Acta Polytechnica CTU Proceedings **36**:33–46, 2022
https://doi.org/10.14311/APP.2022.36.0033

FLOOD RISK ASSESSMENT FOR ROAD INFRASTRUCTURES USING BAYESIAN NETWORKS: CASE STUDY OF SANTAREM - PORTUGAL

Erica Arango[a,*], Monica Santamaria[a], Maria Nogal[b], Hélder S. Sousa[a], José C. Matos[a]

[a] *University of Minho, Institute of Science and Innovation for Bio-Sustainability, Department of Civil Engineering, Guimarães, 4710-057 Braga, Portugal*

[b] *Delft University of Technology, Department of Materials, Mechanics, Management & Design., Gebouw 23, Stevinweg 1, 2628 CN Delft Netherlands*

[*] corresponding author: ericaaarango@gmail.com

Abstract. Assessing flood risks on road infrastructures is critical for the definition of mitigation strategies and adaptation processes. Some efforts have been made to conduct a regional flood risk assessment to support the decision-making process of exposed areas. However, these approaches focus on the physical damage of civil infrastructures without considering indirect impacts resulting from social aspects or traffic delays due to the functionality loss of transportation infrastructures. Moreover, existing methodologies do not include a proper assessment of the uncertainties involved in the risk quantification. This work aims to provide a consistent quantitative flood risk estimation and influence factor modelling for road infrastructures. To this end, a Flood Risk Factor (FRF) is computed as a function of hazard, vulnerability, and infrastructure importance factors. A Bayesian Network (BN) is constructed for considering the interdependencies among the selected input factors, as well as accounting for the uncertainties involved in the modelling process. The proposed approach allows weighting the relevant factors differently to compute the FRF and improves the understanding of the causal relations between them. The suggested method is applied to a case study located in the region of Santarem Portugal, allowing the identification of the sub-basins where the road network has the highest risks and illustrating the potential of Bayesian inference techniques for updating the model when new information becomes available.

Keywords: Bayesian networks, decision-making, flood risk assessment, road networks.

1. Introduction

The efficient and rational management of natural hazards is a concern that has become of increasing importance for decision-makers, generating the need for a strategy that explicitly addresses the uncertainties and consequences involved in the occurrence of these extreme events. A concurrent event of importance for management is the flooding risk of road networks, which is related to a variety of natural, economic, and social factors such as climate change, use and coverage soil, deficient drainage systems, population explosion, among others [1]. Therefore, exploring the relationship between these factors and how they directly or indirectly influence the flood risk of road networks is of great relevance to this study. However, each factor constitutes a complex system, and they also interact with each other. On the other hand, the flood events occurrence leads to social insecurity, undermining the economic development of the impacted regions and affecting ecosystems [2]. Not to mention that its frequency is expected to increase more in the future [3]. Due to that, global economic risks are also increasing, along with the need for proper potential flood risk assessment and management tools to avoid or manage the level of disaster and minimize the potential damage.

Existing research on risk assessment can be classified into two main categories. On the one hand, qualitative assessment methods. For example, the analytical hierarchy process (AHP) or multi-criteria method [4–7] allows integrating multiple factors (e.g., hazard, vulnerability, and exposure) for flood risk assessment. On the other hand, the quantitative evaluation methods, normally based on historical data [8] and scenarios simulation based on the hydrological-hydraulic model [9] or hydrodynamic [10], as well as flood risk simulation models, such as random forest, decision trees, genetic algorithms, among others [11–14]. Qualitative methods are normally criticized for their subjective nature, which results from excessive dependence on empirical knowledge. Therefore, they are considered high uncertainty and low credibility models in risk assessment. On the other hand, quantitative evaluation methods require a large amount of data with high precision, which limits their application in some cases due to the reduced availability of data [15]. Additionally, none of the aforementioned methods define the mutual relationships between the

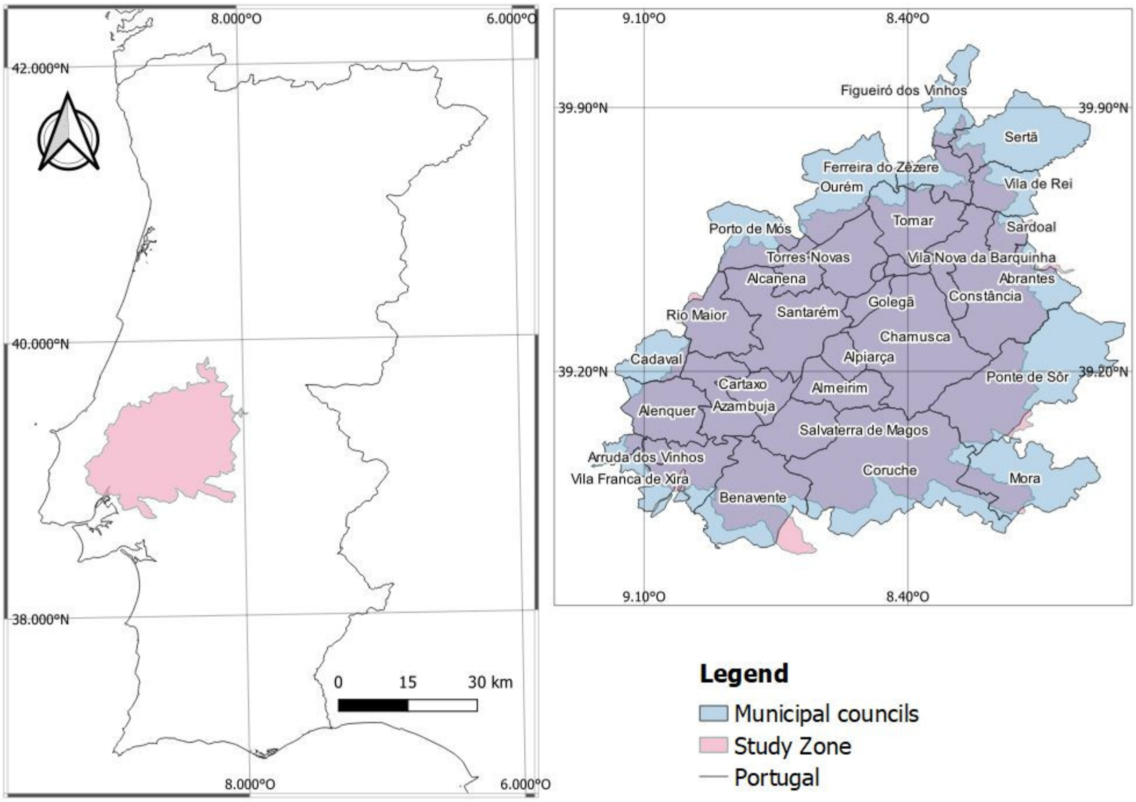

FIGURE 1. Geographical location and specification of the study area.

factors that influence risk, which is often interdependent in a complex and uncertain way [1]. Therefore, it is evident the need for simplifications to establish a model that adequately evaluates the flooding risk in the road network, but that effectively considers the relationship between the different risk factors without requiring much data availability.

A Bayesian network (BN) is considered a simplified model that allows uncertainty estimation during risk assessment, contemplating the probabilities of the random variables and their joint probability distribution [16]. One of its advantages is that it enables the identification of relevant factors and their interrelationships with flood risk using previous knowledge, which can be inferred and reasoned from real data [17]. BNs have been implemented in urban flood risk assessment [1], river basins [18], coastal hazards [19], early warning of floods [20], among others. On the other hand, the geographic information system (GIS) that has powerful spatial data processing techniques allows greater data accessibility. Therefore, there are studies in which the GIS and BN models have been integrated to estimate the probability of hazards, for instance [21–23]. However, there is not enough research on the chains of disaster-causing factors based on the BN model to explore the path through which the factors affect flooding in the road network.

This work aims to provide a consistent quantitative flood risk estimation and influence factor modelling for road infrastructures. Therefore, a flood risk factor (FRF) is computed as a function of hazard, vulnerability, and infrastructure importance factors, using a relative flood risk classification method as a practical and measurable process in terms of inputs and application. Moreover, a Bayesian network (BN) is constructed to consider the interdependencies between the selected input factors, as well as to consider the uncertainties involved in the modelling process. The proposed approach allows the relevant factors to be weighted differently to calculate the FRF and improves the understanding of the causal relationships between them, as shown in the case study application, located in the region of Santarem, Portugal. In this case study, it was possible to identify and quantify the sub-basins where the road network is most at risk and illustrates the potential of Bayesian inference techniques to update the model when new information becomes available.

2. METHODOLOGY

The proposed approach in this study consists of four main steps: (1) identification of the main factors that may affect or influence (directly or indirectly) flood disaster occurrence and the data collection from the different data sources associated with the factors previously identified; (2) assessing flood risk of road infrastructures (FRF index) and (3) development of the BN structural graph and calculation of probability distribution tables based on the BN structure, and (4) sensitivity analysis.

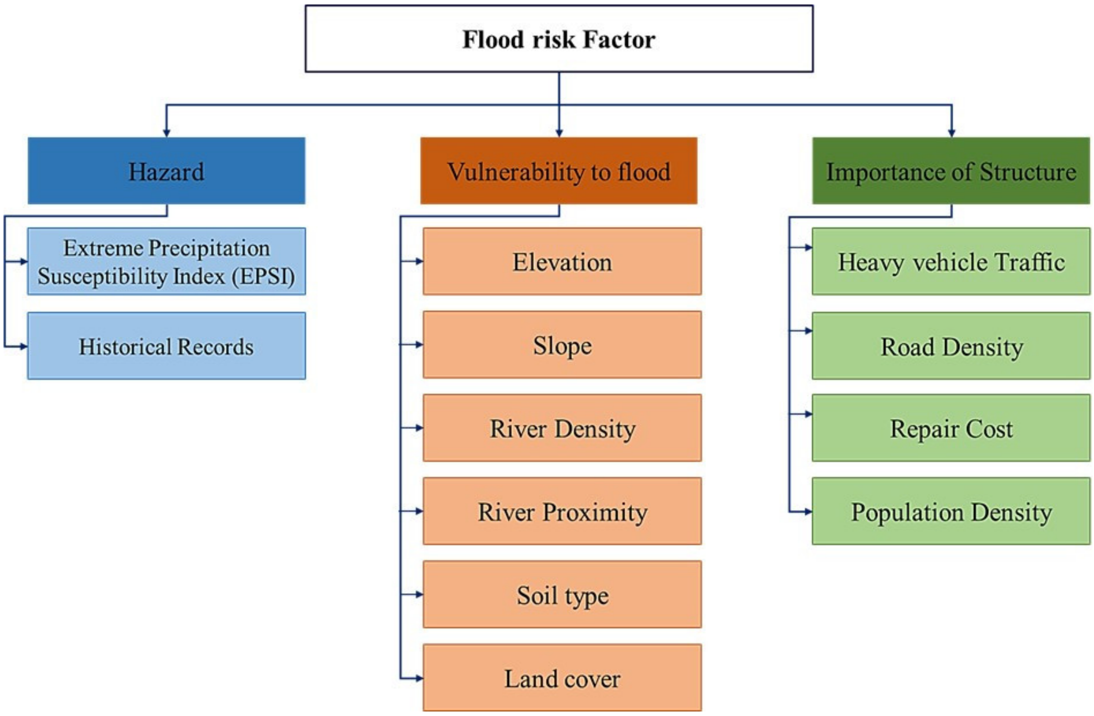

FIGURE 2. Factors affecting the risk of floods on road infrastructures.

The case study is located in Portugal, between 7°57' and 9°9' eastern longitudes and between 38°42' and 39°54' northern latitude, which comprises an area of approximately 7044.4 km^2. The system boundaries that were specified for the implementation of the framework consist of 32 municipal councils from the Santarem district, as shown in Figure 1.

The case study covers a section of one of the most important rivers in Portugal, the Tagus River; therefore, the closest sub-basins are low elevation and tend to be a flood-prone area. These lowlands are characterized by an average altitude of 103.5 m and an average slope of 8.4%. The region has an average population of 40.842 inhabitants, with densely populated regions such as Santarem municipality. The Tagus hydrographic basin presents significant differences in annual precipitation through its spatial distribution, with low rainfall in the upper Tagus and higher rainfall in the middle. Thus, in this latter region, e.g., in the Sorraia sub-basin, flood events occur more frequently. According to the historical flood records dated from 1865 to 2015, the sub-basin with the highest number of flood records are "Rio Tejo (HMWB - Jusante Bs. Castelo do Bode e Belver)" and "Tejo (HMWB - Jusante B. Belver)" (Figure 3), with 21 records and more than 2000 people affected.

2.1. Factors affecting the flood risk of road infrastructures

For flood risk assessment of road infrastructures, it is important to consider the main causing factors, such as the likelihood of the flood occurrence, the structure vulnerability to flooding, and the infrastructure importance. Additionally, each factor is constituted by a series of sub-factors, which directly or indirectly influence flood risks on road infrastructures. For instance, among the causes of flood occurrences, the magnitude of rainfall, the number of storms days, or the rain intensity is usually considered. Moreover, the environment nature where the flooding occurs is focused on the combination of climatic variables and exposure surfaces (i.e., factors such as land cover, topography, rivers distribution in regions). Therefore, elevation, slope, river network, and road density should be considered as some of the most important factors for flood risk assessment.

On the other hand, to consider the damage caused by floods in social terms, factors such as population density are required because the higher the population density, the more serious economic and social damage. Road density is also an important factor, due to regions with greater transport having greater adaptability to disasters. In addition to the factors that reflect the exposure characteristics (buildings, people, or others), past flood records including information on damages and losses is a relevant factor to take into consideration.

The evaluation factors considered in this study are related to the danger of flooding by sub-basin and it is important to clarify that the factors are decided based on the results of previous studies such as [1, 15, 24, 25] and expert judgment. Summarizing, the parameters and sub-parameters that affect the flooding risk of road networks and that are considered in this study are summarized in Figure 2.

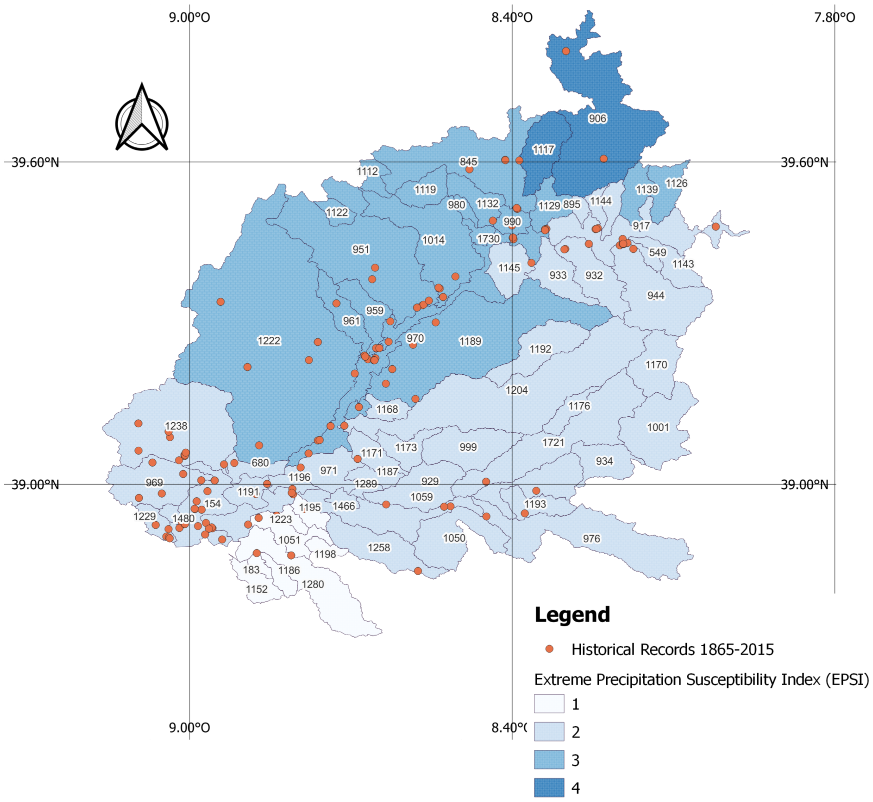

FIGURE 3. Data collection for the likelihood of flood occurrence: EPSI and historical records. Large values of EPSI indicate larger susceptibility.

2.1.1. FACTORS AFFECTING THE LIKELIHOOD OF THE FLOOD RISK OCCURRENCE

The data collection and processing are discussed below.

Extreme precipitation susceptibility index (EPSI)
EPSI is regarded in this study as a principal factor for considering flood hazards. This index which ranges from 1 to 4 was proposed by Santos, Fragoso, and Santos [26], and considers the following extreme precipitation indices:

- $R \times 1$day and $R \times 5$days: Annual highest daily precipitation and annual highest 5 consecutive precipitation days, respectively. Describe the precipitation associated with periods of extreme rain;
- SDII: Annual total precipitation divided by the number of wet days (precipitation ≥ 1 mm) in the year, which reflects daily intensity;
- R20: Number of heavy precipitation days ≥ 20 mm, i.e., moderate to high amounts of daily precipitation;
- CWD: Number of consecutive wet days (daily precipitation ≥ 1 mm), and
- R95PTOT: Fraction of annual total precipitation exceeding the 95th percentile. It is used to evaluate the contribution of extreme daily precipitation to the total precipitation.

EPSI considers a return period of 10 years, (i.e., it has a 10% probability of occurring within any year) which makes it a suitable and balanced indicator of extreme precipitation. Its calculation involves GIS technology justified by the strong seasonality of the precipitation regime, and it was developed for both annual data and the meteorological season considering records between 1950 and 2003. EPSI was calculated from four main stages: (1) selection of the most relevant indices chosen from a sensitivity analysis using different combinations; (2) normalization process for all variables; (3) EPSI calculation using the previously normalized indices, and (4) definition of four classes of susceptibility, based on a classification by quantiles. For more details about EPSI see Santos, Fragoso, and Santos [26]. EPSI indices for each sub-basin are shown in Figure 3. It should be noted that sub-basins are identified in correspondence with the codes used by the Portuguese Environmental Agency [27].

Historical Flood records In addition to EPSI, it is also important to consider the areas where floods occurred in the past, giving greater relevance to basins that have historical records and therefore greater probability of flood risk. The data was obtained from the national database of hydro-geomorphologic-related disasters (landslides and floods) in Portugal, built and maintained since 2009 as part of the DISASTER project [28].

The data of the factors affecting the likelihood of flood risk occurrence are shown in Figure 3.

2.1.2. Vulnerability Factors

Topographic characteristics

Flood occurrences and redistribution are heavily influenced by topography aspects such as elevation and slope. The elevation is characterized by the vertical distance between a given surface and the reference basement, and the slope is an average rate of elevation change in a particular domain. The elevation and slope data were calculated using a Digital Elevation Model (DEM), derived from the Shuttle Radar Topographic Mission (SRTM) in a resolution of 3 arc-seconds, processed for Portugal [29]. Block statistics and grid algebra calculations were used to divide both components, elevation, and slope, into four levels which were then reclassified (as shown in Appendix A - Figure **??**a and Figure **??**b, respectively).

River network

The distribution of the river network has an important role in the flood risk of road networks, especially the river density and the distance to the river. In this work, the study area is mainly focused on the Tagus River, which has its origin in the Sierra de Albarracín, Spain, with a length of 1,100 km, of which 230 km are in Portugal. In the national territory, its main tributaries are the Zezere river and the Sorraia river. Information regarding the river network was collected from the Portuguese Environmental Agency [30]. River density was calculated as the length of rivers per unit area, using Line Density function, and river proximity as the closest distance to a river, which was obtained using the Multiple Buffer operator on rivers of order 3 onwards, as shown in Appendix A - Figure **??** (a, b) respectively.

Land cover and toil type

Environmental factors influencing flood risks are considered in terms of land cover and soil type. The land cover information was obtained from the global map of land use/land cover (LULC), acquired from ESA Sentinel-2 imagery at a 10-meter resolution [31], available in [32] (Figure 11). From this map, the impervious area percentage was calculated, assuming three categories of importance: buildings and paved surfaces as areas with more flooding risk; tree/wooded areas as medium flooding risk; and other surfaces (such as Water (W), grass (G), Flooded vegetation (FV), Crops, (C), Scrub(S) and Bare ground (BG)) as low-risk flood areas. Regarding soil type, the information was obtained from the Topsoil physical properties for Europe, based on the LUCAS topsoil database [33]. The data were classified according to NRCS Soils Triangle and discretized using the hydrologic soil group (HSG) classification according to soil properties proposed by Zeng et al. [34], as shown in Appendix A - Figure **??**.

2.1.3. Infrastructure importance factors

When a high-traffic-density road collapses or becomes blocked, it can affect the whole traffic system due to the use of alternate routes, resulting in social and economic losses. The higher the traffic density, the greater the road importance and the higher the repair cost due to lack of availability. A high road density usually allows alternative routes in case of disaster. On the other hand, the higher the population density, the greater the damage caused by floods. Therefore, in this study, the road traffic density, expected repair cost, road density, and population density are considered sub-parameters for measuring the infrastructure's importance in social and economic terms. The traffic demand and road data of the case study were obtained from *Infraestruturas de Portugal* (IP), the Portuguese transportation infrastructures manager, and the population density data were collected from the National Statistical Institute [35]. Daily traffic density was divided into light and heavy traffic, and road density was calculated as the length of roads per unit area, using the Line Density function. Regarding the repair cost, the average construction costs per kilometer of the European Court of Auditors were taken into account, which depends on the road type and the country, referred to in [36]. These factors are presented in Appendix A - Figure **??**.

The following step consists in discretizing all the considered sub-factors due to each node of the BN model being assigned to a finite set of state values based on discrete probabilities. According to current studies [1, 15], and expert knowledge, the sub-factors impacting flood risk were categorized into three groups, as shown in Table 1.

3. Determining the flood risk of road infrastructure

Expert knowledge and the approach implemented by Kim et al. [37] were used to assign weights according to each factor's importance in the flood risk occurrence. In this case, each evaluation sub-factor was assigned a weight to represent the degree of related flood risk. The weights are assigned using a scale of 1 to 3 through intuitive judgment, where a weighting factor of 1 represented the lowest risk and a weighting value of 3 the highest flood risk. The final risk of each category (hazard, vulnerability, and infrastructure importance) was obtained by adding the weighted average of the total of sub-factors that compose it, also assuming a scale of 1 to 3. Finally, the Flood Risk Factor (FRF) values are assumed as follows: a final risk value less than 2 is a low-risk degree, high risk is considered from a final value greater than 2.5, and medium risk between 2 to 2.5.

Categ.	Factors	Classes		
		Low	Medium	High
Hazard	EPSI	1	2	> 2
	Historical records	0	0 - 3	> 3
Vulnerability	Elevation (m)	> 150	76 - 150	≤ 75
	Slope (%)	> 10	5 - 10	≤ 5
	River density (km/km^2)	≤ 0.15	0.15 - 0.25	> 0.25
	River Proximity (m)	> 1500	500 - 1500	≤ 500
	Soil type[a]	Co, S, LS or SL	L, SiL, Si, SCL	CL, SiCL, SC, SiC or C
	Land Cover	Build area	Trees	Other surfaces
Structure Importance	Light Traffic (vehicles)	$\leq$ 40.000	40.000 - 150.000	$>$ 150.000
	Heavy Traffic (vehicles)	$\geq$ 5.000	5.000 - 20.000	$\geq$ 20.000
	Road density (km/km^2)	≤ 1	0.5 - 1.0	≤ 0.5
	Repair cost (€)	$\leq$ €500 M	€500 M - €1.500 M	$\geq$ €1.500 M
	Population density	$\leq$ 20.000	20.000 - 35.000	$>$ 35.000

[a] Co (coarse), LS (Loamy sand) SL (Sandy loam); L (Loam) SiL (Silt loam) Si (Silt) SCL (Sandy Clay Loam); CL (Clay Loam) SiCL (Silty clay loam) SC (Sandy Slay) SiC (Silty Clay) or C (Clay)

TABLE 1. Classification of different risk factors.

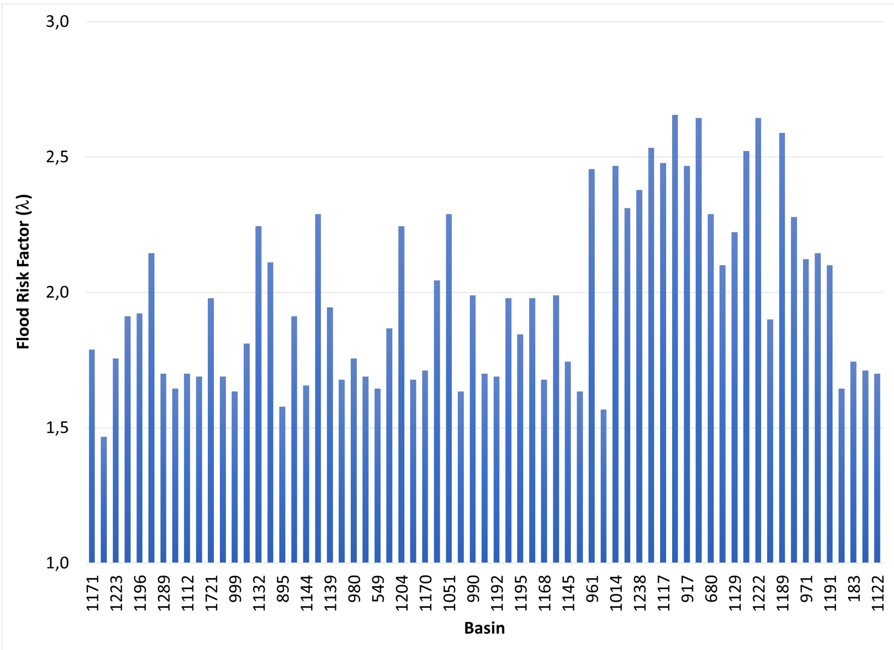

FIGURE 4. FRF results by sub-basin.

3.1. Flood Risk calculation

The average weights for the three main categories (λ_i), namely the likelihood of flood occurrence, vulnerability, and infrastructure importance, were calculated using weights for their sub-factors. Essentially, λ_i can be calculated as Equation 1.

$$\lambda_i = \frac{\sum_{k=1}^{N_k} F_k}{N_k}, \quad k = 1, 2, \ldots N_k \tag{1}$$

where, F_k represents the corresponding sub-factor for each category; N_k represents the number of sub-factors for each category, and the subscript k is the index of each sub-factor. The final FRF is obtained by calculating the average weights of the categories as Equation 2.

$$FRF = \frac{\sum_{i=1}^{N_i} \eta_i \, \lambda_k}{\eta_i}, \quad i = 1, 2, \ldots N_i \tag{2}$$

Where, N_i represents the number of categories and, η_i are the weights of each category. In this case, the weights were set equivalently to 1, assuming that all have equal importance for the final probability of flood risk.

The weights were applied using the classification

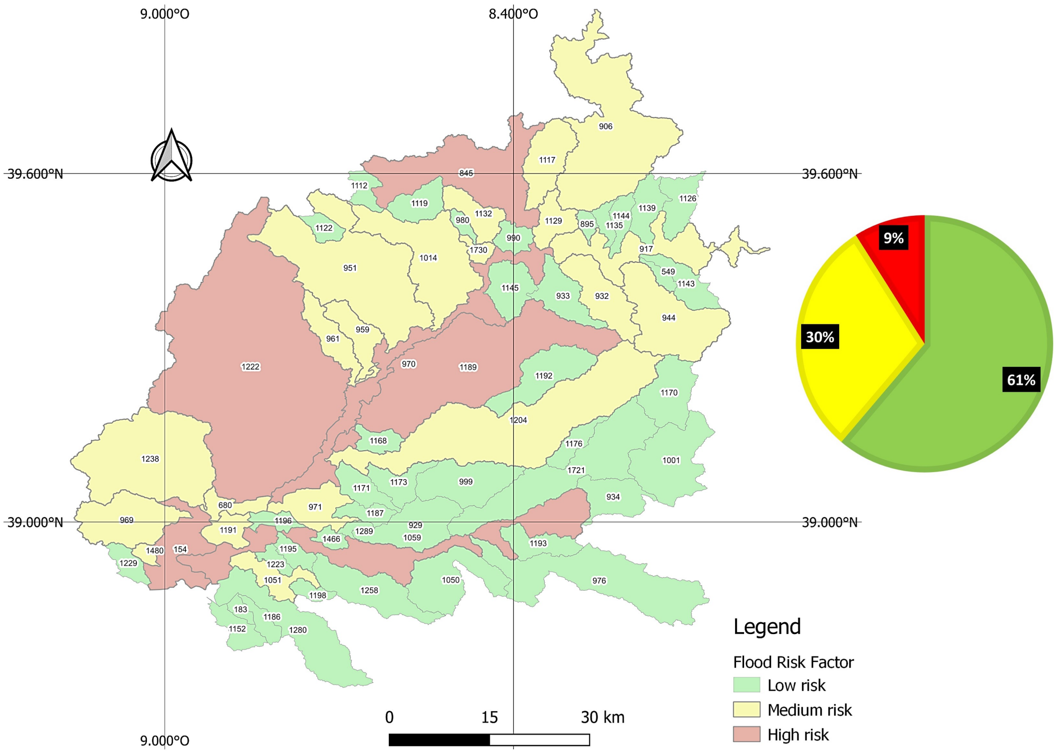

Figure 5. Distribution of the flood risk in the case study.

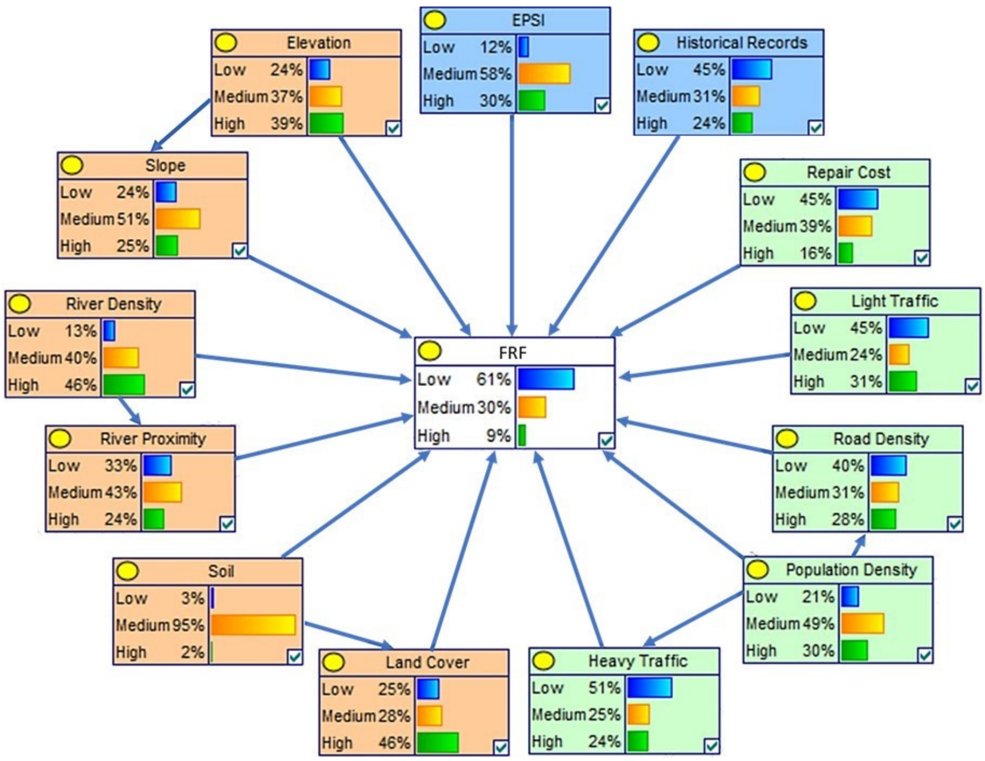

Figure 6. BN model for flood risk assessment of road infrastructures y.

proposed in section 2.2.3, specifically Table 1. For example, a weighting factor 3 was assigned if the EPSI exceeded 2, i.e., it was classified as high; a weighting factor 1 was assigned if the EPSI value was less than or equal to 1, i.e., classified as low; and a weighting factor of 2 was assigned if the EPSI was between 1 and 2, i.e., classified as medium.

Flood risk assessment was conducted on 67 sub-basins in the Center and South of Portugal. Figure 4 illustrates the results of the evaluation, and Figure 5

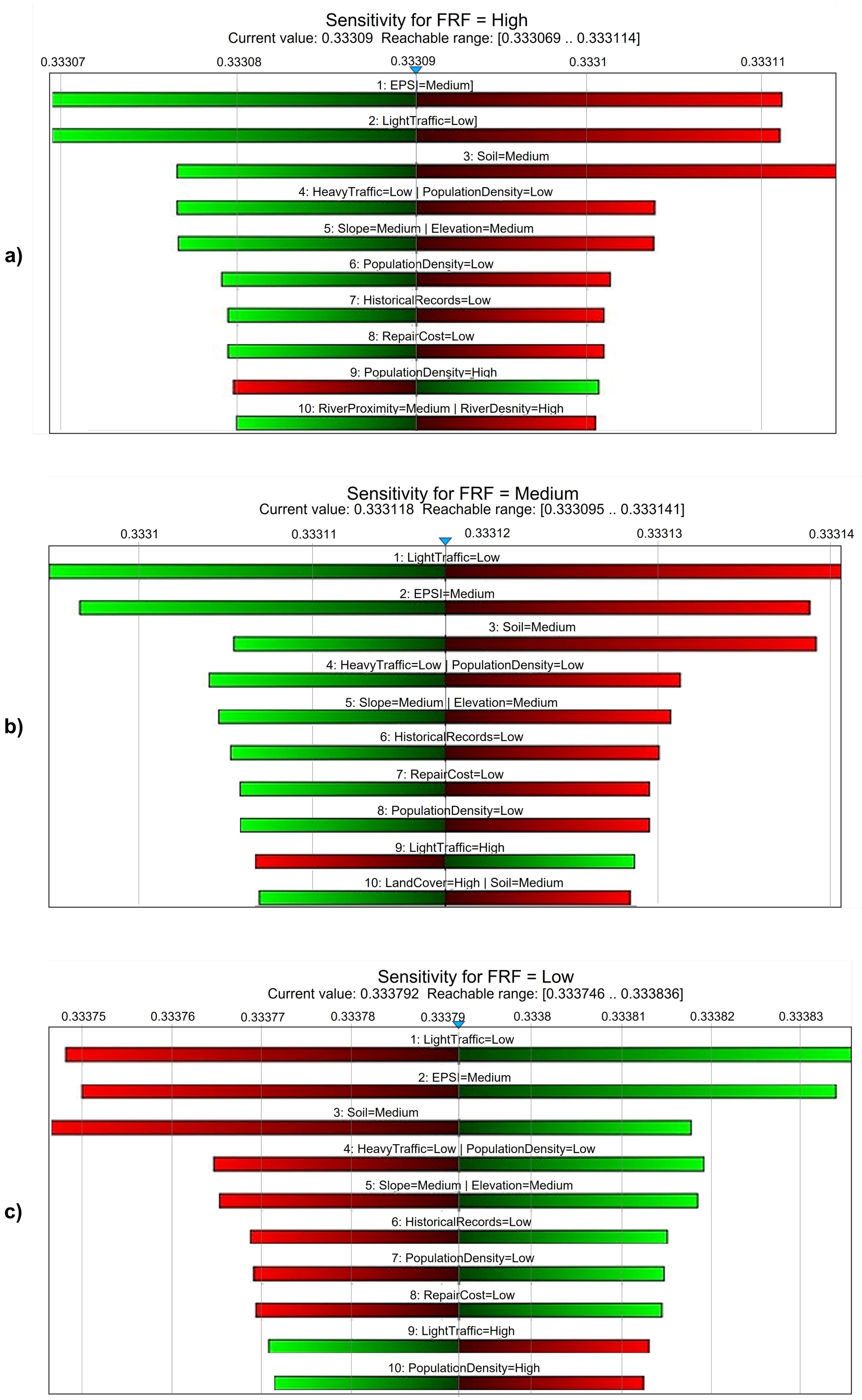

FIGURE 7. FRF Sensitivity Analysis: a) High, b) Medium and c) Low.

illustrates the distribution of the results. From the assessment results, it can be observed that 9% of the sub-basins received a high-risk rating, which suggests that the road network in these regions has a greater probability of suffering large impacts due to flooding events. On the other hand, 61% of the sub-basins (41) received a low-risk rating, while approximately 30% of the total study area is likely to have medium-risk flood events.

4. BN model

The structure of a Bayesian network consists of two components. On the one hand, the Directed Acyclic Graph (DAG) where the random variables are denoted as nodes of eigenvectors and the arc represents the probabilistic dependency between nodes. The other component is Conditional Probability Tables (CPTs) used to specify dependency relationships encoded in a DAG [22], [38]. The BN model construction for the flood risk evaluation of the road network is based on the defined hazardous factors in Section 2.2, which are used as nodes in the BN structure graph. The network construction procedure consisted of three main steps: (1) review of the structure and potential relationships between the different factors (nodes), (2) parameter learning to obtain conditional probability table (CPT), and (3) sensitivity analysis.

For the first step, the potential relationships between different factors were first determined based on the scientific literature and expert knowledge and were subsequently verified using learning algorithms such as Bayesian estimation and maximum likelihood estimation (MLE). Regarding the parameter learning, the training was made from the sample data collected from the factors in the study area, once the respective discretization was used to determine the conditional probability distribution for each sub-factor. For this methodology, the GeNIe software [39] was utilized, which allows importing data directly from CSV or txt format. Additionally, GeNIe software has a great variety of useful algorithms in the learning structures and parameters process, as well as facilitating sensitivity analysis in simple graphs to calculate their impact on the results. For this study PC and Bayesian Search structure learning algorithm was implemented .

The developed BN model is presented in Figure 6, which consists of all nodes (sub-factors) with their corresponding states.

The categorical classification employed in Section 2.2 can be also visualized, i.e., flood risk is represented with the white node, hazard factors with blue nodes, vulnerability factors with orange nodes, and factors to consider the infrastructure importance with green nodes. The nodes states (discretization) are established according to the classes shown in Table 1 (e.g., the "Elevation" node has different altitudes, such as low, medium, or high) and each square represents the state probabilities in terms of percentage. Figure 6 also shows that the 13 sub-factors considered affect directly or indirectly the occurrence of flood risk in the road network. The nodes connected with arrows represent a potential relationship between sub-factors, in which the initial node (arrow beginning) is the cause, while the pointed node (arrow end) is the effect. Each potential relationship is expressed through conditional probabilities calculated from the data collected.

The developed BN model can objectively evaluate links among diverse influencing factors and statistically infer the likelihood of flood risk based on information obtained from QGIS, presenting a novel method for assessing flood risk for road networks.

4.1. Sensitivity Analysis

The sensitivity in the flood risk assessment was evaluated to identify the influence of sub-factors on the obtained results. For this, another tools available in GeNle was used, which performs a simple sensitivity analysis of Bayesian networks using the algorithm proposed by Kjaerulff and van der Gaag [40]. Which efficiently calculates a complete set of derivatives of the posterior probability distributions on the target nodes on each of the numerical parameters of the Bayesian network from a set of target nodes. It means, the BN model learning and recording the relative importance of the input variables to predict the output.

The sensitivity value of the 10 most dominant sub-factors influencing each of the three states of the target node "Flood Risk" are shown in the tornado graph in Figure 7 a, b, c, for high, medium, and low levels, respectively. It is worth mentioning that the bar color indicates the change direction in the target node; red indicates a negative change, while green indicates a positive change. For instance, according to the sensitivity values of different factors, the greatest impact on the high FRF is the EPSI factor (related to the causes of the hazard) with a state "Medium" and the Light Traffic (Low), relative to the importance of the infrastructure, and Soil (Medium), related to vulnerability. Meanwhile, variables such as Population Density and Repair Cost have a lower influence on the high flood probability.

On the other hand, in Figure 7b, the sensitivity analysis for FRF with state "Medium" is shown, presenting a different order of the sensitive factors, in which the most sensitive factor is Light Traffic (Low), followed by EPSI (Medium) and Soil type (Medium). The same sub-factors order is present for the more influents of FRF (Medium) but not for the less sensitive, which are population density (Low) and Light Traffic (High) as shown in Figure 7c.

Once the network has been established and validated, it is possible to evaluate cases. For instance, the road network of the *Vala de Azambuja* basin presents a final probability of 17% of low flood risks, 17 % moderate, and 67 % probability of high risk (rounded values), based on the data evidence introduced in the BN as shown in Figure 8. In the analyzed scenario, the flood risk of the road network is most likely to be

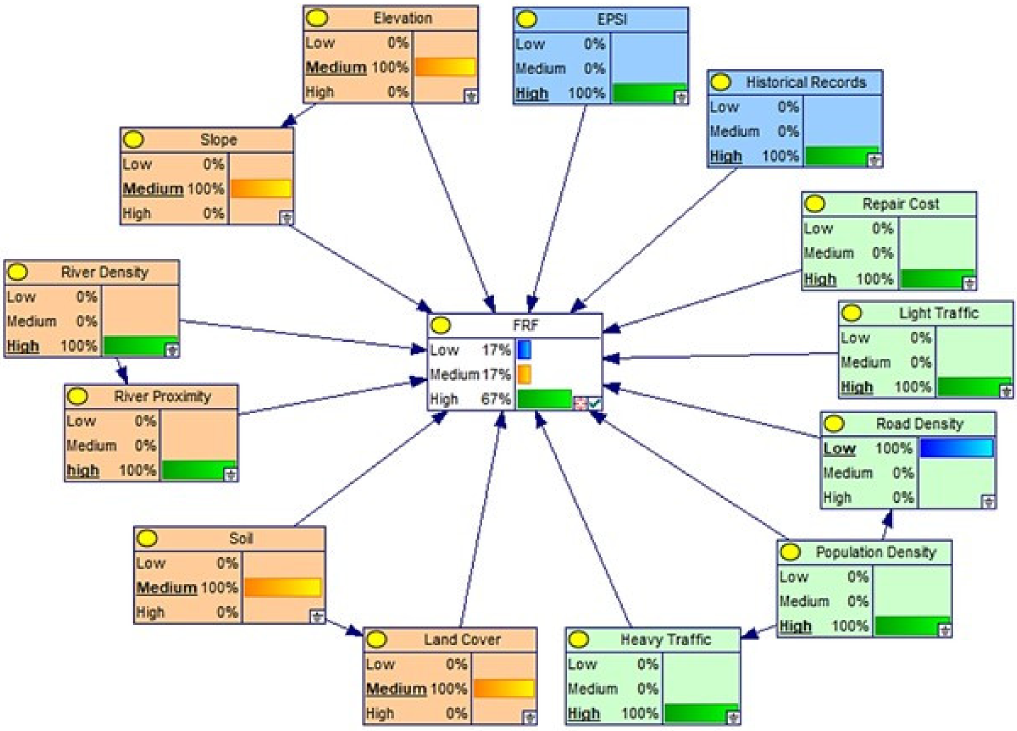

FIGURE 8. BN analysis FRF results of *Vala da Azambuja* road network .

"High" given that this state has obtained the highest probability. If additional information regarding any sub-factor is collected, this new evidence can be easily used for updating the model and reassessing the flood risk.

5. CONCLUSIONS

This work developed a consistent quantitative flood risk estimation and influence factor modelling for road infrastructures. Several factors affecting the flood risks were identified, and weights were assigned to each factor based on expert knowledge and relevant literature in order to quantify a flood risk factor (FRF). Moreover, a BN model was introduced to examine the potential relationships between the factors. It was constructed based on the strength of the relationships found and the probability distribution tables were obtained employing parameter learning algorithms. By using the BN model, it is possible to account for the uncertainties in the input factors and their relationships and propagate them throughout the model to obtain a probability of flood risk. This characteristic is distinct from some previous works where the final risk estimation is computed deterministically. Moreover, the proposed approach has the advantages of enhancing the understanding of the interdependencies among the influencing factors and being flexible for updating the model when new information becomes available. This framework may provide a basis for a decision analysis and decision support with the quantification of expected categorized consequences.

The flood risk in road infrastructures was evaluated based on the proposed methodology in the region of Santarém, Portugal. The case study showed that 61% of the sub-basins present a low flood risk, 30% present a medium flood risk, and 9% present a high flood risk. Moreover, the sensitivity analysis revealed that the most influential sub-factors on a high flood risk were the EPSI factor, light traffic vehicles, and soil type. Thus, the obtained results enabled the identification of the sub-basins where the road infrastructures are heavily exposed to the risk of flooding and provided insights to decision-makers on how the network can be improved to prevent high impacts during flood events and prioritize available resources. The proposed approach can be implemented in other case study areas with similar data availability and can be revised for those with less information available.

There are some limitations to the proposed methodology. First, the established BN model accounted for the most influencing factors of flood risks in road infrastructures, but it was constrained by data availability. In other words, more influencing factors can be included for a more comprehensive risk assessment. For instance, only the reconstruction cost of the roads is included, while indirect impacts originated due to the lack of availability of the roads (e.g., additional travel time) are not accounted for. Additional limitations from the proposed approach are data quality and the fact that certain steps from the method involve subjectivity. For the former issue, the focus of future research will be on how to gather data at a higher spatial resolution to improve the accuracy of the obtained results. For the latter, future work can be oriented towards performing subsequent analyses regarding the effect of the decisions made by the experts in the model.

Acknowledgements

This work was partly financed by FCT / MCTES through national funds (PIDDAC) under the R&D Unit Institute for Sustainability and Innovation in Engineering Structures (ISISE), under reference UIDB / 04029/2020. The first author would like to thank FCT - Portuguese Scientific Foundation for the research grant 2020.05755.BD. The second author would like to thank FCT - Portuguese Scientific Foundation for the research grant SFRH/BD/144749/2019.

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 769255. This document reflects only the views of the author(s). Neither the Innovation and Networks Executive Agency (INEA) nor the European Commission is in any way responsible for any use that may be made of the information it contains.

References

[1] Z. Wu, Y. Shen, H. Wang, et al. Assessing urban flood disaster risk using Bayesian network model and GIS applications. *Geomatics, Natural Hazards and Risk* **10**(1):2163-84, 2019. https://doi.org/10.1080/19475705.2019.1685010.

[2] A. A. Komolafe, S. Herath, R. Avtar. Establishment of detailed loss functions for the urban flood risk assessment in Chao Phraya River basin, Thailand. *Geomatics, Natural Hazards and Risk* **10**(1):633-50, 2019. https://doi.org/10.1080/19475705.2018.1539038 .

[3] W. Duan, B. He, K. Takara, et al. Changes of precipitation amounts and extremes over Japan between 1901 and 2012 and their connection to climate indices. *Climate Dynamics* **45**(7-8):2273-92, 2015. https://doi.org/10.1007/s00382-015-2778-8.

[4] Y. Chen, R. Liu, D. Barrett, et al. A spatial assessment framework for evaluating flood risk under extreme climates. *Science of The Total Environment* **538**:512-23, 2015. https://doi.org/10.1016/j.scitotenv.2015.08.094.

[5] L. Lin, C. Hu, Z. Wu. Assessment of Flood Hazard Based on Underlying Surface Change by Using GIS and Analytic Hierarchy Process. textitGeo-Spatial Knowledge and Intelligence. Communications in Computer and Information Science, pp. 589-99, 2017. https://doi.org/10.1007/978-981-10-3966-9_65.

[6] A. Ghosh, S. K. Kar. Application of analytical hierarchy process (AHP) for flood risk assessment: a case study in Malda district of West Bengal, India. *Natural Hazards* **94**(1):349-68, 2018. https://doi.org/10.1007/s11069-018-3392-y.

[7] L. Bertilsson, K. Wiklund, I. de Moura Tebaldi, et al. Urban flood resilience - A multi-criteria index to integrate flood resilience into urban planning. *Journal of Hydrology* **573**:970-82, 2019. https://doi.org/10.1016/j.jhydrol.2018.06.052.

[8] Y. Sado-Inamura, K. Fukushi. Empirical analysis of flood risk perception using historical data in Tokyo. *Land Use Policy* **82**:13-29, 2019. https://doi.org/10.1016/j.landusepol.2018.11.031.

[9] W. Li, K. Lin, T. Zhao, et al. Risk assessment and sensitivity analysis of flash floods in ungauged basins using coupled hydrologic and hydrodynamic models. *Journal of Hydrology* **572**:108-20, 2019. https://doi.org/10.1016/j.jhydrol.2019.03.002.

[10] G. Zhao, B. Pang, Z. Xu, et al. Assessment of urban flood susceptibility using semi-supervised machine learning model. *Science of The Total Environment* **659**:940-9, 2019. https://doi.org/10.1016/j.scitotenv.2018.12.217.

[11] Z. Wang, C. Lai, X. Chen, et al. Flood hazard risk assessment model based on random forest. *Journal of Hydrology* **527**:1130-41, 2015. https://doi.org/10.1016/j.jhydrol.2015.06.008.

[12] Z. Wu, Y. Zhou, H. Wang, et al. Depth prediction of urban flood under different rainfall return periods based on deep learning and data warehouse. *Science of The Total Environment* **716**, 2020. https://doi.org/10.1016/j.scitotenv.2020.137077.

[13] W. Chen, Y. Li, W. Xue, et al. Modeling flood susceptibility using data-driven approaches of naïve Bayes tree, alternating decision tree, and random forest methods. *Science of The Total Environment* **701**, 2020. https://doi.org/10.1016/j.scitotenv.2019.134979.

[14] H. Darabi, B. Choubin, O. Rahmati, et al. Urban flood risk mapping using the GARP and QUEST models: A comparative study of machine learning techniques. *Journal of Hydrology* **569**:142-54, 2019. https://doi.org/10.1016/j.jhydrol.2018.12.002.

[15] S. Huang, H. Wang, Y. Xu, et al. Key Disaster-Causing Factors Chains on Urban Flood Risk Based on Bayesian Network. *Land* **10**(2), 2021. https://doi.org/10.3390/land10020210.

[16] P. A. Aguilera, A. Fernández, R. F. Ropero, et al. Groundwater quality assessment using data clustering based on hybrid Bayesian networks. *Stochastic Environmental Research and Risk Assessment* **27**(2):435-47, 2012. https://doi.org/10.1007/s00477-012-0676-8.

[17] A. Couasnon, A. Sebastian, O. Morales-Nápoles. A Copula-Based Bayesian Network for Modeling Compound Flood Hazard from Riverine and Coastal Interactions at the Catchment Scale: An Application to the Houston Ship Channel, Texas. *Water* **10**(9):20, 2018. https://doi.org/10.20944/preprints201808.0072.v4.

[18] M. J. Flores, R. F. Ropero, R. Rumí. Assessment of flood risk in Mediterranean catchments: an approach based on Bayesian networks. *Stochastic Environmental Research and Risk Assessment* **33**(11-12):1991-2005, 2019. https://doi.org/10.1007/s00477-019-01746-3.

[19] M. Sanuy, J. A. Jiménez, N. Plant. A Bayesian Network methodology for coastal hazard assessments on a regional scale: The BN-CRAF. *Coastal Engineering* **157**, 2020. https://doi.org/10.1016/j.coastaleng.2019.103627.

[20] S. Balbi, F. Villa, V. Mojtahed, et al. A spatial Bayesian network model to assess the benefits of early warning for urban flood risk to people. *Natural Hazards and Earth System Sciences* **16**(6):1323-37, 2016. https://doi.org/10.5194/nhess-16-1323-2016.

[21] W. M. Dlamini. A Bayesian belief network analysis of factors influencing wildfire occurrence in Swaziland. *Environmental Modelling & Software* **25**(2):199-208, 2010. https://doi.org/10.1016/j.envsoft.2009.08.002.

[22] R. Liu, Y. Chen, J. Wu, et al. Assessing spatial likelihood of flooding hazard using naïve Bayes and GIS: a case study in Bowen Basin, Australia. *Stochastic Environmental Research and Risk Assessment* **30**(6):1575-90, 2015. https://doi.org/10.1007/s00477-015-1198-y.

[23] R. Liu, Y. Chen, J. Wu, et al. Integrating Entropy-Based Naïve Bayes and GIS for Spatial Evaluation of Flood Hazard. *Risk Analysis* **37**(4):756-73, 2017. https://doi.org/10.1111/risa.12698.

[24] H. Joo, C. Choi, J. Kim, et al. A Bayesian Network-Based Integrated for Flood Risk Assessment (InFRA). *Sustainability* **11**(13), 2019. https://doi.org/10.3390/su11133733.

[25] N. Kazakis, I. Kougias, T. Patsialis. Assessment of flood hazard areas at a regional scale using an index-based approach and Analytical Hierarchy Process: Application in Rhodope-Evros region, Greece. *Science of The Total Environment* **538**:555-63, 2015. https://doi.org/10.1016/j.scitotenv.2015.08.055.

[26] M. Santos, M. Fragoso, J. A. Santos. Regionalization and susceptibility assessment to daily precipitation extremes in mainland Portugal. *Applied Geography* **86**:128-38, 2017. https://doi.org/10.1016/j.apgeog.2017.06.020.

[27] I. P. (APA), Agencia Portuguesa do Ambiente, Bacias Hidrográficas das Massas de Agua de Portugal Continental: CDG SNIAmb, 2018. https://sniambgeoviewer.apambiente.pt/Geodocs/shpzips/wise_vw_surfacewaterbody_basin_ptcont.zip.

[28] J. L. Zêzere, S. Pereira, A. O. Tavares, et al. DISASTER: a GIS database on hydro-geomorphologic disasters in Portugal. *Natural Hazards* **72**(2):503-32, 2014. https://doi.org/10.1007/s11069-013-1018-y.

[29] J. Goncalves, J. C. Fernandes. *Assessment of SRTM-3 DEM in Portugal with topographic map data.* in EARSeL 3D Remote Sensing Workshop, 2005. shorturl.at/muOZ8. .

[30] I. P. Agência. *Portuguesa do Ambiente (APA), Massas de água superficiais Rios de Portugal continental: conjunto de dados geográfico SNIAmb*, 2018. https://sniambgeoviewer.apambiente.pt/Geodocs/shpzips/wise_vw_surfacewaterbody_river_line_ptcont.zip.

[31] K. Karra, C. Kontgis, Z. Statman-Weil, et al. Global land use / land cover with Sentinel 2 and deep learning. *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS*. pp. 4704-7, 2021. https://doi.org/10.1109/igarss47720.2021.9553499.

[32] Esri. *Global map of land use/land cover* (LULC), 2020. https://www.arcgis.com/home/item.html?id=d6642f8a4f6d4685a24ae2dc0c73d4ac.

[33] C. Ballabio, P. Panagos, L. Monatanarella. Mapping topsoil physical properties at European scale using the LUCAS database. *Geoderma* **261**:110-23, 2016. https://doi.org/10.1016/j.geoderma.2015.07.006.

[34] Z. Zeng, G. Tang, Y. Hong, et al. Development of an NRCS curve number global dataset using the latest geospatial remote sensing data for worldwide hydrologic applications. *Remote Sensing Letters* **8**(6):528-36, 2017. https://doi.org/10.1080/2150704x.2017.1297544.

[35] INE. Reultados censos, *National Statistical Institute*, 2021. https://ine.pt/scripts/db_censos_2021.html.

[36] D. Persyn, J. Díaz-Lanchas, J. Barbero. Estimating road transport costs between and within European Union regions. *Transport Policy*, 2020. https://doi.org/10.1016/j.tranpol.2020.04.006.

[37] M. O. Kim, K. Kim, J. H. Yun, et al. Fire risk assessment of cable bridges for installation of firefighting facilities. *Fire Safety Journal* **115**, 2020. https://doi.org/10.1016/j.firesaf.2020.103146.

[38] X. Wu, Z. Jiang, L. Zhang, et al. Dynamic risk analysis for adjacent buildings in tunneling environments: a Bayesian network based approach. *Stochastic Environmental Research and Risk Assessment* **29**(5):1447-61, 2015. https://doi.org/10.1007/s00477-015-1045-1.

[39] L. BayesFusion, GeNIe Modeler Software Package. 2020. https://www.bayesfusion.com/genie/.

[40] U. Kjaerulff, L. C. Van der Gaag. Making Sensitivity Analysis Computationally Efficient, *Proceedings of the Sixteenth Conference on Uncertainty in Artificial Intelligence (UAI2000)*, pp. 317-325, 2013. https://doi.org/10.48550/arXiv.1301.3868.

A. Appendices

A.1. Data collection of vulnerability factors

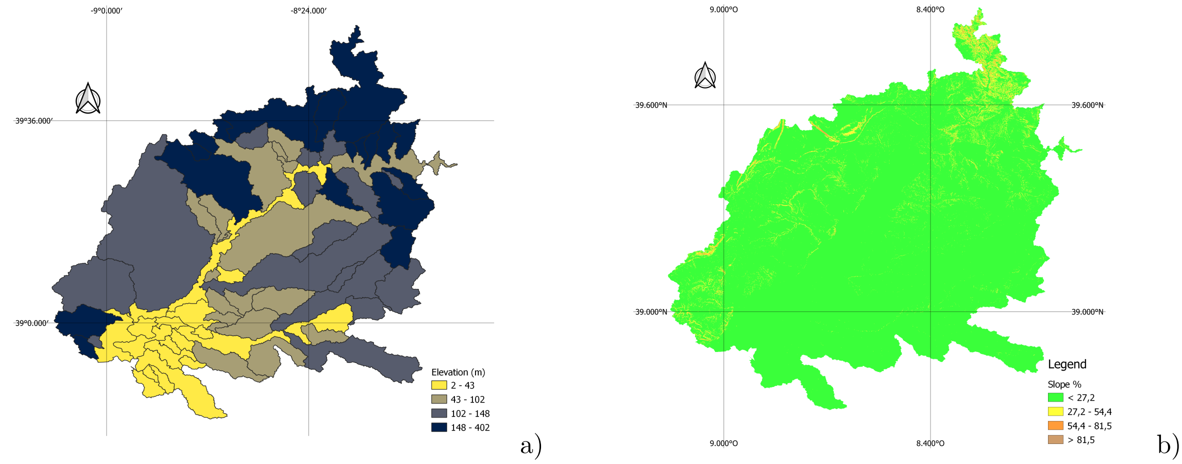

Figure 9. Data collection of vulnerability factors: a) Elevation and b) Slope.

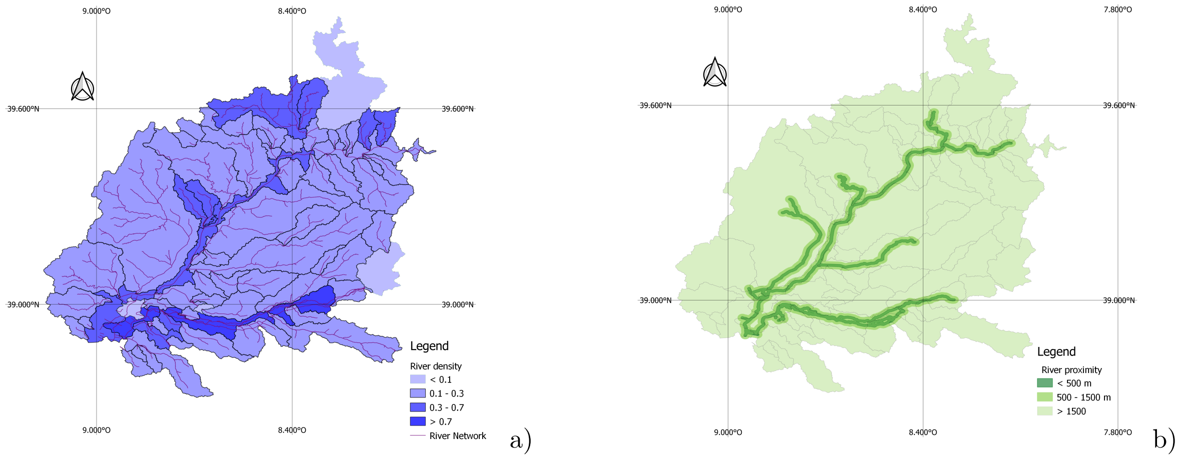

Figure 10. Data collection of vulnerability factors: a) River density and b) River proximity.

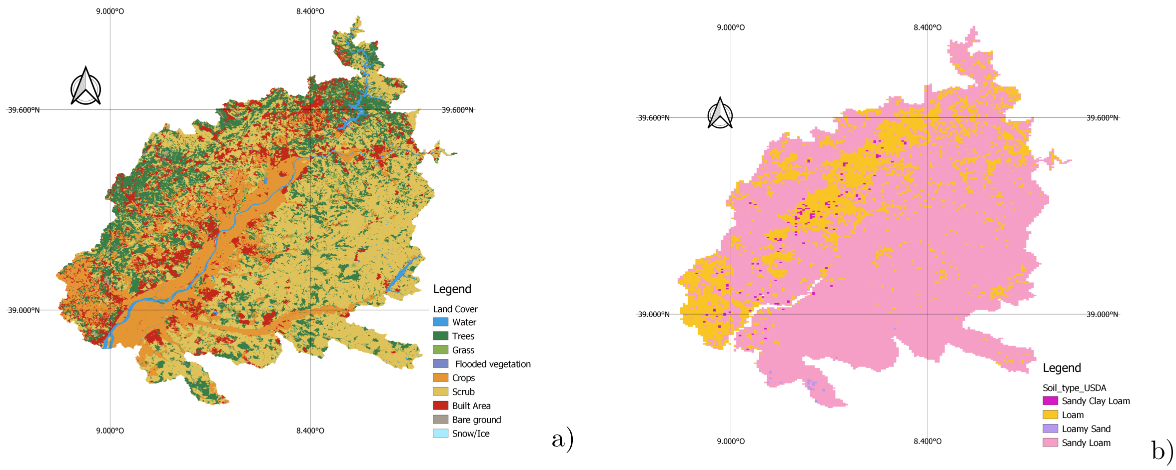

Figure 11. Data collection of vulnerability factors: a) Land Cover and b) Soil Type.

A.2. Data collection of infrastructure importance factors

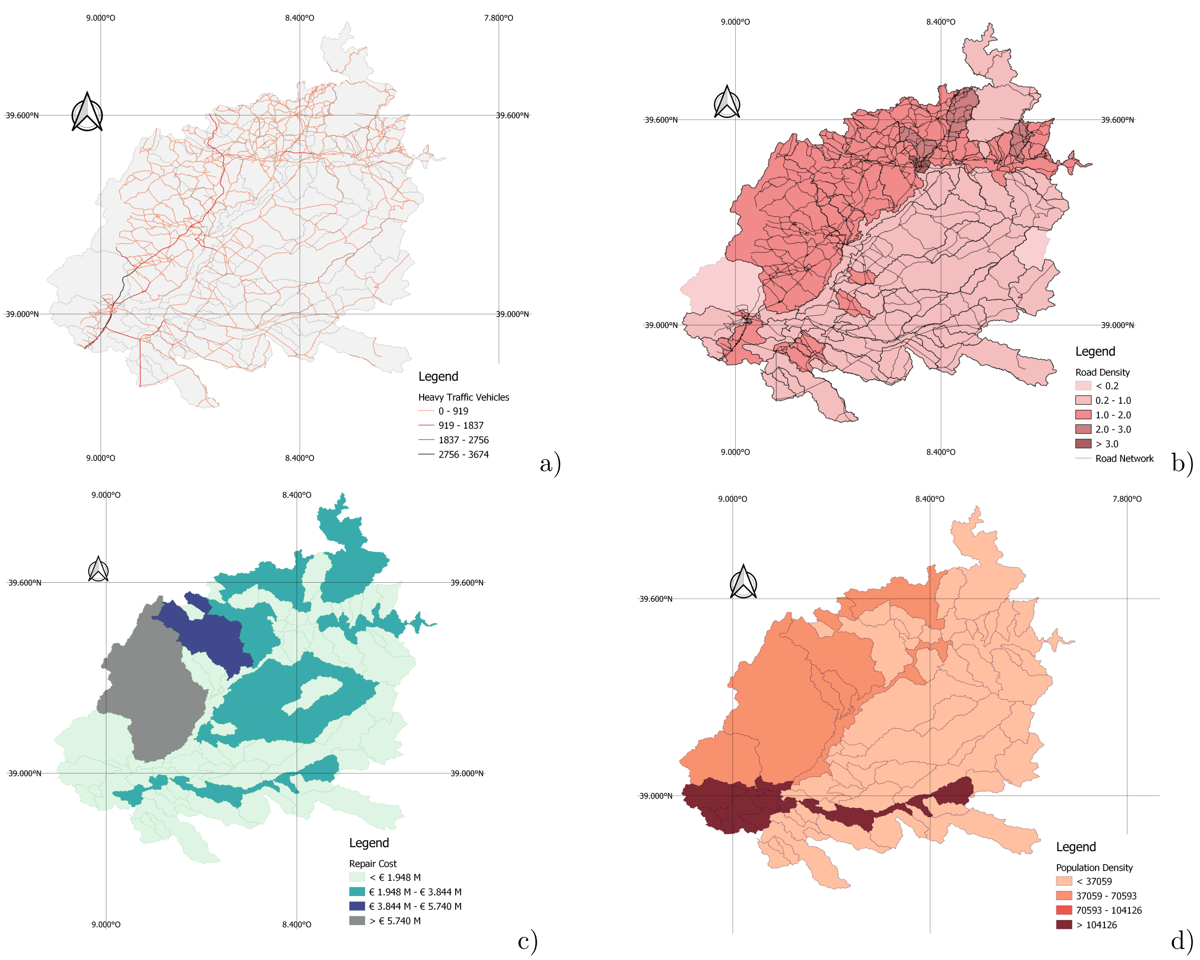

Figure 12. Data collection of infrastructure importance factors: a) Heavy Traffic Vehicles, b) Road density, c) Repair Cost and d) Population Density.