Urban flood risk mapping using the GARP and QUEST models: A comparative study of machine learning techniques

Hamid Darabi[a], Bahram Choubin[b], Omid Rahmati[c], Ali Torabi Haghighi[a], Biswajeet Pradhan[d,e], Bjørn Kløve[a]

[a]Water Resources and Environmental Engineering, University of Oulu, P.O. Box 4300, FIN-90014 Oulu, Finland.

[b]Department of Watershed Management, Sari Agriculture Science and Natural Resources University, P.O. Box 737, Sari, Iran.

[c]Department of Watershed Management, Faculty of Natural Resources and Agriculture, Lorestan University, Iran.

[d]Centre for Advanced Modelling and Geospatial Information Systems (CAMGIS), School of Information, Systems and Modelling, Faculty of Engineering and IT, University of Technology Sydney, 2007 NSW, Australia

[e]Department of Energy and Mineral Resources Engineering, Choongmu-gwan, Sejong University, 209 Neungdong-ro, Gwangjin-gu, Seoul 05006, Republic of Korea

Corresponding author: Email address: bjorn.klove@oulu.fi Tel: +358 40 594 4514

Abstract

Flood risk mapping and modeling is important to prevent urban flood damage. In this study, a flood risk map was produced with limited hydrological and hydraulic data using two state-of-the-art machine learning models: Genetic Algorithm Rule-Set Production (GARP) and Quick Unbiased Efficient Statistical Tree (QUEST). The flood conditioning factors used in modeling were: precipitation, slope, curve number, distance to river, distance to channel, depth to groundwater, land use, and elevation. Based on available reports and field surveys for Sari city (Iran), 113 points were identified as flooded areas (with each flooded zone assigned a value of 1). Different conditioning factors, including urban density, quality of buildings, age of buildings, population density, and socio-economic conditions, were taken into account to analyze flood

vulnerability. In addition, the weight of these conditioning factors was determined based on expert knowledge and Fuzzy Analytical Network Process (FANP). An urban flood risk map was then produced using flood hazard and flood vulnerability maps. The area under the receiver-operator characteristic curve (AUC-ROC) and Kappa statistic were applied to evaluate model performance. The results demonstrated that the GARP model (AUC-ROC=93.5%, Kappa=0.86) had higher performance accuracy than the QUEST model (AUC-ROC=89.2%, Kappa=0.79). The results also indicated that distance to channel, land use, and elevation played major roles in flood hazard determination, whereas population density, quality of buildings, and urban density were the most important factors in terms of vulnerability. These findings demonstrate that machine learning models can help in flood risk mapping, especially in areas where detailed hydraulic and hydrological data are not available.

Keywords: Urban planning, Flood risk management, GIS, FANP, Data-mining.

1. Introduction

Urban areas can flood due to intense precipitation events, rapid snowmelt, and rises in sea, lake, river, and groundwater levels. Major reasons for urban flooding are poor drainage systems, lack of maintenance, and poorly controlled growth of urban areas, especially in developing countries. Low capacity for infiltration or storage during high-intensity rainfall also triggers floods, especially in urban areas. Urban flooding risks will increase with future climate and land use changes (Stocker, 2013), a factor that must be considered in flood management (Fernández and Lutz, 2010). Flood impacts can be mitigated by improved prediction, awareness (early warning), and mapping. Predicting the location of flood-prone areas using flood hazard maps is critical for improved city planning (Büchele et al., 2006). However, due to the complexities of urban

environments, urban flood modeling and prediction of flood-prone areas face many challenges (Chen et al., 2009). Previous studies have used different types of models to assess flood risks in urban areas. These models include Hydrological Simulation Program-FORTRAN (HSPF) (Bicknell et al., 1993), Illinois Urban Drainage Area Simulator (ILLUDAS) (Terstriep and Stall, 1974), Technical Release 55 (TR-55) (USDA, 1986), Hydrologic Engineering Center-River Analysis System (HEC-RAS) (Wiles and Levine, 2002), Storm Water Management Model (SWMM) (Cole and Shutt, 1976), and Urban Flood Cell Model (MODCEL) (Gomes Miguez et al., 2017). Additionally, GIS-based multi-criteria decision analysis tools, such as Analytic Hierarchy Process (Fernández and Lutz, 2010), have been extensively used.

A challenge associated with several urban flood models is that they rely on detailed hydrological and hydraulic data. Therefore, they cannot be used directly in data-scarce environments, such as in developing countries where data availability is still a major challenge. The aim of the present study was to develop new approaches based on Genetic Algorithm Rule-Set Production (GARP) and Quick Unbiased Efficient Statistical Tree (QUEST) to model and map urban flood risk. The GARP modeling tool analyzes the relationship between spatial data and environmental parameters by a training and testing process. The advantages of this model over other models are the genetic algorithm rules set and the stochastic approach, allowing different outputs to be run to reach an optimum result (Stockwell, 1999). The QUEST algorithm has many advantages over other models as it is a quick, impartial, and efficient statistical tree, it employs a linear or unbiased variable selection model, and it uses imputation instead of substitute splits to deal with missing values (Sut and Simsek, 2011). In this study, we adapted the association rules between machine learning techniques with some conditioning hydrological factors, in order to identify flood-prone areas in the urban environment.

We developed a new machine learning technique for analysis of occurrence of flooded areas in urban environments. Specific objectives of the study were to: i) assess the role of different factors resulting in floods in the urban area of Sari city, Iran; ii) predict and validate the flood hazard map by two models, GARP and QUEST; and iii) produce a flooding vulnerability map using Fuzzy Analytical Network Process (FANP), by considering the interconnections among criteria through Fuzzy Decision Making Trial and Evaluation Laboratory (Fuzzy DEMATEL). The novel component of the work lies in applying the GARP and QUEST models together with hydrological variables to predict occurrence of flooded areas, by learning and computing the relationship between all (flood) events. These models are the first automated tools for predicting flooded areas in urban environments.

2. Material and methods

2.1. Study area

Sari city (35°58′39″-36°50′12″N; 52°56′42″-53°59′32″E) is one of the largest cities in northern Iran, and is located at an altitude of between 9 and 82 m above sea level (asl) (Fig. 1). The city has a population of 296,417, making it the second largest city on the southern coast of the Caspian Sea (Zali et al., 2016). It covers an area about 42 km^2 and is located at the outlet of the Tajan river watershed (about 4352 km^2 in area), with the Tajan river passing through the east of Sari and then discharging into the Caspian Sea. The residential areas of the city are surrounded mainly by agricultural land, orchards, and high mountains (the Alborz range) covered by forest. The climate type in the region is dry semi-humid (aridity index 0.73; Sahin, 2012; Choubin et al., 2018), with 734 mm annual precipitation and 1004 mm potential evapotranspiration. For the purposes of the present study, rainfall data (for the period 1986-2016) recorded at Sari weather station were obtained from the Iranian Meteorological Organization (IRIMO).

94

95 **Fig. 1.** SOMEWHERE HERE

96

97 ## 2.2. Materials

98 ### 2.2.1. Conditioning factors of urban flood hazard

99 There are no universal guidelines for selecting flood conditioning factors in urban areas. In the
100 present study, eight different factors were selected, based on the literature, to evaluate flood hazard.
101 These were: rainfall, land use/land cover (LULC), elevation, slope percent, curve number (CN),
102 distance to river, distance to channel, and depth to groundwater (Thieken et al., 2005; Chen et al.,
103 2009; Fernández et al., 2010; Ouma and Tateishi, 2014; Choubin et al., 2019).
104 - **Rainfall:** Daily rainfall data were obtained from the IRIMO to prepare the rainfall amount map.
105 The recorded amount varies from 722 mm in the east of the study area to 745 mm in the west (Fig.
106 2a).
107 - **Land use/land cover:** Runoff conditions vary considerably under different LULC patterns. A
108 LULC map for 2015 was obtained from Sari city authority (Fig. 2b) and, based on this map, the
109 study area was subdivided into three main group types: open spaces (including orchard, parks, and
110 agricultural area), urban districts (residential buildings, commercial, business, and industrial
111 buildings), and water body (river).
112 - **Elevation:** A Digital Elevation Model (DEM) with resolution 5 m was obtained from Sari city
113 authority (Fig. 2c). It confirmed that the elevation of the study area ranges from 9 to 82 m asl.
114 - **Slope percent:** The slope percent factor plays a major role in flooding, as it affects the water
115 velocity. In addition, flatlands or lowlands have gentle slopes that reflect a constant threat of
116 flooding (Wang et al., 2015; Rahmati et al., 2016; Pirnia et al., 2018; Torabi Haghighi et al., 2018).
117 The slope map was extracted from the DEM of the study area in ArcGIS 10.3 to quantify

topographic controls on hydrological processes. The slope varies from more than 6% in the south to less than 1% in the center and north of the study area (Fig. 2d).

- **Curve Number (CN):** Curve number (CN), a parameter developed by the United State Soil Conservation Service (USCS), is a function of land use treatments and hydrological condition, antecedent soil moisture, and soil type. Land use and Hydrologic Soil Group (HSG) maps were used here to estimate the contribution of rainfall to runoff. The CN map for the study area (Fig. 2e) was extracted based on the land use map, the HSG map, and a lumped CN value, using the ArcCN-runoff tool in ArcGIS software (Zhan & Huang, 2004; Darabi et al., 2014; Menberu et al., 2014). As shown in Fig. 2e, different CN class were given corresponding codes, with larger values indicating stronger runoff generation capability.

- **Distance to river:** The banks of the Tajan River are major flood-prone areas in Sari city. Hence, distance to the river plays an important role in urban flood mapping in this city. The Euclidean distance to the river was calculated using Euclidian Distance module in ArcGIS 10.3 (Fig. 2f).

- **Distance to channel:** Channels or drainage systems in the urban environments collect surface water. The map of 'distance to channel' was also prepared using the Euclidian Distance module in ArcGIS 10.3 (Fig. 2).

- **Depth to groundwater level:** Infiltration capacity generally depends on soil moisture and depth to groundwater, which directly affects the surface runoff volume during high-intensity precipitation (Fernández and Lutz, 2010). Some studies have documented that depth to groundwater level is an influential factor in initial storage capacity of a basin (Yin and Li, 2001; Fernández and Lutz, 2010). The groundwater level data used in this study were obtained from Iranian Water Resources Management Company (IWRMC). As can be seen in Fig. 2h, the depth to groundwater ranges from 1.9 m (in the north of the study area) to 20.8 m (in the south).

Among the eight conditioning factors on urban flood inundation hazard, elevation, slope, depth to groundwater, distance to channel, distance to river, and rainfall are continuous factors, while land use and curve number are categorical factors.

Fig. 2. SOMEWHERE HERE

2.2.2. Flood inventory map

Mapping the flood locations in an area is vital in explaining the correlation between flooding and the conditioning factors. In this study, a flood inventory map was prepared based on both multiple field surveys and available documents (flood historical database) obtained from Sari city authority during 2015 to 2017. The location of flooded sites was recorded using a Global Positioning System (GPS) device. In this geographical region, floods occur in the rainy period (December-May). Fig. 1 shows the severity of the flooding that occurred in 2015-2017. In order to develop the urban flood hazard map, flooded and non-flooded areas were assigned a code of 1 and 0, respectively. In this step, the historical records on flood occurrence and inspection provided essential information (Fig. 1). Based on field surveys in urban Sari, a total of 113 points were identified as flooded areas, while 76 non-flooded points were randomly chosen in non-flooded zones. In flood hazard analysis, the flooded locations were randomly divided into two groups, comprising 70% (79 locations) and 30% (34 locations), for the purpose of training and validation, respectively. The non-flooded locations were also randomly split into two groups, for training (70% of locations) and validation (30% of locations).

2.2.3. Vulnerability factors

Vulnerability has been defined as *"the conditions determined by physical, social, economic and environmental factors or processes, which increase susceptibility to the impact of hazards"* (UNDP, 2004). Ouma and Tateishi (2014) describe flood vulnerability assessment as the process of determining the degree of susceptibility of a given location to flooding if information on its exposure to floods is known. There are various socio-environmental factors that influence vulnerability in urban areas and their inclusion may depend on available data (Dayal et al., 2018). In the present study, the factors urban density, quality of buildings, age of buildings, population density, and socio-economic conditions were taken into account.

The *urban density* of Sari was divided into four classes (high, medium, low, very low), according to suggestions by Güneralp et al. (2017) (Fig. 3a). Quality and age of buildings have significant impacts on the damage caused by urban floods. *Quality of buildings* was divided into five classes (very high, high, medium, low, very low), which reflect building condition. There are also areas without buildings, which were included in the quality of buildings map (Fig. 3b). Low quality buildings are unsafe for residents and are more vulnerable to natural hazards, especially earthquakes and floods (Schubert and Sanders, 2012, Gerl et al., 2014). As regards *age of buildings*, the current architectural structure of Sari city reflects its historical development. Densely built-up old and modern buildings with administrative and commercial functions are located in Sari city center. In surrounding districts, settlement areas with multi-age residential buildings and open spaces are abundant. Buildings were divided into five classes based on their age (very old, old, medium, new, newest), and there were also some areas with no buildings (Fig. 3c). Data on the age and quality of buildings in the study area in 2015 were obtained from Sari city authority. *Population density* and *socio-economic conditions* for residents are key issues in preparing a flood vulnerability map. Population density refers to the number of people inhabiting

187 a given urban area, and high levels of population density are usually associated with good
188 economic conditions and higher productivity, but also higher susceptibility to natural hazards such
189 as earthquakes and floods (Güneralp et al., 2017). In this study, population density was divided
190 into four classes (high, medium, low, very low) (Fig. 3d). Socioeconomic data contain in-depth
191 information on the inherent properties and behavior of humans and society within a specific
192 geographical region. These types of information are valuable when considering indirect and
193 intangible impacts of natural hazards such as flooding (Kaspersen and Halsnæs, 2017). In the
194 present study, socio-economic conditions were divided into five classes (A, B, C, D, E, reflecting
195 very good, good, moderate, weak, and very weak socio-economic conditions respectively) (Fig.
196 3e). There are also natural areas surrounding Sari city, which were not considered residential areas.
197 Underlying data on population density and socio-economic conditions in Sari in 2015 were
198 obtained from Sari city authority. The classification of all five variables was carried out by Sari
199 city authority.
200
201 **Fig. 3.** SOMEWHERE HERE
202
203 **2.3. Risk prediction**
204 Risk is a function of hazard and vulnerability. Therefore, the urban flood risk map for Sari was
205 produced through flood hazard and vulnerability maps, using equation 1 (Dewan, 2013):
206 $$Risk = Hazard \times Vulnerability \quad (1)$$
207
208 The models used for predicting the flood hazard, model performance, the process of preparing the
209 vulnerability map based on FANP, and extraction of the flood risk map are described below.

2.3.1. Flood hazard prediction

Two state-of-the-art machine learning models, GARP and QUEST, were applied to produce the flood hazard map.

2.3.1.1. Genetic Algorithm for Rule-set Prediction (GARP)

GARP is a machine learning algorithm that has shown excellent predictive capability in different fields such ecological modeling (Stockwell, 1999; Peterson et al., 2002a). The GARP algorithm was selected to predict flood inundation hazard in the study area. It is inspired by models of genetic evolution as a presence-only modeling tool that analyzes the relationship between flood inundation dataset and topo-hydrological variables through an iterative process and conditional rules for model building (Zhu et al., 2007; Sánchez-Flores, 2007; Boeckmann and Joyner, 2014; Qin et al., 2015). The GARP algorithm (Boeckmann and Joyner, 2014) produces a number of flood inundation predictions for urban areas through an iterative process to improve the stability of model output. Performing multiple runs to obtain different outputs of model and using the best-subset method are important in selecting the best output with optimum parameters. The set of models that achieves harmony between omission (sensitivity) and commission (specificity) error thresholds is defined by the user (Anderson, 2003; Boeckmann and Joyner, 2014). The GARP output is a collection of grids of the study area, which can be used in a GIS environment to identify flood-prone areas (Boeckmann and Joyner, 2014). In this study, the GARP model was run using DesktopGARP software. Based on optimum combinations of error components, the 10 best-subset models were chosen out of the 100 repeats (Anderson, 2003; Sobek-Swant et al., 2012). In addition, the importance of conditioning factors (ICF) (precipitation, slope, curve number, distance

to river, distance to channels, depth to groundwater, land use, and elevation) for urban flood hazard was analyzed using the GARP model. A complete mathematical and technical description of GARP model can be found in Peterson et al. (2002b) and Fitzpatrick et al. (2007).

2.3.1.2. Quick, unbiased, and efficient statistical tree (QUEST)

QUEST (Loh and Shih, 1997) is a popular data-mining model which produces subsets of the data that are as homogeneous as possible with respect to the response variable (Rattray et al., 2009; Ture et al., 2009). QUEST is a tree-structured classification algorithm that yields a growing binary-split decision tree (Lee and Park, 2013). It employs a sequential tree growing method, which utilizes a linear discriminate analysis method in splitting tree nodes. This has many advantages over recursive tree construction methods such as classification and regression tree (CART) (Ierodiaconou et al., 2011). In addition, it is unbiased in choosing splitting rules and does not use an exhaustive variable search routine (Sut and Simsek, 2011).

The QUEST algorithm was selected as the second model to predict flood inundation. Moreover, the QUEST algorithm applies imputation instead of surrogate splitting to deal with missing values. According to Ture et al. (2005), QUEST has a negligible bias because it uses an unbiased variable selection technique in modeling. Therefore, QUEST can easily handle categorical and continuous factors (Chou, 2012; Lee and Park, 2013; Lee and Lee, 2015).

2.3.2. Evaluating the predictive performance of models

The receiver-operator characteristic (ROC) was used to evaluate of the performance of models (Gorsevski, 2006; Jiménez-Valverde et al., 2012). The area under the curve (AUC-ROC) has been widely used for evaluating model accuracy (Frattini et al., 2010). The AUC-ROC value is the

probability that a test record is accurately differentiated from a random point in the predetermined context of the study area (Phillips and Dudík, 2008; Kornejady et al., 2017). Area under the curve values range from 0 to 1; at an AUC-ROC range of 0.5-0.6, 0.6-0.7, 0.7-0.8, 0.8-0.9, and 0.9-1, models are classified as weak, average, good, very good, and excellent, respectively (Yesilnacar, 2005). The AUC-ROC is one of the most popular evaluation criterion to assess the performance of different models that produce success and prediction rates (Tehrany et al., 2015).

The Kappa statistic was also applied to evaluate model performance in this study. It uses model classification probabilities to calculate the likelihood of agreement by chance based on null hypothesis investigation (Monserud and Leemans, 1992). According to Monserud and Leemans (1992), the Kappa statistic can be classified into five classes of performance: $k<0.4$, $0.4<k<0.55$, $0.55<k<0.85$, $0.85<k<0.99$, and $0.99<k<1.00$, corresponding to poor, moderate, good, excellent, and perfect, respectively. All performance analyses were carried out in R software.

2.3.3. Urban flood vulnerability map

Urban density, quality and age of buildings, population density, and socio-economic conditions (Fig. 4) were used to determine vulnerable areas to flood inundation events. The relative weights of these factors were determined using FANP, which is one of many decision-making techniques that incorporate the Analytical Network Process (ANP) with Fuzzy set theory. FANP is conducted in five steps (Buyukozkan and Cifci, 2012; Sajedi-Hosseini et al., 2018a): i) Transformation of the problem to a network structure. This first step was done using Fuzzy DEMATEL to design a network structure from factors influencing vulnerability to floods. Fuzzy DEMATEL has been widely used to solve the structure of complex problems through a visual structural model and to assess the causal relationship between factors (Wu and Lee, 2007). In Fuzzy DEMATEL, the

279 directed influential degrees between pair-wise criteria are expressed as Fuzzy interval numbers
280 (Table 1). For more details of Fuzzy DEMATEL, see Chang et al. (2011), Dalalah et al. (2011),
281 and Sajedi-Hosseini et al. (2018a). ii) Pairwise comparisons of criteria based on their importance,
282 using Fuzzy extent analysis as described in Chang et al. (2011). Triangular Fuzzy numbers to
283 pairwise comparisons are presented in Table 1. iii) Calculation of the initial super-matrix based on
284 the weights obtained from the previous step. iv) Computation of the weighted super-matrix through
285 multiplying the initial super-matrix by cluster weights. v) Conversion of the weighted super-matrix
286 into a limit super-matrix and determination of priorities and the importance of factors. The FANP
287 and Fuzzy DEMATEL methods were used in the Super Decision and MATLAB software,
288 respectively, and then output layers were overlaid in the GIS environment.
289
290 **Table 1:** SOMEWHERE HERE
291
292 **3. Results**
293 **3.1. Performance of models**
294 The efficiency and precision of the GARP and QUEST models were assessed using AUC-ROC
295 and Kappa evaluation criteria (Fig. 4 and Table 2). According to the validation results (Fig. 4 and
296 Table 2), GARP and QUEST achieved 93.50% and 89.20% prediction rates, respectively.
297 Therefore, the efficiency of the GARP model was somewhat better than that of QUEST. The
298 Kappa value was 0.86 (excellent) for GARP and 0.79 (good) for QUEST (Table 2).
299
300 **Fig. 4.** SOMEWHERE HERE
301 **Table 2.** SOMEWHERE HERE

3.2. Urban flood hazard mapping

The urban flood maps produced by the GARP and QUEST models (Fig. 5a and 5b) illustrated the flood inundation probability over the study area. Both models demonstrated that zones with high hazard probability are mostly located in the north and center of Sari city (along the Tajan River). As Fig. 5 shows, both modeling approaches were similar in terms of flood inundation prediction. Both models provided valuable information for mapping flood hazard, even though there were no detailed hydrological and hydraulic data available.

Fig. 5. SOMEWHERE HERE

3.3. Contribution analysis of predictive factors

The results of the GARP model indicated that distance to channel (ICF=1.00), land use (ICF=0.96), and elevation (ICF=0.89) were the most important factors, followed by curve number (ICF=0.83), distance to river (ICF=0.74), depth to groundwater (ICF=0.57), rainfall (ICF=0.38), and slope (ICF=0.22) (Fig. 6). Therefore, all thematic layers made a significant contribution in flood inundation modeling, and hence all were used as independent variables in the GARP and QUEST models to generate the urban flood hazard map.

Fig. 6 SOMEWHERE HERE

Probability curves created by the GARP model for each of the continuous factors are presented in Fig. 7. As Fig. 7a shows, with altitude up to 30 m asl, the probability of flood inundation hazard increased, while above an altitude of 30 m asl it decreased. Altitudes above 50 m asl were

designated as the most favorable class from a flood hazard point of view. As can be seen in Fig. 7b, with increasing slope up to 1.4%, flood inundation hazard increased, while with slope >1.4% the probability of flood inundation hazard decreased. Also, when the depth to groundwater increased, the flood inundation hazard significantly decreased. In this study, the average probability of flood inundation hazard decreases by 0.02 with increasing depth of groundwater for every 2 meters (Fig. 7c). As can be seen from Figs. 7d and 7e, flooding inundation hazard increased with decreasing distance to river and distance to channel, and these two variables showed similar responses to the probability of flood inundation occurrence. Increasing the distance from the river and channels for every 200 meters, the average probability of flood inundation hazard decreases by 0.045 and 0.025, respectively. Furthermore, as shown in Fig. 7f, the probability of flood inundation occurrence increased as rainfall increased. By increasing the rainfall for every 2 millimeters, the average probability of flood inundation hazard decreases by 0.017.

Fig. 7. SOMEWHERE HERE

3.4. Vulnerability map

Based on the FANP results on the relative importance of urban flood vulnerability factors, population density (0.370), quality of buildings (0.185), and urban density (0.148) were the most important factors, followed by age of buildings (0.148) and socio-economic conditions (0.147). The weights assigned to each class of the urban density, quality of buildings, age of buildings, population density, and socio-economic factors (based on the expert knowledge and the FANP method) are summarized in Table 3.

Table 3: SOMEWHERE HERE

The urban flood vulnerability value for each part of the city obtained using the FANP method is shown in Fig. 8a. The results showed that the most vulnerable flooding zones are located in the center of Sari city. For better visual interpretation of urban flood vulnerability, the vulnerability map was classified into five classes (Fig. 8b): very low, low, moderate, high, and very high, occupying 35.49%, 15.25%, 13.52%, 22.61%, and 13.10% of the study area, respectively.

Fig. 8. SOMEWHERE HERE

3.5. Risk map

In the flood risk index map, the risk value ranged from 0.05 to 0.76 (Fig. 9a). The flood risk map was classified into five classes using the natural break method: very low, low, moderate, high, and very high, occupying 32.90%, 21.71%, 21.98%, 7.70%, and 15.69% of the study area, respectively (Fig. 9b). The flood risk map indicated that central and northern sites in Sari city are most exposed to flood risk.

Fig. 9. SOMEWHERE HERE

3.6. Data scarcity and flood risk modelling

Although there are several hydrologic/hydraulic models to analyze flood hazard, robust flood hazard maps are still lacking in data-scarce regions (Gigović et al., 2017; Rahmati and Pourghasemi, 2017; Samela et al., 2017). As discussed by Fernández and Lutz (2010) and Ouma and Tateishi (2014), flood risk assessment in data scarce regions, especially urban environments, using hydraulic models still remains challenging. In this study, urban flood hazard was

quantitatively predicted using GARP and QUEST models which have simple requirements in terms of input data, computational time, and costs.

5. Conclusions

Accurate flood risk assessment is vital for effective urban water management and sustainable urban development. Urban floods are influenced by different factors, which in developing countries are often related to unplanned urban development along lowlands and river banks (e.g., river floods), poor maintenance, and clogging of urban drains. In this study, two models (GARP and QUEST) were applied for the first time to produce urban flood risk maps. The results confirm that machine learning techniques can be applied to urban flood zoning. The hazard and vulnerability map revealed that distance to channel and population density are important factor in risk mapping. Data mining methods were applied here for flood risk assessment using GARP and QUEST, demonstrating that reliable results can be obtained without expensive field surveys or complex hydrodynamic modelling. The approach presented is particularly useful in regions with little data, to quickly define and expose flood hazards. Hazard and vulnerability mapping can also serve as a first step in developing flood risk reduction strategies and in allocating resources for flood defenses and forecasting and warning systems. Our analysis showed that proper design and maintenance of drainage systems is essential for sustainable urban management, and that planning to distribute urban residents equally in all areas can be important in reducing flood risks.

Acknowledgments

394 The authors would like to thank Sari city authority and Moretza Shabani, GIS expert in the
395 Department of Watershed Management and Engineering, Sari Agriculture Science and Natural
396 Resources University, for their contributions.

397

398 **References**

399 Anderson, R.P., 2003. Real vs. artefactual absences in species distributions: tests for Oryzomys
400 albigularis (Rodentia: Muridae) in Venezuela. J. Biogeogr. 30(4), 591-605.

401 Becknell, B.R., Imhoff, J. C., Kittle, J. L., Donigian, A. S., Johanson, R.C., 1993. Hydrological
402 simulation program: FORTRAN. User's manual for release 10 (No. PB-94-114865/XAB).
403 AQUA TERRA Consultants, Mountain View, CA (United States).

404 Behrens, T., Scholten, T., 2006. A comparison of data-mining techniques in predictive soil
405 mapping. Dev. Soil. Sci. 31, 353-617.

406 Boeckmann, M., Joyner, T.A., 2014. Old health risks in new places? An ecological niche model
407 for I. ricinus tick distribution in Europe under a changing climate. Health & place. 30, 70-
408 77.

409 Büchele, B., Kreibich, H., Kron, A., Thieken, A., Ihringer, J., Oberle, P., Nestmann, F., 2006.
410 Flood-risk mapping: contributions towards an enhanced assessment of extreme events and
411 associated risks. Nat. Hazard Earth Sys. 6(4), 485-503.

412 Büyüközkan, G., Çifçi, G., 2012. A novel hybrid MCDM approach based on fuzzy DEMATEL,
413 fuzzy ANP and fuzzy TOPSIS to evaluate green suppliers. Expert Syst. Appl. 39(3), 3000-
414 3011.

415 Chang, B., Chang, C. W., Wu, C.H., 2011. Fuzzy DEMATEL method for developing supplier
416 selection criteria. Expert Syst. Appl. 38(3), 1850-1858.

417 Chapi, K., Singh, V.P., Shirzadi, A., Shahabi, H., Bui, D.T., Pham, B.T., Khosravi, K., 2017. A
418 novel hybrid artificial intelligence approach for flood susceptibility assessment. Environ.
419 Modell. Softw. 95, 229-245.

420 Chen, J., Hill, A.A., Urbano, L.D., 2009. A GIS-based model for urban flood inundation. J. Hydrol.
421 373(1), 184-192.

422 Chen, Y., Zhou, H., Zhang, H., Du, G., Zhou, J., 2015. Urban flood risk warning under rapid
423 urbanization. Environ Res. 139, 3-10.

424 Cherqui, F., Belmeziti, A., Granger, D., Sourdril, A., Le Gauffre, P., 2015. Assessing urban
425 potential flooding risk and identifying effective risk-reduction measures. Sci. Total Environ.
426 514, 418-425.

427 Chou, J.S., 2012. Comparison of multilabel classification models to forecast project dispute
428 resolutions. . Expert Syst. Appl. 39(11), 10202-10211.

429 Choubin, B., Darabi, H., Rahmati, O., Sajedi-Hosseini, F., Kløve, B., 2018. River suspended
430 sediment modelling using the CART model: A comparative study of machine learning
431 techniques. Sci. Total Environ. 615, 272-281.

432 Choubin, B., Moradi, E., Golshan, M., Adamowski, J., Sajedi-Hosseini, F., Mosavi, A., 2019. An
433 Ensemble prediction of flood susceptibility using multivariate discriminant analysis,
434 classification and regression trees, and support vector machines. Sci. Total Environ. 651,
435 2087-2096.

436 Cole, G.D., Shutt, J.W., 1976. SWMM as a predictive model for runoff. In: Proceedings of the
437 National Symposium on Urban Hydrology, Hydraulics and Sediment Control. Univ.
438 Kentucky, Lexington, KY, USA, pp. 193-201.

439 Dalalah, D., Hayajneh, M., Batieha, F., 2011. A fuzzy multi-criteria decision making model for
440 supplier selection. Expert Syst. Appl. 38(7), 8384-8391.

441 Darabi, H., Shahedi, K., Solaimani, K., Miryaghoubzadeh, M., 2014. Prioritization of
442 subwatersheds based on flooding conditions using hydrological model, multivariate analysis
443 and remote sensing technique. Water. Environ. J. 28(3), 382-392.

444 Dayal, K.S., Deo, R.C. and Apan, A.A., 2018. Spatio-temporal drought risk mapping approach
445 and its application in the drought-prone region of south-east Queensland, Australia. Nat.
446 Hazards. 93, 823–847.

447 De Meyer, M., Robertson, M. P., Peterson, A. T., Mansell, M.W., 2008. Ecological niches and
448 potential geographical distributions of Mediterranean fruit fly (Ceratitis capitata) and Natal
449 fruit fly (Ceratitis rosa). J. Biogeogr. 35(2), 270-281.

Dewan, A., 2013. Floods in a megacity: geospatial techniques in assessing hazards, risk and vulnerability. Springer Science & Business Media.

Ernst, J., Dewals, B.J., Detrembleur, S., Archambeau, P., Erpicum, S., Pirotton, M., 2010. Micro-scale flood risk analysis based on detailed 2D hydraulic modelling and high resolution geographic data. Nat. Hazards. 55(2), 181-209.

Fernández, D. S., Lutz, M. A. (2010). Urban flood hazard zoning in Tucumán Province, Argentina, using GIS and multicriteria decision analysis. Eng. Geol. 111(1-4), 90-98.

Fitzpatrick, M.C., Weltzin, J.F., Sanders, N.J. and Dunn, R.R., 2007. The biogeography of prediction error: why does the introduced range of the fire ant over-predict its native range?. Global Ecol Biogeogr. 16(1), 24-33.

Frattini, P., Crosta, G. and Carrara, A., 2010. Techniques for evaluating the performance of landslide susceptibility models. Eng. Geol. 111(1-4), 62-72.

Gerl, T., Bochow, M., Kreibich, H., 2014. Flood damage modeling on the basis of urban structure mapping using high-resolution remote sensing data. Water. 6(8), 2367-2393.

Gigović, L., Pamučar, D., Bajić, Z. and Drobnjak, S., 2017. Application of GIS-interval rough AHP methodology for flood hazard mapping in urban areas. Water, 9(6), p.360. doi:10.3390/w9060360.

Gomes Miguez, M., Peres Battemarco, B., Martins De Sousa, M., Moura Rezende, O., Pires Veról, A., Gusmaroli, G., 2017. Urban Flood Simulation Using MODCEL—An Alternative Quasi-2D Conceptual Model. Water. 9(6), 445

Gorsevski, P.V., Gessler, P.E., Foltz, R.B., Elliot, W.J., 2006. Spatial prediction of landslide hazard using logistic regression and ROC analysis. T. GIS. 10(3), 395-415.

Güneralp, B., Zhou, Y., Ürge-Vorsatz, D., Gupta, M., Yu, S., Patel, P. L., Seto, K.C., 2017. Global scenarios of urban density and its impacts on building energy use through 2050. P. Natl. Acad. Sc. 201606035.

Hong, H., Tsangaratos, P., Ilia, I., Liu, J., Zhu, A.X., Chen, W., 2018. Application of fuzzy weight of evidence and data mining techniques in construction of flood susceptibility map of Poyang County, China. Sci. Total Environ. 625, 575-588.

Ierodiaconou, D., Monk, J., Rattray, A., Laurenson, L. and Versace, V.L., 2011. Comparison of automated classification techniques for predicting benthic biological communities using hydroacoustics and video observations. Cont. Shelf. Res. 31(2), 28-38.

Jiménez-Valverde, A., 2012. Insights into the area under the receiver operating characteristic curve (AUC) as a discrimination measure in species distribution modelling. Global Ecol. Biogeogr. 21(4), 498-507.

Kaspersen, P.S., Halsnæs, K., 2017. Integrated climate change risk assessment: A practical application for urban flooding during extreme precipitation. Clim. Serv. 6, 55-64.

Koks, E.E., Jongman, B., Husby, T.G., Botzen, W.J., 2015. Combining hazard, exposure and social vulnerability to provide lessons for flood risk management. Environ. Sci. Policy. 47, 42-52.

Kornejady, A., Ownegh, M., Rahmati, O., Bahremand, A., 2017. Landslide susceptibility assessment using three bivariate models considering the new topo-hydrological factor: HAND. Geocarto Int. doi:10.1080/10106049.2017.1334832.

Lee, S., Lee, C. W. 2015. Application of decision-tree model to groundwater productivity-potential mapping. Sustainability. 7(10), 13416-13432.

Lee, S., Park, I. 2013. Application of decision tree model for the ground subsidence hazard mapping near abandoned underground coal mines. J. Environ. Manage.127, 166-176.

Loh, W.Y., Shih, Y.S., 1997. Split selection methods for classification trees. Statistica sinica, 815-840.

Menberu, M. W., Haghighi, A. T., Ronkanen, A. K., Kværner, J., Kløve, B., 2014. Runoff curve numbers for peat-dominated watersheds. J. Hydrol. Eng. 20(4), 04014058.

Monserud, R.A., Leemans, R., 1992. Comparing global vegetation maps with the Kappa statistic. Ecol. model. 62(4), 275-293.

Otto, A., Hornberg, A., Thieken, A., 2018. Local controversies of flood risk reduction measures in Germany. An explorative overview and recent insights. J. Flood Risk Manag. 11, 382-394.

Ouma, Y.O., Tateishi, R., 2014. Urban flood vulnerability and risk mapping using integrated multi-parametric AHP and GIS: methodological overview and case study assessment. Water. 6(6), 1515-1545.

Peterson, A.T., Ball, L.G. and Cohoon, K.P., 2002a. Predicting distributions of Mexican birds using ecological niche modelling methods. Ibis, 144(1), 27-32.

Peterson, A.T., Ortega-Huerta, M.A., Bartley, J., Sánchez-Cordero, V., Soberón, J., Buddemeier, R.H. and Stockwell, D.R., 2002b. Future projections for Mexican faunas under global climate change scenarios. Nature, 416(6881), p.626. https://doi.org/10.1038/416626a

Phillips, S.J., Dudík, M., 2008. Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. Echography, 31(2), 161-175.

Pirnia, A., Golshan, M., Darabi, H., Adamowski, J., Rozbeh, S., 2018. Using the Mann–Kendall test and double mass curve method to explore stream flow changes in response to climate and human activities. Journal of Water and Climate Change, jwc2018162.

Qin, H.P., Li, Z.X., Fu, G., 2013. The effects of low impact development on urban flooding under different rainfall characteristics. J. Environ. Manage. 129, 577-585.

Qin, Z., Zhang, J. E., DiTommaso, A., Wang, R. L., Wu, R.S., 2015. Predicting invasions of Wedelia trilobata (L.) Hitchc. With Maxent and GARP models. J. Plant Res. 128(5), 763-775.

Rahmati, O., Pourghasemi, H.R., 2017. Identification of critical flood prone areas in data-scarce and ungauged regions: A comparison of three data mining models. Water Resour. Manag. 31(5), 1473-1487.

Rahmati, O., Zeinivand, H., Besharat, M., 2016. Flood hazard zoning in Yasooj region, Iran, using GIS and multi-criteria decision analysis. Geomat. Nat. Haz. Risk 7(3), 1000-1017.

Rattray, A., Ierodiaconou, D., Laurenson, L., Burq, S. and Reston, M., 2009. Hydro-acoustic remote sensing of benthic biological communities on the shallow South East Australian continental shelf. Estuar.Coastal. Shelf Sci. 84(2), 237-245.

Sahin, S., 2012. An aridity index defined by precipitation and specific humidity. J. Hydrol. 444, 199-208.

Sajedi-Hosseini, F., Choubin, B., Solaimani, K., Cerdà, A., Kavian, A., 2018a. Spatial prediction of soil erosion susceptibility using FANP: application of the Fuzzy DEMATEL approach. Land Degrad. Dev. https://doi.org/10.1002/ldr.3058

Sajedi-Hosseini, F., Malekian, A., Choubin, B., Rahmati, O., Cipullo, S., Coulon, F., Pradhan, B., 2018b. A novel machine learning-based approach for the risk assessment of nitrate groundwater contamination. Sci. Total Environ. 644, 954-962.

Samela, C., Troy, T.J., Manfreda, S., 2017. Geomorphic classifiers for flood-prone areas delineation for data-scarce environments. Adv. Water Resour. 102, 13-28.

Sánchez-Flores, E., 2007. GARP modeling of natural and human factors affecting the potential distribution of the invasive Schismus arabicus and Brassica tournefortii in 'El Pinacate y Gran Desierto de Altar'Biosphere Reserve. Ecol. Model. 204(3), 457-474.

Schubert, J. E., Sanders, B.F., 2012. Building treatments for urban flood inundation models and implications for predictive skill and modeling efficiency. Adv. Water Resour. 41, 49-64.

Sobek-Swant, S., Kluza, D. A., Cuddington, K., Lyons, D.B., 2012. Potential distribution of emerald ash borer: What can we learn from ecological niche models using Maxent and GARP?. Forest Ecol. Manag. 281, 23-31.

Stieglitz, M., Rind, D., Famiglietti, J., Rosenzweig, C., 1997. An efficient approach to modeling the topographic control of surface hydrology for regional and global climate modeling. J. Climate. 10(1), 118-137.

Stocker, T.F., Qin, D., Plattner, G.K., Tignor, M., Allen, S. K., Boschung, J., Midgley, P.M. 2013. Climate change 2013: the physical science basis. Intergovernmental panel on climate change, working group I contribution to the IPCC fifth assessment report (AR5). New York.

Stockwell, D., 1999. The GARP modelling system: problems and solutions to automated spatial prediction. International journal of geographical information science, 13(2), 143-158.

Sut, N., Simsek, O., 2011. Comparison of regression tree data mining methods for prediction of mortality in head injury. Expert Syst. Appl. 38(12), 15534-15539.

Tehrany, M. S., Pradhan, B., Mansor, S., Ahmad, N., 2015. Flood susceptibility assessment using GIS-based support vector machine model with different kernel types. Catena. 125, 91-101.

Tehrany, M.S., Pradhan, B., Jebur, M.N., 2013. Spatial prediction of flood susceptible areas using rule based decision tree (DT) and a novel ensemble bivariate and multivariate statistical models in GIS. J. Hydrol. 504, 69-79.

Tehrany, M.S., Pradhan, B., Jebur, M.N., 2014. Flood susceptibility mapping using a novel ensemble weights-of-evidence and support vector machine models in GIS. J. Hydrol. 512, 332-343.

Terstriep, M.L., Stall, J.B., 1974. The Illinois urban drainage area simulator, ILLUDAS. Bulletin (Illinois State Water Survey) no. 58.

Thieken, A.H., Müller, M., Kreibich, H., Merz, B., 2005. Flood damage and influencing factors: New insights from the August 2002 flood in Germany. Water Resour. Res. 41(12). https://doi.org/10.1029/2005WR004177

Torabi Haghighi, A., Menberu, M.W., Darabi, H., Akanegbu, J., Kløve, B., 2018. Use of remote sensing to analyse peatland changes after drainage for peat extraction. Land Degradation & Development, 29(10), 3479-3488.

Ture, M., Kurt, I., Kurum, A.T., Ozdamar, K., 2005. Comparing classification techniques for predicting essential hypertension. Expert Syst. Appl. 29(3), 583-588.

Ture, M., Tokatli, F. and Kurt, I., 2009. Using Kaplan–Meier analysis together with decision tree methods (C&RT, CHAID, QUEST, C4. 5 and ID3) in determining recurrence-free survival of breast cancer patients. . Expert Syst. Appl. 36(2), 2017-2026.

UNDP (United Nations Development Program), 2004. Reducing Disaster Risk: A Challenge for Development. UNDP Bureau for Crisis Prevention and Recovery, New York, 146pp.

USDA, 1986. Urban Hydrology for Small Watersheds, Technical Release 55. USDA Natural Resources Conservation Service, Washington, DC.

Wang, Z., Lai, C., Chen, X., Yang, B., Zhao, S., Bai, X., 2015. Flood hazard risk assessment model based on random forest. J. Hydrol. 527, 1130-1141.

Wasko, C., Sharma, A., 2015. Steeper temporal distribution of rain intensity at higher temperatures within Australian storms. Nat. Geosci. 8(7), 527-529.

Wiles, J. J., Levine, N. S., 2002. A combined GIS and HEC model for the analysis of the effect of urbanization on flooding; the Swan Creek watershed, Ohio. Environ. Eng. Geosci. 8(1), 47-61.

Woodward, M., Kapelan, Z. and Gouldby, B., 2014. Adaptive flood risk management under climate change uncertainty using real options and optimization. Risk Analysis. 34(1), 75-92.

Wu, W.W. and Lee, Y.T., 2007. Developing global managers' competencies using the fuzzy DEMATEL method. . Expert Syst. Appl. 32(2), 499-507.

Yesilnacar, E.K., 2005. The application of computational intelligence to landslide susceptibility mapping in Turkey. University of Melbourne, Department, 200.

Yin, H., Li, C., 2001. Human impact on floods and flood disasters on the Yangtze River. Geomorphology. 41(2), 105-109.

Zali, N., Ganji, R., Hosseine Amiri. H., 2016. Population Balance Planning for Metropolitan Area Network (MAN) in North of Iran in the 1400 Horizon. J. Geog. Eng. Terr. 1 (1), 54-71.

Zerger, A., 2002. Examining GIS decision utility for natural hazard risk modelling. Environ. Modell. Softw. 17(3), 287-294.

Zerger, A., Smith, D. I., Hunter, G. J., Jones, S. D. 2002. Riding the storm: a comparison of uncertainty modelling techniques for storm surge risk management. Appl. Geogr. 22(3), 307-330.

Zhan, X., Huang, M.L., 2004. ArcCN-Runoff: an ArcGIS tool for generating curve number and runoff maps. Environ. Modell. Softw. 19(10), 875-879.

Zhu, L., Sun, O. J., Sang, W., Li, Z., Ma, K., 2007. Predicting the spatial distribution of an invasive plant species (Eupatorium adenophorum) in China. Landscape Ecol. 22(8), 1143-1154.

Table 1 Linguistic scale and corresponding Fuzzy values (Sajedi-Hosseini et al., 2018)

Fuzzy DEMATEL		FANP		
Linguistic term	TFN	Linguistic term	TFN	Reciprocal of TFN
No effect	(0,0,0.25)	Equally important (EI)	(1, 1, 1)	(1, 1, 1)
Very low effect	(0,0.25,0.5)	Weekly Important (WI)	(1, 3, 5)	(1/5, 1/3, 1)
Low effect	(0.25,0.5,0.75)	Strongly Important (SI)	(3, 5, 7)	(1/7, 1/5, 1/3)
High effect	(0.5,0.75,1)	Very important (VI)	(5, 7, 9)	(1/9, 1/7, 1/5)
Very high effect	(0.75,1,1)	Absolutely important (AI)	(7, 9, 9)	(1/9, 1/9, 1/7)

TFN: Triangular fuzzy number

640 **Table 2** Model performance based on area under the receiver-operator characteristic curve (AUC-
641 ROC) and the Kappa statistic

Models	AUC-ROC	Kappa
GARP	0.935 (93.5%)	0.864
QUEST	0.892 (89.2%)	0.793
Difference between areas	0.0436 (4.36%)	-
z statistic	1.390	-

642

643

644 **Table 3** Limit super-matrix of conditioning factors of urban flood vulnerability

Criteria	Class	Weights	Criteria	Class	Weights
Socio-economic conditions	A	0.0147	Quality of buildings	Very high	0.0271
	B	0.0234		High	0.0298
	C	0.0291		Medium	0.0321
	D	0.0350		Low	0.0347
	E	0.0392		Very low	0.0364
	Natural	0.0068		No building	0.0252
Population density	High	0.1343	Age of buildings	Very old	0.0470
	Medium	0.1089		old	0.0394
	Low	0.0741		Medium	0.0350
	Very low	0.0530		New	0.0224
Urban density	High	0.0482		Newest	0.0022
	Medium	0.0414		No building	0.0021
	Low	0.0321			
	Very low	0.0265			

645

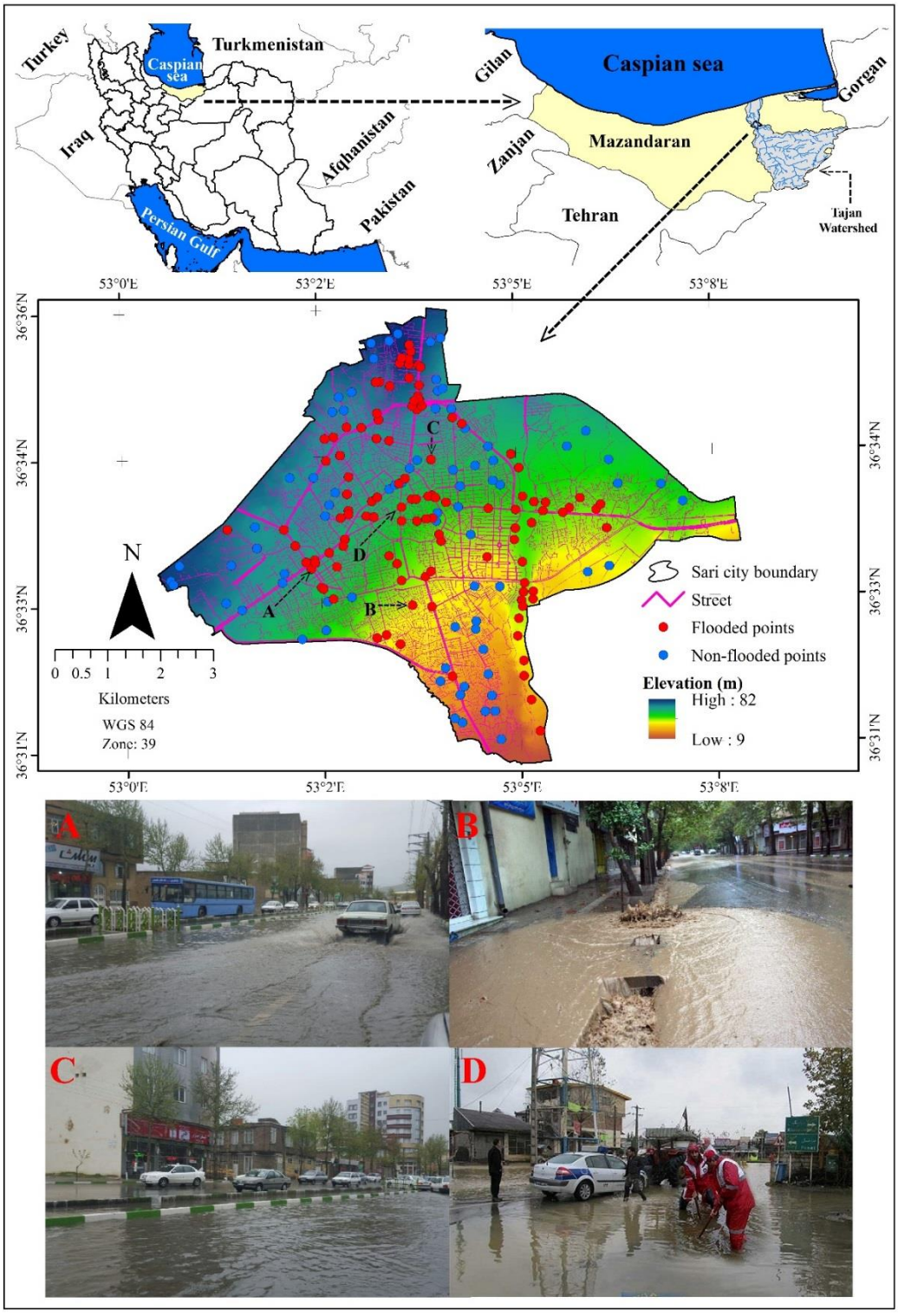
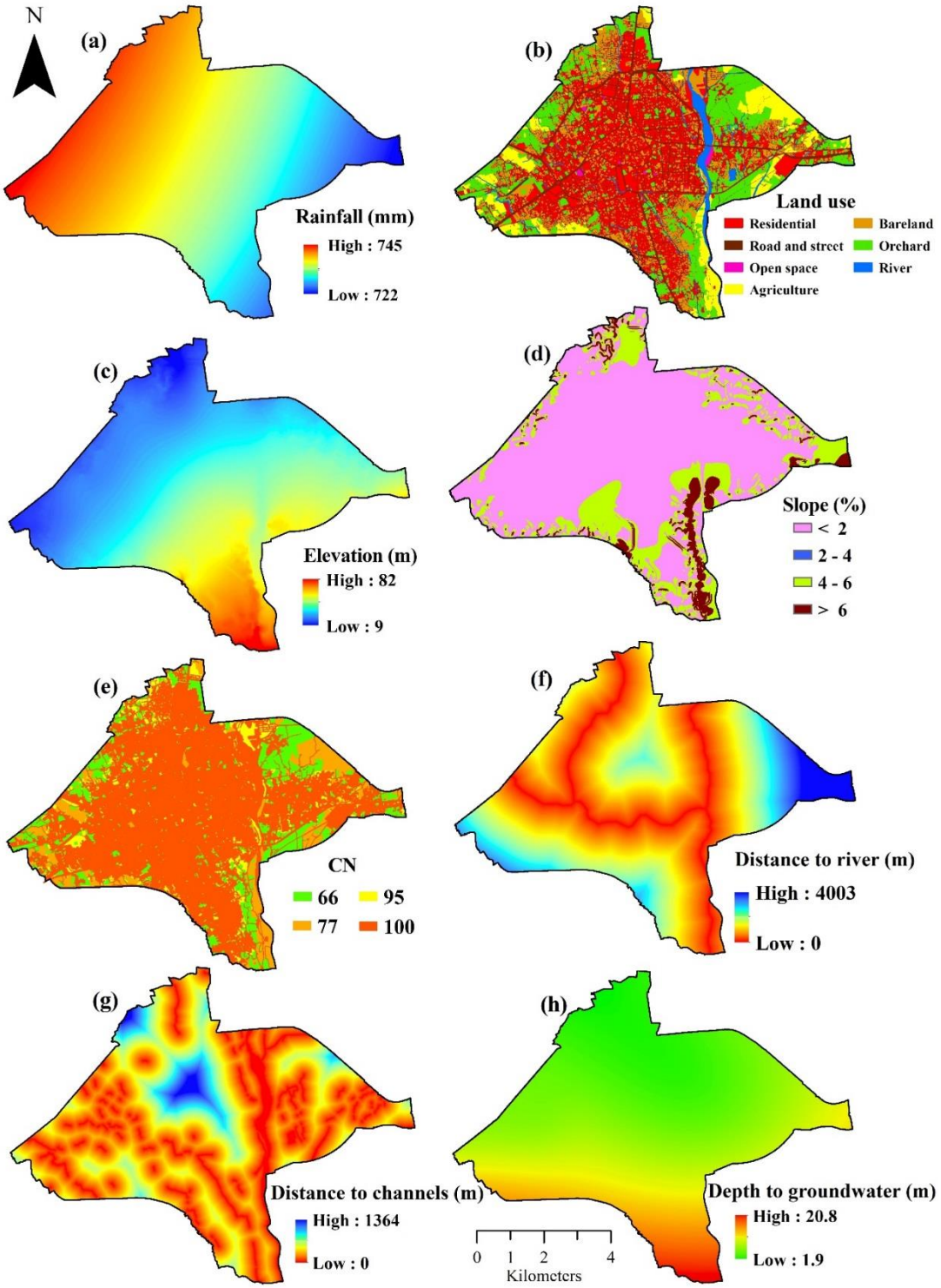


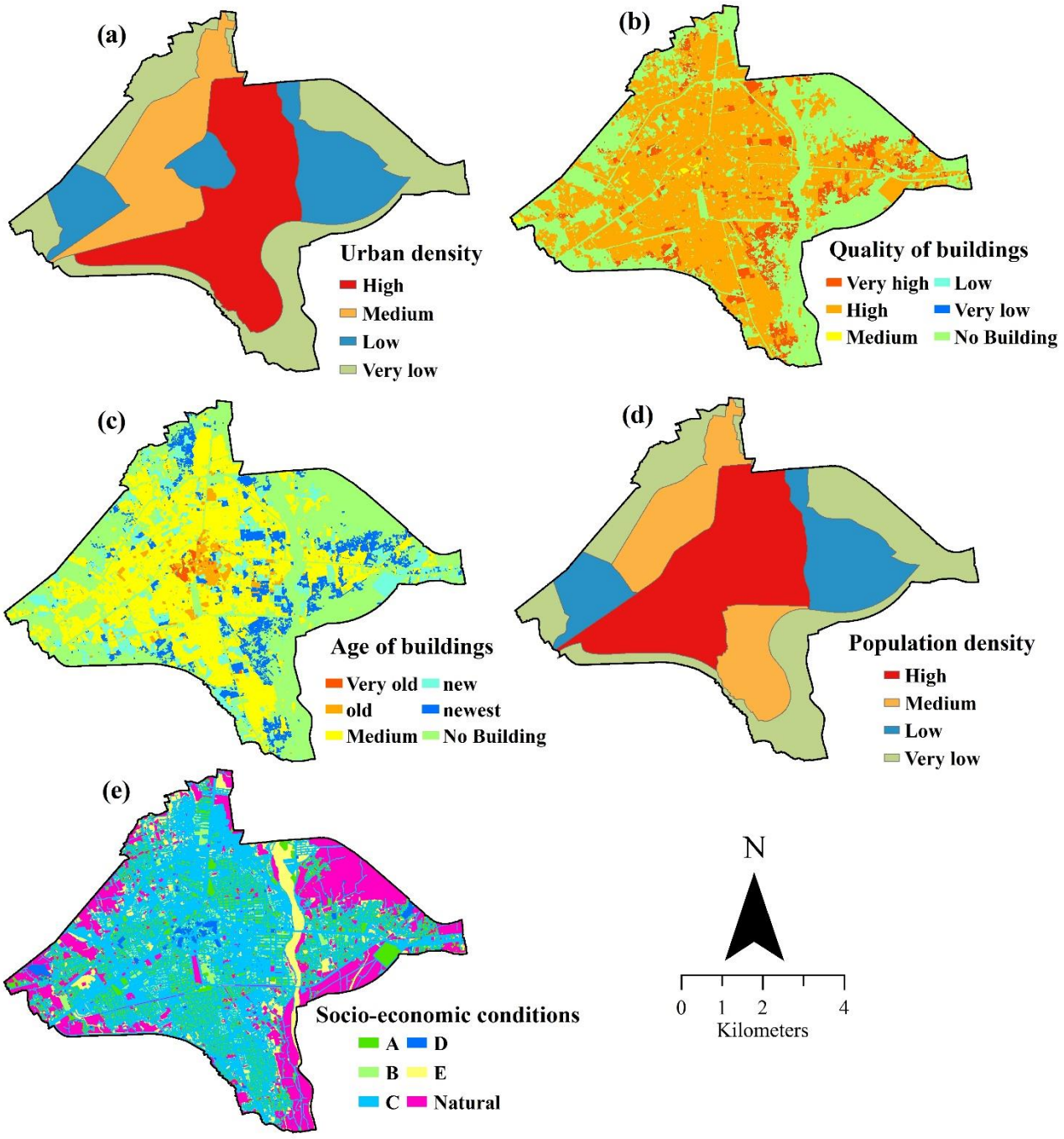
Fig. 1. Location of Sari city in Iran, and map of the city showing points with a history of flooding. A, B, C and D show previous flooding in the Jame Jam, Dokhaniat, Danesh, and Modarres districts, respectively.



649

650 **Fig. 2.** Conditioning factors of urban flood hazard: a) Rainfall, b) land use, c) elevation, d) slope, e) curve
651 number (CN), f) distance to river, g) distance to channel, and h) depth to groundwater.

652



653

654 **Fig. 3.** Conditioning factors of urban flood vulnerability: a) Urban density, b) quality of buildings, c) age
655 of buildings, d) population density, and e) socio-economic conditions.

656

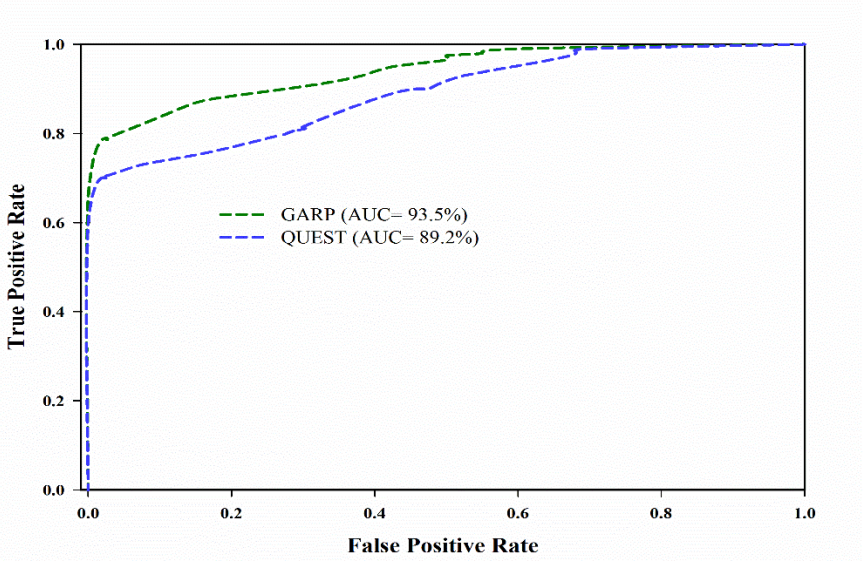


Fig. 4. Validation of the GARP and QUEST models using the area under the receiver-operator characteristic curve (AUC-ROC) method.

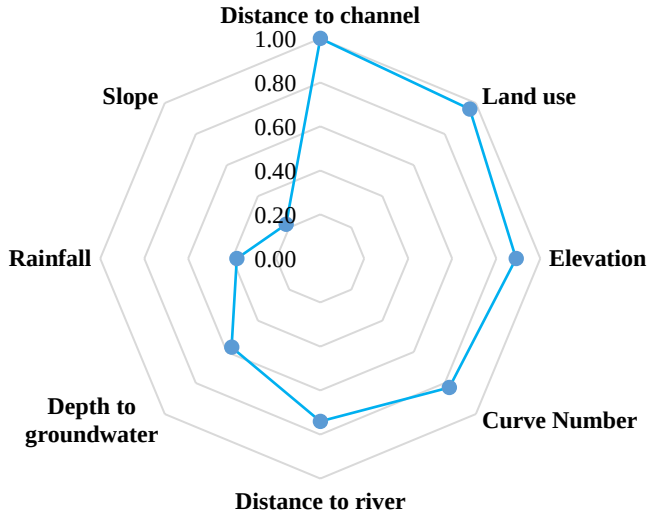
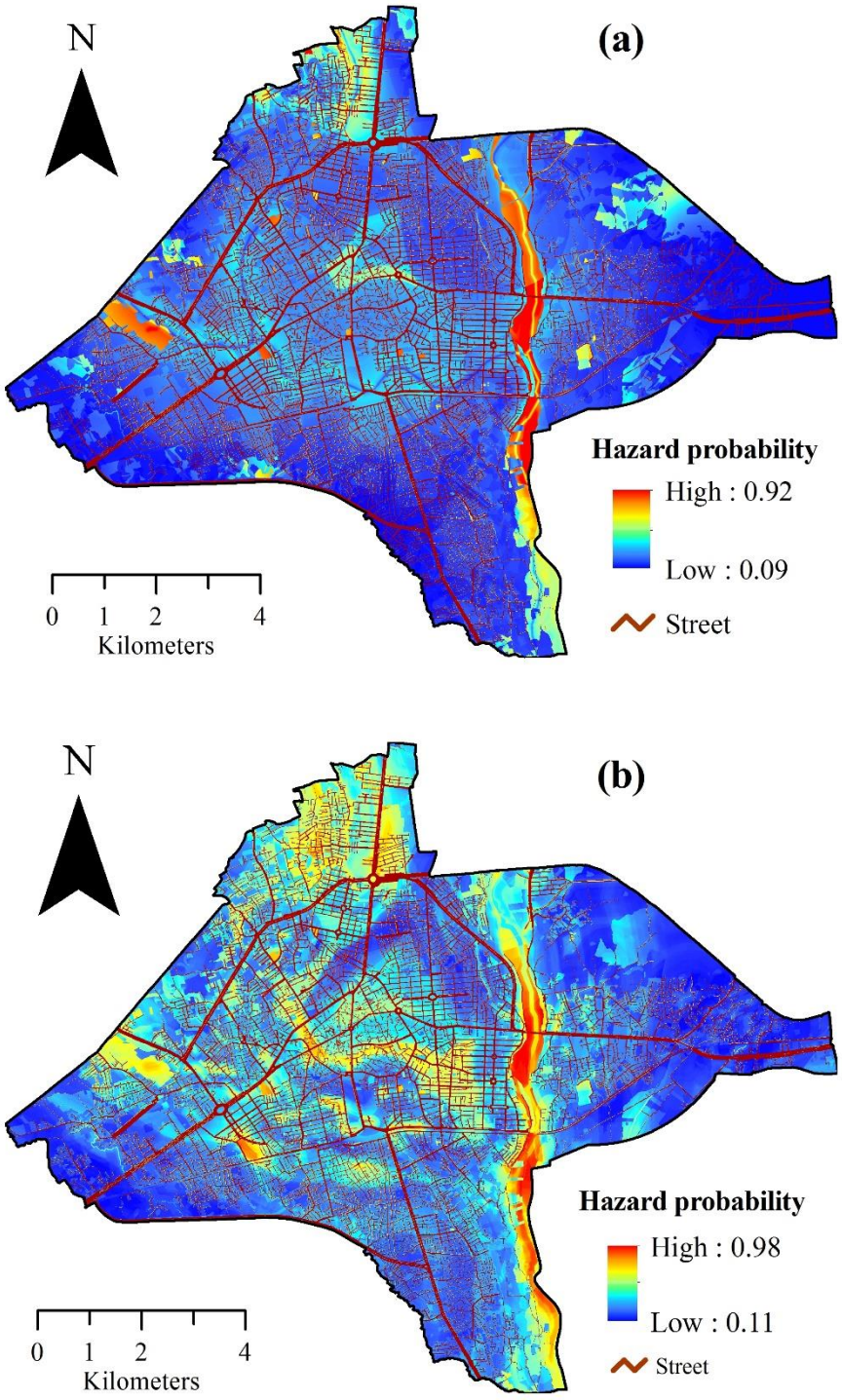


Fig. 5. Importance of conditioning factors (ICF) in urban flood hazard, based on the GARP model.



665

666 **Fig. 6.** Probability maps of urban flood hazard obtained using: a) the GARP model and b) the QUEST

667 model.

668

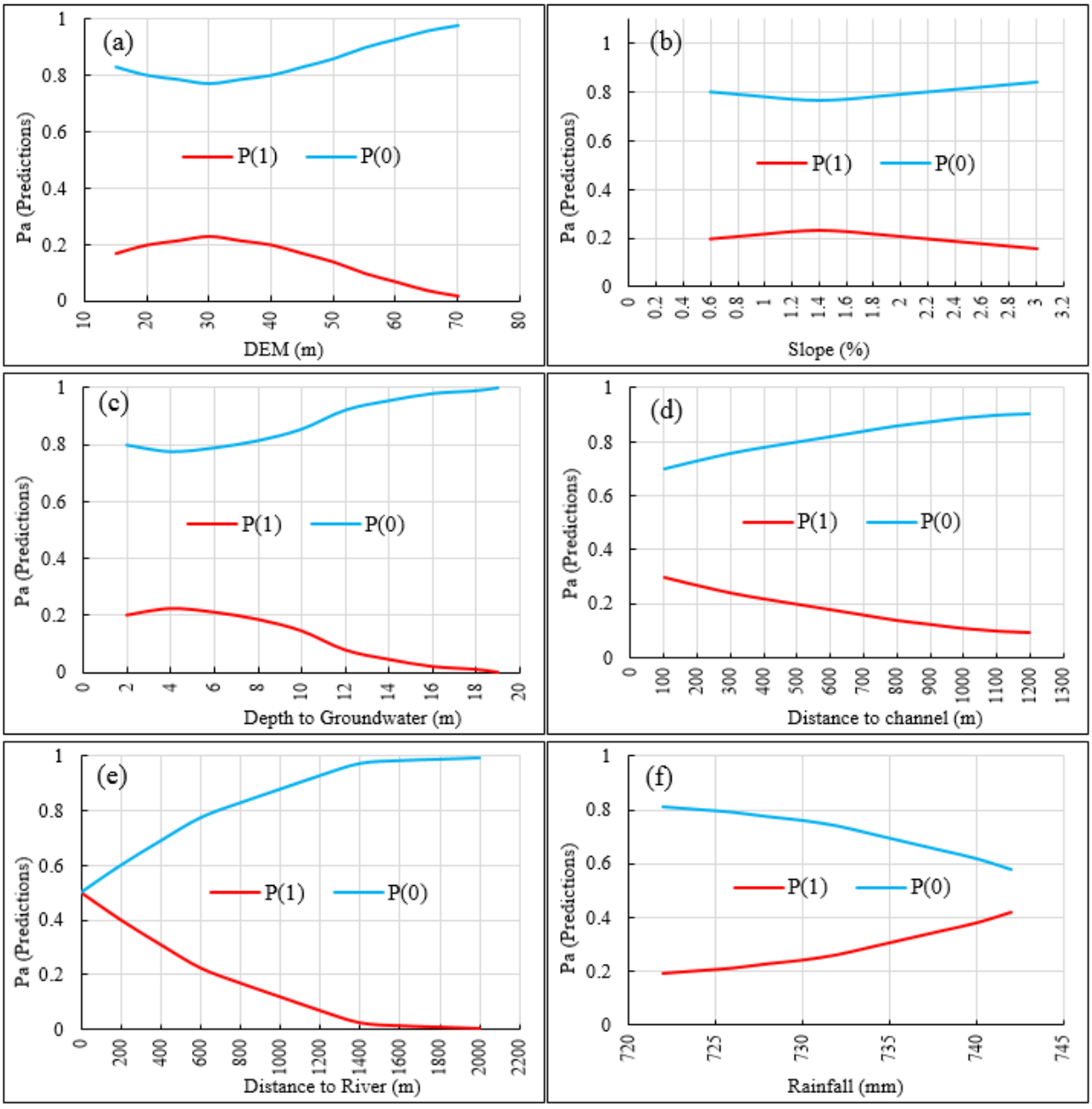


Fig. 7. Probability charts of flood hazard related to conditioning factors: a) elevation, b) slope, c) depth to groundwater, d) distance to channel, d) distance to river, and e) rainfall. — P(1): probability of occurrence, — P(0): probability of non-occurrence.

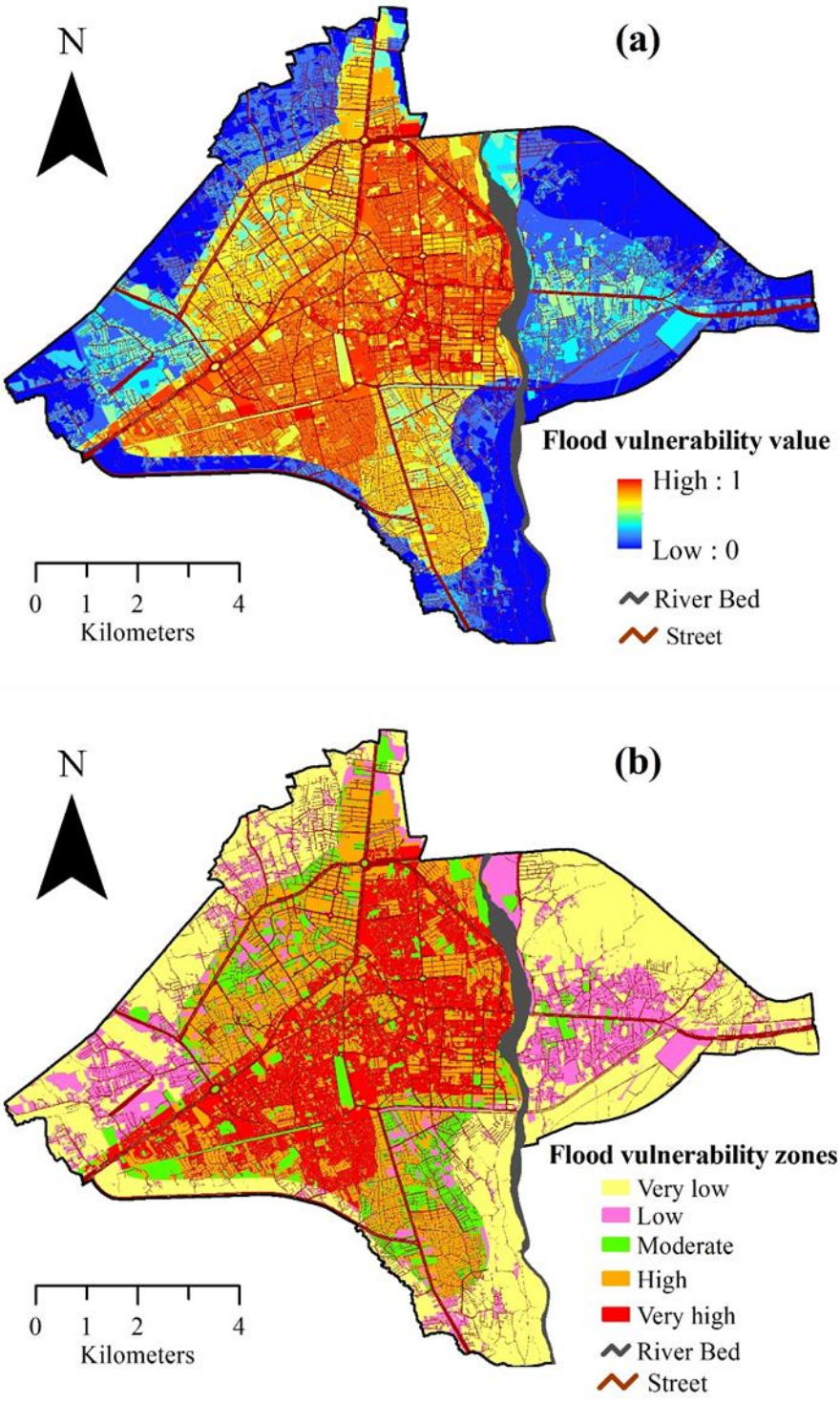
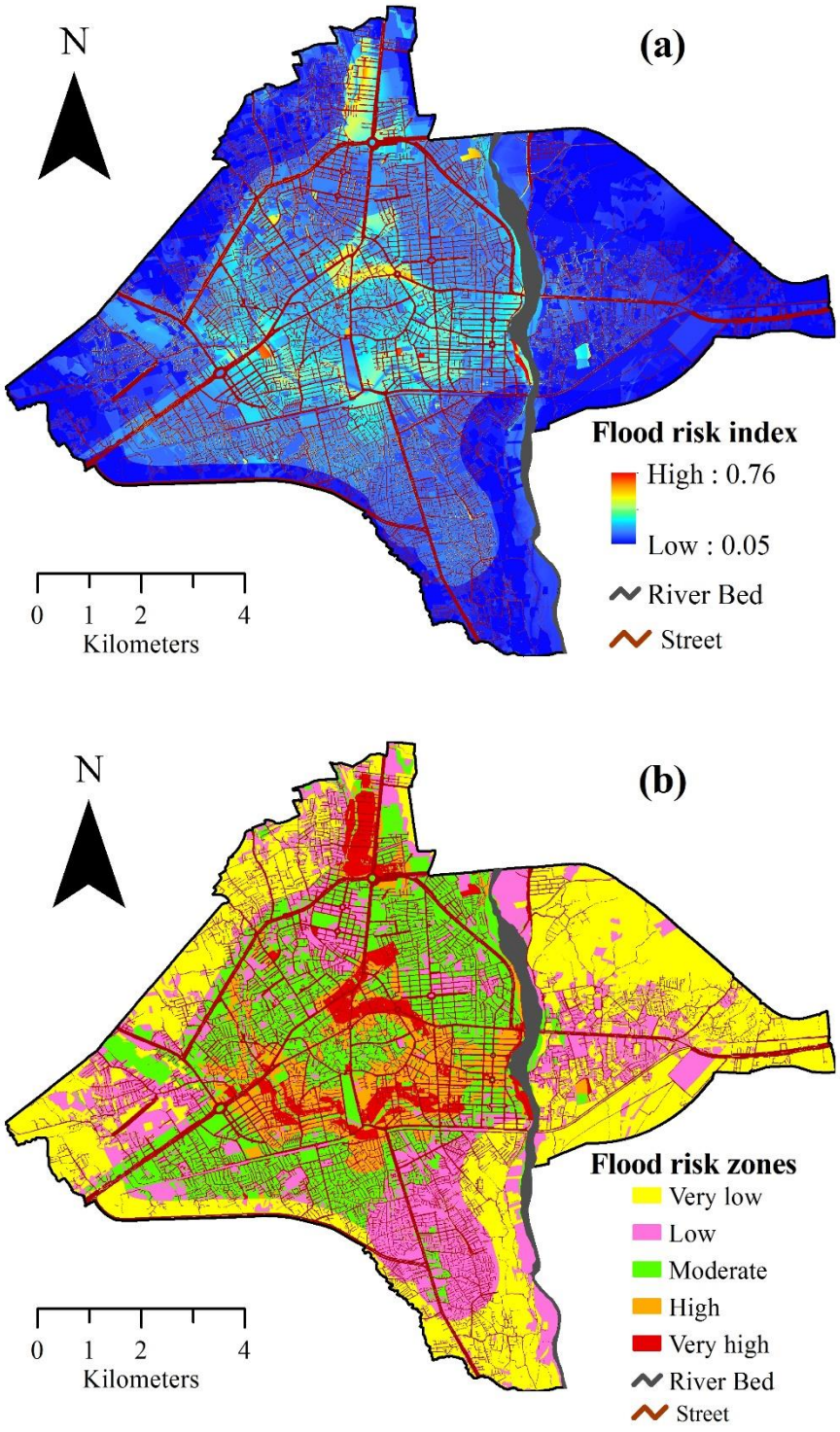


Fig. 8. Urban flood vulnerability maps. a) Flood vulnerability values and b) flood vulnerability zones.



683

684 **Fig. 9.** Urban flood risk maps. a) Urban flood risk index and b) urban flood risk zones.

685

686