathrm{P}(\mathrm{A}|\mathrm{B}) = \frac{\mathrm{P}(\mathrm{B}|\mathrm{A})\mathrm{P}(\mathrm{A})}{P(B)}$$

Where:P(A) and P(B) are the prior probability distribution of the observed events A and B.P(A|B) is the posterior probability of observing the event A given that B is true.P(B|A) is the probability of observing the event B given that A is true.

The prior distribution of each event is computed practically by dividing the input data into a series of discrete bins according to their value, and hence represent the probability of an individual element falling within one of these states. Through the application of Bayes’ theorem, the posterior probability of events occurring are propagated through the nodes of the system according to the values that are input in the training of the statistical model (Sperotto et al., 2019). In this way, the model can be adapted considering new evidence in the form of observational data or under a narrative scenario, and calculating the posterior probability of all other nodes within the BN.

In comparison to, for example, a GIS overlay method, a major added value of the BN model derives from its ability to learn from the data to make a probabilistic assessment of current and future risk, and understand how the parameters of hazard, exposure, and vulnerability interact and influence each other within the same assessment framework (and case study). As opposed to a classical GIS overlay method, this information is extracted by the model itself to produce an output that can be interpreted to understand the risk across the whole case study area, and against future scenario analysis, as implemented here.

3.2. Model design and configuration testing

The practical implementation of the BN requires a multi-stage approach, beginning with the design of the BN model and the subsequent parametrization of the variables at stake in the network. 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 for 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).

Building on the IPCC risk framework, BN variables have been selected and classified in terms of hazard, exposure and vulnerability, through the considered judgment of the authors and an associated literature review (Merz et al., 2010; Paprotny et al., 2020b), according to their potential influence on the overall flooding risk. This causal relationship has been depicted in Fig. 2 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 the identified variables. The boxes of the diagram correspond to the BN nodes that in turn represent the system variables, with unidirectional arrows between the boxes depicting the arcs that determine the causal relationships between variables in the model (Sperotto et al., 2019).

An alternative approach to understanding the optimal model performance is by analyzing various configurations of the network, as defined through knowledge-based judgement and pertinent literature (Poelhekke et al., 2016). By setting these different configurations and observing the respective model outputs, also in terms of the accuracy of the model simulation performance, it is possible to identify which models perform best in comparison to one another, and define, in a relative sense, the "optimal" model performance.

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 (e.g. the hazard-related characteristic linked to the Secchia river 2014 flood event), 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 (Poelhekke et al., 2016).

The second stage looks at reconfiguring the model (to improve its performance) by identifying, through literature review and considered 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 testing of the multiple BN configurations, 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. Specifically, these variable states can be defined in three different ways, namely i) into

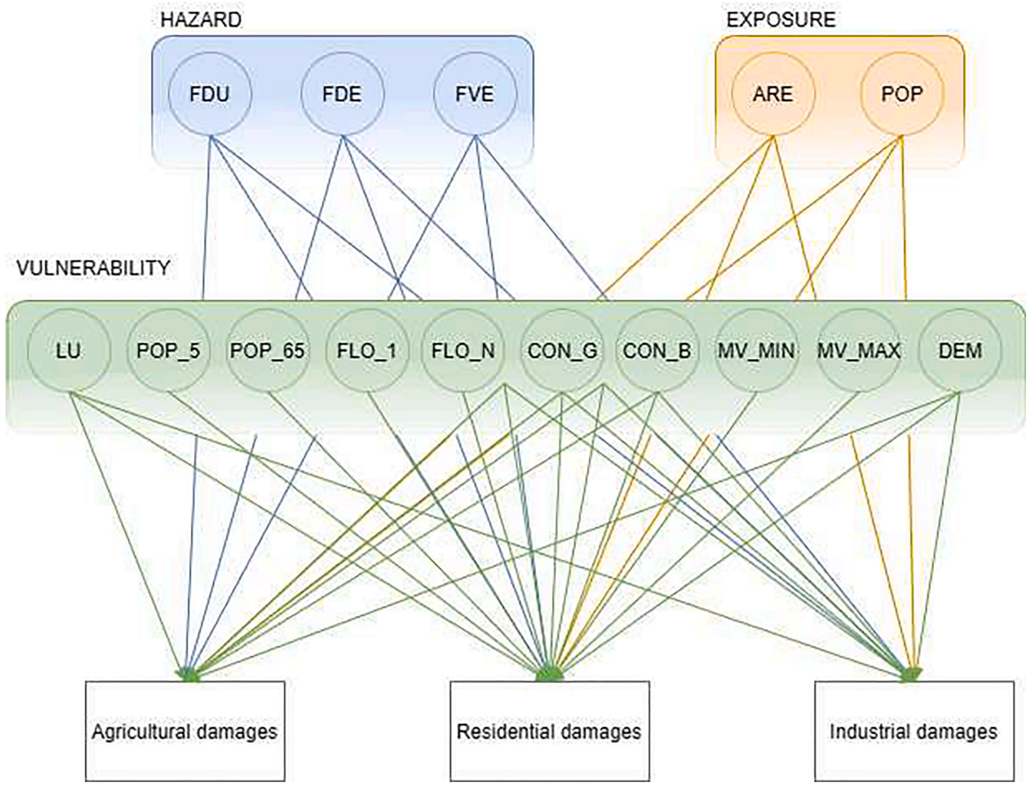

Fig. 2. Risk-based BN Conceptual Framework. List of acronyms used for hazard variables: FDU: Flood duration; FDE: Maximum flood depth; FVE: Maximum flood velocity; Exposure variables: ARE: Area of reported damage; POP: Population density; Vulnerability variables: LU: Land Use Cover; POP_5: Population under 5; POP_65: Population over 65; FLO_1: Number of houses with 1 story; FLO_N: Number of houses with greater than 1 story; 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.

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.

The model training and validation relies on deterministic data, this approach being borne of the data available for analysis within the research. The model is trained to learn from empirically observed data, that is assumed to be reliable, and suitable to validate the model performance. While this means that the model, as trained, is specific to the 2014 flooding event, there is also a probabilistic element to the analysis, in particular concerning the two projected future scenarios where the data applied are drawn from probabilistic models. This approach was considered necessary to build the future scenario analysis on modelled projections, rather than applying deterministic narratives.

3.3. Model validation

When the simulated 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 the 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). In this way, the model error corresponds to a direct comparison between the values obtained from simulation and data from empirical observation, which is measured in terms of the probability of label misclassification (Peña et al., 2005; Rodriguez et al., 2010).

3.4. Scenario analysis

Once the BN model has been trained and validated it can be used for scenario analysis, in which various potential scenarios are studied in order to simulate their respective impacts across the variables at stake in the network. The conditions of these scenarios are simulated by 'setting' different evidence for one or more nodes (variables) within the BN model, and then propagating that information through the system, thus inferring the behaviour of the variables in order to observe changes in the 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).

Within this study, for the analysis of future '*what-if*' scenarios, variables were 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 (Land Use-based Integrated Sustainability Assessment) model (Lavalle, 2014), against the expected land cover changes up to 2050 from the same model projection. This involves classification of the land into three main categories, these being urban fabric, industry/commercial/services, and agriculture. Specifically, the aggregated changes mostly concern a loss of agricultural land, replaced by urban areas. There are also changes in some areas to and from industrial zoning within the case study area, however the overall share of industrial land remains roughly equal. Data for the 2010 timeframe was used for the baseline scenario to represent land use associated to the 2014 flooding event, being the closest in time. The 40-year time frame used for scenario analysis was therefore built upon land use changes as observed from 2010 to 2050, where 2050 was the widest time frame available, providing the most long-term future insight. The projected changes in land use zoning were minimal in general, but particularly so in the short-term projections; impacts of the land use changes on the model output were more evident under the most distant projection.

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 (Vacondio et al., 2016). This model was sourced from the JRC LISFLOOD-FP flood hazard model dataset (Dottori et al., 2016), which contained projections of flood depth under return periods from 10 to 500 years. The 2014 Secchia flooding event was described by Sairam et al. (2020), as a 5-year return period flooding event, and as such the 10-year return period projection of the JRC model represented the closest available dataset to train the baseline scenario in the BN model. Then, to model the scenario SC_FDE, we compared the distribution of flood depth under the 10-year return period, with that of a 200-year return period, representing a medium probability event, as defined by the European Flood Directive (EC, 2007; Ward et al., 2011). The configuration of the LISFLOOD-FP model (Distributed Water Balance and Flood Simulation Model) itself considers the non-linear dynamic of land use and related parameters using sub- grid variability in land use, as well as the non-linear patterns in rainfall runoff processes and their associated effects on different surface types (Burek et al., 2013). This comparative analysis was chosen in order to best represent a relative increase in flooding depth, by finding 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.

3.5. Sensitivity analysis

A sensitivity analysis could then be performed on the output of the BN model, to provide information on the sensitivity of the BN assessment endpoints (i.e. damages to the residential, agricultural, and industrial sectors), in relation to changes in their various explanatory nodes.

Under this analysis, the stepwise modification of individual input parameters is used to observe changes in the damage assessment endpoint probabilities, and thus determine the impacts of each parameter on the output (Kragt, 2009; Pollino et al., 2007). As such, it would be possible to interpret the relative importance of the various input nodes in determining the highest class of flood damages (Furlan et al., 2020).

Specifically, in each iteration of the sensitivity analysis, the likelihood of the highest class of each variable was in turn assigned a 100% probability, and the changes in the probability distributions were observed for the different assessment endpoint nodes. This analysis, as well as the scenario analysis, give an output that cannot be interpreted in a spatially explicit manner. However their remains the ability to analyze these results for each individual sector, retaining an element of disaggregation.

4. Results

4.1. Model design and configurations testing

As described in Section 3.2, an initial risk-based conceptual framework was established that highlighted the cause-effect relationships between the various components of disaster risk (i.e. hazard, exposure, and vulnerability), and the resulting damages.

According to the data available, this framework was then translated into an expert-based BN conceptual model, as shown in Fig. 2, that provided the basis for the practical BN model implementation. 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 their related child nodes, i.e. the multi-sectoral damages.

In order to introduce the data into the model, variable states were defined, 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 into three intervals with similar data frequencies. Other qualitative variables were categorized by the intrinsic characteristics of the data, such as the 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.

Limitations in the quantity of damage data for training would constrain the ability of the model to incorporate the full set of input variables. As such, building on the risk-based conceptual framework as shown in Fig. 2, a two-tiered approach to the BN model configuration was conceptualized (as also discussed in section 3.2), allowing 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 notably restricted. Specifically, the first iteration of this new model under the first tier of model reconfiguration (hereafter CONF_1A) utilized only the hazard variables (i.e. maximum flood depth (FDE), velocity (FVE), and duration (FDU)) as explanatory metrics of multi-sectoral damages. The impacts of the input variables showed 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 then 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 tend to flatten. This evolution was particularly pronounced for the agricultural sector, again due to the data-poor condition. Accordingly, 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 CONF_1 of the BN model are shown in Fig. 3.

Then, the second tier of model configuration involved the 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 (ARE), which then leads to the multi-sectoral damages (as also developed in Paprotny et al., 2020b; Sayers et al., 2002). Fig. 4 shows the new **CONF_2** as selected for analysis in the validation stage (Section 4.2).

4.2. Model validation

Once the different configurations of the BN model were set, with the appropriate parametrization of the chosen variables, the models can be validated in order to give an estimation of their prediction error, and then the model with the highest performance in the estimation of multi-sectoral damages can be selected. As explained in Section 3.3, data-based validation was performed to determine the probability of observational data being misclassified in the three damage-related assessment endpoints. Specifically, the k-fold validation was applied for this analysis, and carried out through the *bn.cv* function in R with the “k-fold” method (Scutari, 2010). Due to the limited amount of observational data available for the agricultural sector, the number of folds used for the validation process was limited to five, with the results of this analysis for the most accurate model, **CONF_1C,** illustrated in Fig. 5.

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. 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 ‘what-if’ scenario analysis.

4.3. Scenario analysis

After validation of the model, it was then possible to apply it for inferential purposes, including the analysis of two potential scenarios as described in Section 3.4. For each scenario, evidences were set based on the projected dataset for the variables of concern (i.e. land use/cover, flood depth, detailed in Table 1), and the changes were then propagated downward, with the output (posterior probabilities) recorded for comparison with the original conditional probabilities.

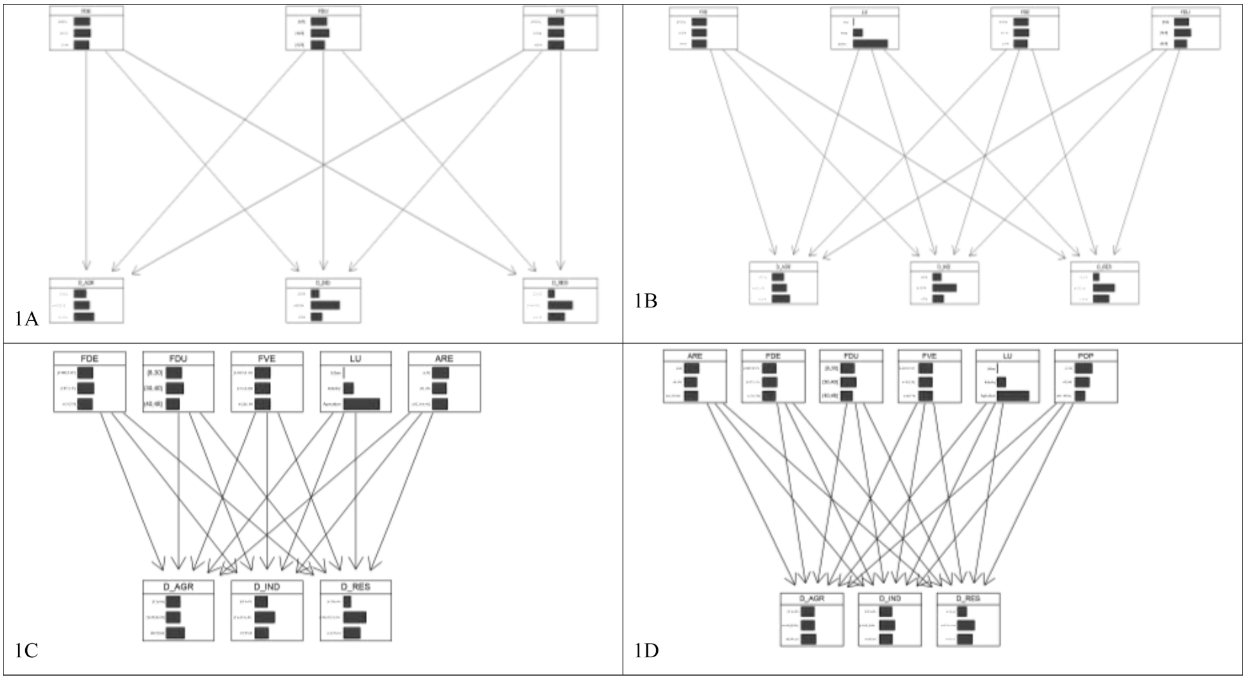

Fig. 3. Outputs for the definition of the BN model from CONF_1A (top left) to CONF_1D. List of acronyms used for Explanatory nodes: FDU: Flood duration; FDE: Maximum flood depth; FVE: Maximum flood velocity; ARE: Area of reported damage; POP: Population density; LU: Land Use Cover; Damage nodes: D_AGR: Agricultural; D_IND: Industrial; D_RES: Residential.

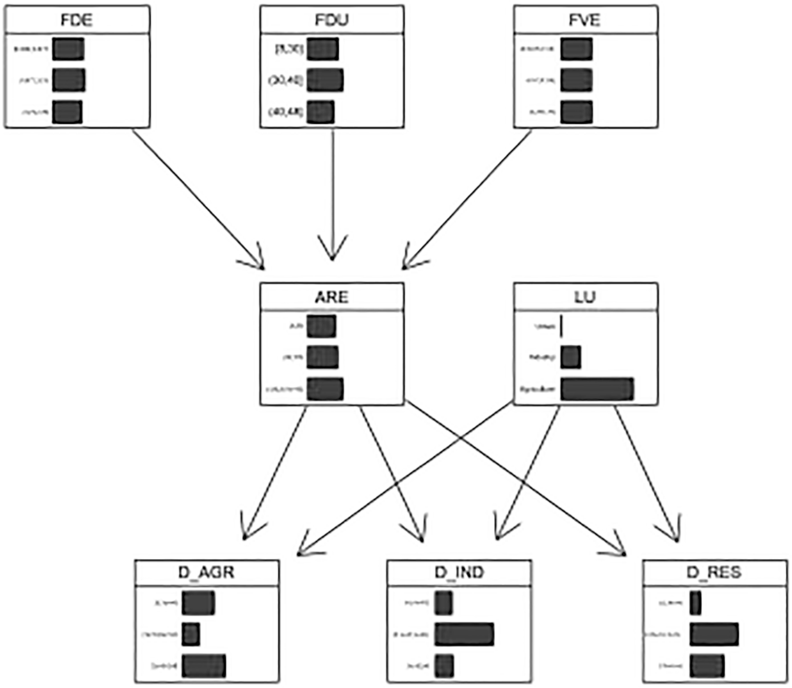

Fig. 4. CONF_2 - BN model for multi-sectoral flooding damages assessment with associated variable marginal distributions.

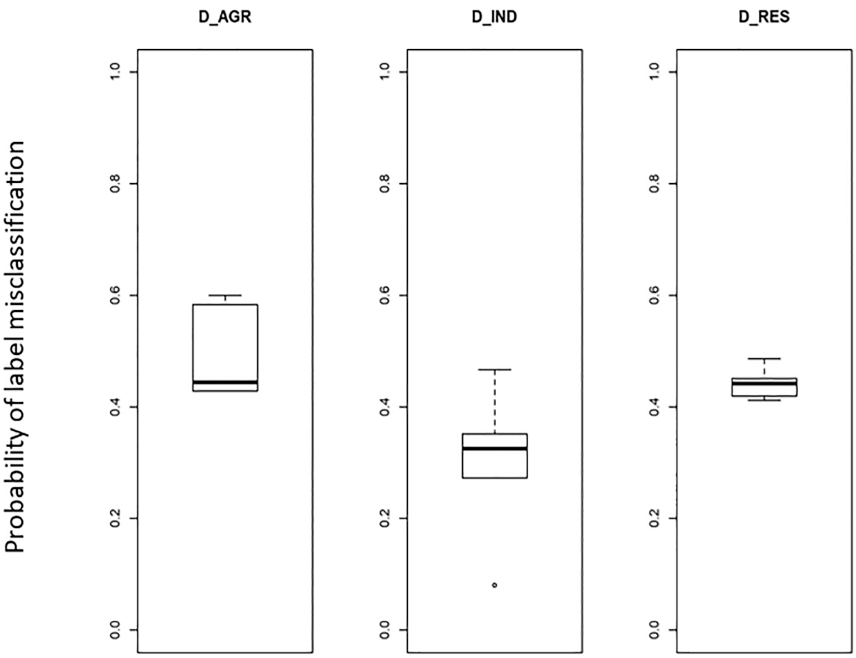

Fig. 5. Boxplots representing the probability of damages misclassification for the i) agricultural (D_AGR), ii) industrial (D_IND), and iii) residential (D_RES) sectors under the CONF_1C of the BN model.

Particularly, 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 (Lavalle, 2014). For **Scenario 2** (SC_FDE), the comparative changes in flood depth classes between the JRC LISFLOOD-FP model 10-year and 200-year return period flooding event projections were used to build the future scenario, representing the relative probabilistic increases in flooding depths across the case study area that could be seen under a more severe river flood event. The impacts of these evidences were then propagated through the BN to each damage-

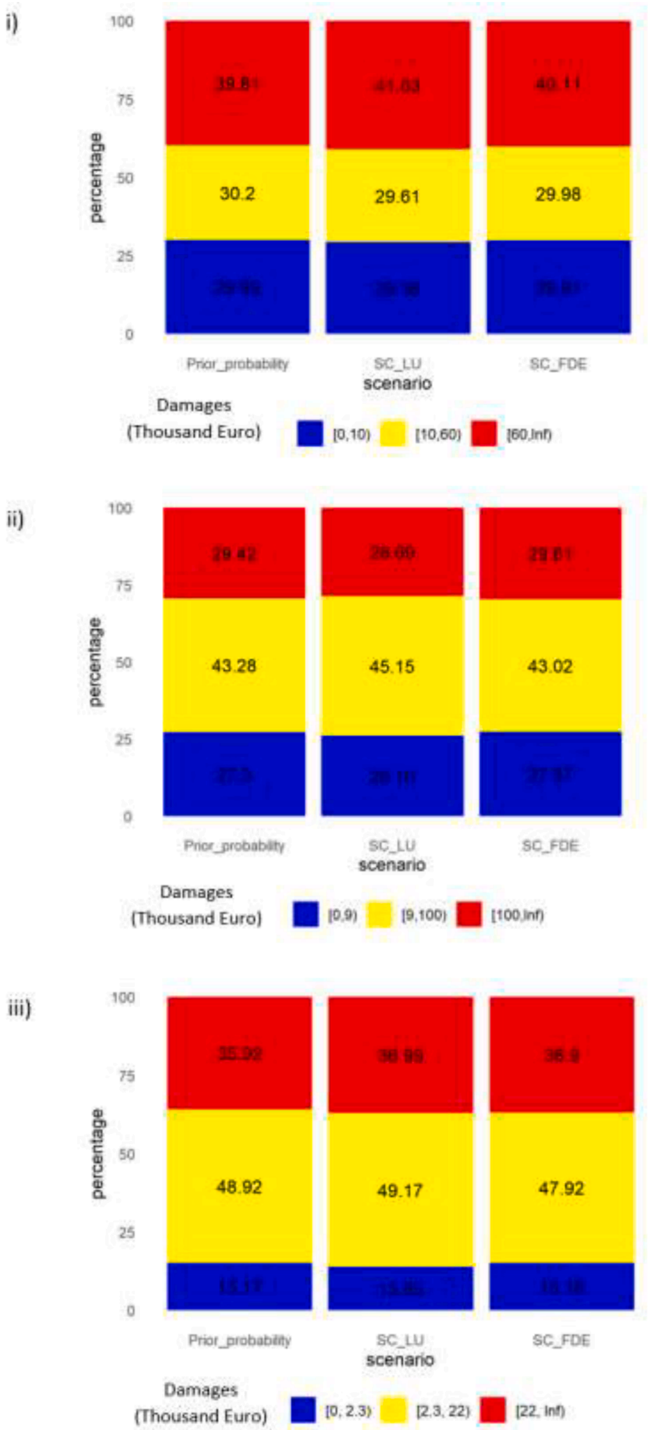

Fig. 6. Outputs of the BN model, comparing the prior probability of the BN model against the two simulated scenarios (SC_LU) and (SC_FDE), for the evaluation of potential damages in the i) agricultural, ii) industrial, and iii) residential sectors.

related assessment endpoint as a form of prognostic inference.

The resulting model outputs for the agricultural, industrial and residential sectors respectively are shown in Fig. 6. Specifically, the coloured boxes represent the frequency of labels classed under the three classifications of monetary damage for the different scenarios, with the value of damages increasing from blue to red.

Scenario 1: Understanding the change in damages under changing land use patterns (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 (over €22,000 and €6,000 respectively). For these two sectors in particular, these results indicate that there is an expected increase in damages 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 potential changes in total industrial area in the case study of concern. 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. 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: Understanding the change in damages at different flood-related return periods (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, with more damages in the highest range over €22,000, the changes are smaller for the residential and industrial damages.

These limited changes are again likely a result of the input dataset used for the BN training and scenario analysis, where the flood depth does not increase too significantly between the 10-year and 200-year return periods, used to represent changes in the probability of the magnitude of flood depth at each location. Specifically, under the 200-year return period (more severe flood event), there is an approximately 8% increase in the number of cases classified under the highest flood depth (i.e., over 1.21 m). 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 changes in hazard characteristics, and thus the resulting effects in damages.

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 3.5. This was performed individually for each of the explanatory nodes in the model, by setting a 100% probability of their 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 Fig. 7, with the relative probability distributions of damage 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.

The results indicate that the importance of each variable in terms of contributing to the potential damages varies by sector.

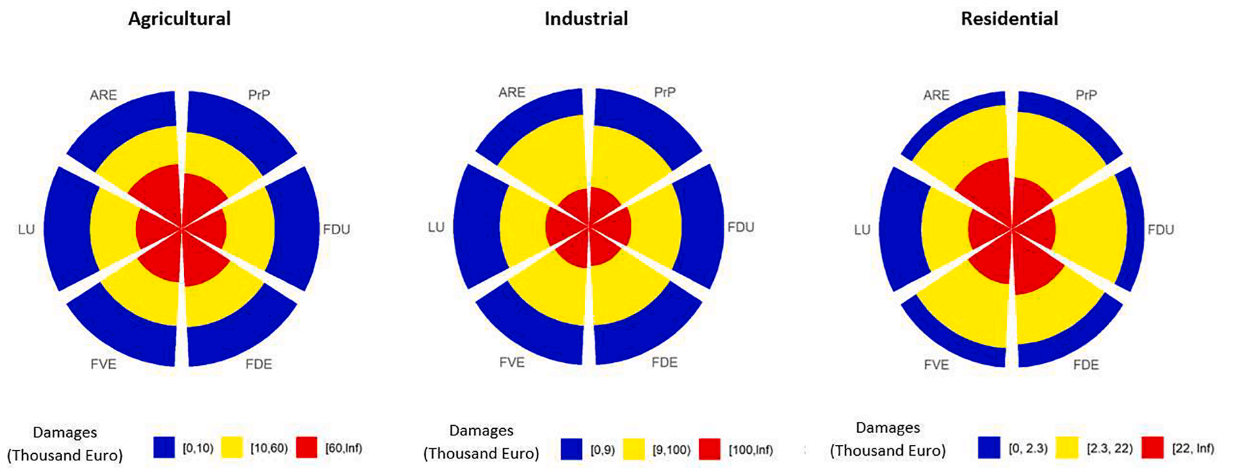

Fig. 7. Sensitivity analysis for the explanatory nodes of the constructed BN model for the agricultural, industrial, and residential sectors respectively.

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 explain the particular response of the industrial sector to the identified future scenarios, relative to the agricultural and residential sectors.

5. Conclusions

This work 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 and enhanced with hazard, vulnerability and exposure-related data for the case study area. 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 simulation of future damages under possible changes in hazard and exposure patterns. Specifically, the methodology as presented offers a more complete picture on multi-sectoral damages by including not only residential damage simulation as an assessment endpoint, but also for the industrial and agricultural sectors. Further, various approaches for the design and configuration of the BN model are deeply explored, alongside a sensitivity analysis, providing greater insight into how to optimize the design of BN models for multi-sectoral damage assessment. The final model, as constructed, showed promising capability for damage projection for all three sectors investigated within the case study and under two different scenarios. It also provides an analysis of two '*what-if*' scenarios for the examination of potential future damages 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, and this analysis also showed the potential capacity of the BN model to better understand possible future damages. Specifically, we observed higher expected monetary damages to the three sectors studied in comparison to the 2014 flooding event, under simulations of increased flooding depths that would result from a lower probability event. Similarly, damages were expected to increase for the agricultural and residential sectors under a projected scenario of changing land use patterns in the case study area. As such, it can be more clearly demonstrated how changing patterns in hazard, exposure, and vulnerability factors may likely influence damages that can be expected under future events.

Moreover, the stepwise model configuration provided insight on how to optimize 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 selected 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 capability for damage simulation, 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 was lower than might have been expected, as a result of limitations linked to the input training datasets. Further, the authors acknowledge the difference across the flood models-related data (i.e. flood depth data from the hydraulic model by Vacondio et al. 2016, and the LISFLOOD-FP model) used for the BN training and scenario analysis, and as a result the limitations presented. It was felt, however, that the relative changes in flood depth from the available data projections (LISFLOOD-FP data) represented the most realistic scenario in the absence of more suitable and compatible alternatives.

However, 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 increasingly frequent and severe flooding events, and show 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 spatialization of the output of the model, building on the GIS-based structure of the training dataset. A major advantage in the application of the BN approach lies in the possibility to combine heterogenous data from multiple sources and across different domains, which was vital in this study. In this case, while the training of the

BN model used GIS spatialization to extract the relevant data at each location, the output was purely probabilistic and could not be used to capture the extent of damages in specific locations. It may, for example, be possible to design the BN around aggregated zones rather than training the model on the entirety of the case study, and therefore allow for such a spatialized output that can capture more granular dynamics. Further approaches to overcome these limitations can also be explored in a more sophisticated framework, for example through the use of other ML methods, such as random forest, that may simplify the model outcomes' spatialization. While the model could provide some disaggregation in the output and validation, in terms of the multiple sectors analyzed, a more sophisticated model, either expanding on the BN approach or through other ML methods, that could perform a spatially explicit application would present an important improvement of the approach. Further, a spatial model could also be applied with the aim of capturing 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 the analysis to potential damages that have not yet been observed and recorded.

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 simulation 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.

Specifically, we envision that this application will support decision makers in the fields of disaster risk reduction and climate change adaptation in extracting increased knowledge from flooding disaster and loss data, capitalizing on the full potential of increased volumes and variety of available data.

These results can support the definition of risk reduction measures and preparedness and emergency plans, particularly in the face of unexpected risk scenarios which can be difficult to evaluate and assess. We believe that the presentation of these results could also be used to promote and improve risk awareness, through the demonstration of the possible increased future flooding damages for multiple sectors. This multi-sectoral approach will also give to decision makers a more complete picture of these damages, while also providing a unique tool to demonstrate the range of impacts to various sectoral stakeholders.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The research leading to these results has been funded by the LODE project (*Loss Data Enhancement for DRR & CCA management*, https://www.lodeproject.polimi.it/) funded by the European Union Humanitarian Aid and Civil Protection (DG ECHO, contract number: 825567). The authors gratefully acknowledge their colleagues, Dr. Mattia Amadio for the contribution on multi-sectoral damage data for the case study area, and Dr. Arthur H. Essenfelder for his valuable advice during the river flood-related scenario analysis.

References

Amadio, M., Mysiak, J., Carrera, L., Koks, E., 2016. Improving flood damage assessment models in Italy. Nat. Hazards 82 (3), 2075–2088. 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.

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 (6), 1323–1337. https://doi.org/10.5194/nhess-16-1323-2016.

Burek, P., Van Der Knijff, J., De Roo, A., 2013. LISFLOOD – Distributed Water Balance and Flood Simulation Model – Revised User Manual 2013. Publications Office of the European Union. https://publications.jrc.ec.europa.eu/repository/handle/JRC78917.

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. Eng. Geol. 66 (3–4), 241–255. https://doi.org/10.1016/S0013-7952(02)00031-5.

Coppola, E., Verdecchia, M., Giorgi, F., Colaiuda, V., Tomassetti, B., Lombardi, A., 2014. Changing hydrological conditions in the Po basin under global warming. Sci. Total Environ. 493, 1183–1196. https://doi.org/10.1016/j.scitotenv.2014.03.003.

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. Prog. Disaster Sci. 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. Agric. Water Manag. 77 (1–3), 21–36. https://doi.org/10.1016/j.agwat.2004.09.029.

Directive 2007/60/EC of the European Parliament and of the Council of 23 October 2007 on the assessment and management of flood risks. https://eur-lex.europa.eu/eli/dir/2007/60/oj.

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.

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. Sci. Total Environ. 703, 134972. https://doi.org/10.1016/j.scitotenv.2019.134972.

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. Nat. Hazards Earth Syst. Sci. 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 popolazione 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. Nat. Hazards Earth Syst. Sci. 12 (12), 3733–3752. https://doi.org/10.5194/nhess-12-3733-2012.

Joyce, J. (2021). Bayes' Theorem. In E. N. Zalta (Ed.), The Stanford Encyclopedia of Philosophy (Fall 2021 Edition). https://plato.stanford.edu/archives/fall2021/entries/bayes-theorem/%3E).

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/.

Landuyt, D., Broekx, S., D'hondt, R., Engelen, G., Aertsens, J., Goethals, P.L.M., 2013. A review of Bayesian belief networks in ecosystem service modelling. Environ. Model. Softw. 46, 1–11. https://doi.org/10.1016/j.envsoft.2013.03.011.

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.

Li, Y., Wu, F.-X., Ngom, A., 2016. A review on machine learning principles for multi-view biological data integration. Briefings Bioinf. 19 (2), bbw113. https://doi.org/10.1093/bib/bbw113.

Lombardi, A., Tomassetti, B., Colaiuda, V., Verdecchia, M., 2018. Increasing of severe hydrological events in the Po basin under global warming. Geophys. Res. Abstracts 20. https://ui.adsabs.harvard.edu/abs/2018EGUGA.20.6354L/abstract.

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.

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.

Montanari, A., 2012. Hydrology of the Po River: Looking for changing patterns in river discharge. Hydrol. Earth Syst. Sci. 16 (10), 3739–3747. https://doi.org/10.5194/hess-16-3739-2012.

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 Trans. Built Environ. 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 Italy. Water Resour. Res. 51 (10), 7995–8011. https://doi.org/10.1002/2015WR017426.

Paprotny, D., Kreibich, H., Morales-Nápoles, O., Terefenko, P., Schröter, K., 2020a. Estimating exposure of residential assets to natural hazards in Europe using open data. Nat. Hazards Earth Syst. Sci. 20 (1), 323–343. https://doi.org/10.5194/nhess-20-323-2020.

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.

Peña, J.M., Björkegren, J., Tegnér, J., 2005. Learning dynamic Bayesian network models via cross-validation. Pattern Recognit. Lett. 26 (14), 2295–2308. https://doi.org/10.1016/j.patrec.2005.04.005.

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. Coast. Eng. 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. Environ. Modell. Software 22 (8), 1140–1152. https://doi.org/10.1016/j.envsoft.2006.03.006.

Rodriguez, J.D., Perez, A., Lozano, J.A., 2010. Sensitivity analysis of k-Fold cross validation in prediction error estimation. IEEE Trans. Pattern Anal. Mach. Intell. 32 (3), 569–575. https://doi.org/10.1109/TPAMI.2009.187.

Sairam, Nivedita, Schröter, Kai, Carisi, Francesca, Wagenaaar, Dennis, Domeneghetti, Alessio, Molinari, Daniela, Brill, Fabio, Priest, Sally, Viavattene, Christophe, Merz, Bruno, Kreibich, Heidi, 2020. Bayesian Data-Driven approach enhances synthetic flood loss models. Environ. Model. Softw. 132 https://doi.org/10.1016/J.ENVSOFT.2020.104798.

Sairam, N., Schröter, K., Rözer, V., Merz, B., Kreibich, H., 2019. Hierarchical Bayesian approach for modeling spatiotemporal variability in flood damage processes. Water Resour. Res. 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. *Proc. Inst. Civ. Eng.: Civ. Eng. 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 Resour. Res. 50 (4), 3378–3395. https://doi.org/10.1002/2013WR014396.

Scutari, M., 2010. Learning Bayesian networks with the bnlearn R package. J. Stat. Softw. 35 (3), 1–22.

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

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. *J. Environ. Manage.* 202, 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. Environ. Sci. 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. J. Risk Res. 23 (11), 1424–1439. 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).

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. Environ. Modell. Software 57, 60–75. https://doi.org/10.1016/j.envsoft.2014.02.003.

Vacondio, R., 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. Nat. 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. Nat. Hazards Earth Syst. Sci. Discuss. 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 (9), 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 Resour. Res. 54 (5), 3688–3703. https://doi.org/10.1029/2017WR022233.

Ward, P.J., De Moel, H., Aerts, J., 2011. How are flood risk estimates affected by the choice of return-periods? Nat. Hazards Earth Syst. Sci. 11 (12), 3181–3195. https://doi.org/10.5194/NHESS-11-3181-2011.

Wing, O.E.J., Pinter, N., Bates, P.D., Kousky, C., 2020. New insights into US flood vulnerability revealed from flood insurance big data. Nat. Commun. 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. Clim. Change 89 (3–4), 411–433. https://doi.org/10.1007/s10584-008-9395-z.