

Flood and Flash Flood Hazard Mapping Using the Frequency Ratio, Multilayer Perceptron and Their Hybrid Ensemble

Mihnea Cristian Popa ^{1,2}, Daniel Constantin Diaconu ^{1,3,*}

¹ Centre for Integrated Analysis and Territorial Management, University of Bucharest, Bucharest, 010041, Romania; popamihneacristian@gmail.com (M.C.P.)

² Simion Mehedinți “Nature and Sustainable Development” Doctoral School, University of Bucharest, Bucharest, 010041, Romania

³ Department of Meteorology and Hydrology, Faculty of Geography, University of Bucharest, Bucharest, 010041, Romania

* Correspondence: daniel.diaconu@unibuc.ro; Tel.: +40-745-965-302

Abstract: The importance of identifying the areas vulnerable for both floods and flash-floods is an important component of risk management. The assessment of vulnerable areas is a major challenge in the scientific world. Adaptation and mitigation have generally been treated as two separate issues, both in public politics and in practice, in which mitigation is seen as the attenuation of the cause, and studies of adaption look into dealing with the consequences of climate change [1]. Studies on the impact of climate change on flood risk are mostly conducted at the river basin or regional scale [2,3]. Remote sensing and GIS technologies, together with the latest modelling techniques, can contribute to our ability to predict and manage floods [4-6]. Various methods are commonly used to map flood sensitivity. Recent methods such as multicriteria evaluation [7], decision tree analysis (DT) [8], fuzzy theory [9,10], weight of samples (WoE) [11], artificial neural networks (ANN) [12-14], frequency ratio (FR) [15] and logistic regression (LR) approaches [16] have been widely used by many researchers.

Keywords: Flash-Flood Potential Index; Multilayer Perceptron Neural Networks; hybrid model

The aim of this study is to provide a methodology-oriented study of how to identify the areas vulnerable to floods and flash-floods in the catchment by computing two indices: the Flash-Flood Potential Index (FFPI) for the mountainous and the Sub-Carpathian areas, and the Flood Potential Index (FPI) for the low-altitude areas, using the frequency ratio (FR), a bivariate statistical model, the Multilayer Perceptron Neural Networks (MLP), and the ensemble model MLP-FR. A database containing historical flood locations and the areas with torrentiality was created and used to train and test the models. The resulting models were computed using GIS techniques, thus resulting the flood and flash-flood vulnerability maps. The use of the two indices represents a preliminary step in creating flood vulnerability maps, which could represent an important tool for local authorities and a support for flood risk management policies.

In order to compute the proposed models, we created a database which contains the historical flood locations from 1970–2012 and the locations of the areas affected by torrentiality. The areas affected by torrentiality were identified based on satellite imagery and from the RUSLE (Revised Universal Soil Loss Equation) model, which contains the areas where soil is affected by water erosion [17]. The database contains the locations of the historical floods, which were obtained from the National Administration “Romanian Waters” and the locations affected by torrentiality. According to previous studies which use Machine Learning techniques [18–23], the training and testing data were split in a 70% ratio for the training samples and 30% for the testing samples.

The selection of the flood and flash-flood conditioning variables represents a key step in running the proposed models. The present study proposes the use of 14 flood conditioning variables, used to compute the Flood Potential Index (FPI), and 13 flash-flood conditioning variables, used to compute the Flash-Flood Potential Index (FFPI). The 14 variables used for the FPI are as follows: slope,

elevation, hydrological soil groups (HSG), slope aspect, elevation above channel (EaC), distance from rivers (DfR), saturated hydraulic conductivity (SHC), land-use, drainage density (DD), plan curvature (PLC), Topographic Position Index (TPI), Topographic Wetness Index (TWI), multi-annual precipitations (MaP) and the Convergence Index (CI). The 13 variables used for the FFPI are as follows: slope, profile curvature (PC), HSG, slope aspect, slope length and steepness factor (L-S), Curve number (CN), CI, land-use, soil erodibility by water (SEW), DD, TPI, TWI, and MaP.

The training and testing samples also hold the values of the factors which overlap the locations with 1 and 0 and were extracted using the Extract Multi Values to Points tool in ArcGIS. The analysis of the samples was carried out in Microsoft Excel and in Weka 3.9 (open-source Machine Learning software). The resulting values were computed using ArcGIS, thus resulting in the hazard maps.

The frequency ratio (*FR*) model represents a bivariate statistical method which is widely used in research for landslide and flood prediction mapping [15,20,24-27]. The frequency ratio is a probabilistic model. It is simple, easy to understand and apply, and it aims to determine the ratio of the area in which the occurrence of a phenomenon is present in the study area and also the probability ratio of an occurrence to a non-occurrence for given attributes [64]. The *FR* method is based on the association of the flood and flash-flood conditioning variables and the locations of the historical floods or areas affected by torrentiality.

The Multilayer Perceptron is an artificial neural network (ANN) used in function approximation and pattern recognition and is made up of three components [28]. Artificial neural networks represent a simple way to mimic the neural system of the human brain, in which, through various samples – in this case, the training samples – one can recognize data which were previously unseen, and make decisions and solve problems regarding the spatial relationship/association between input variables and the presence or absence of a certain phenomenon [29-31]. An MLP is based on the backpropagation algorithm – a supervised learning technique [28,32]. The neurons, represented by the variables/factors used in the analysis, are known as “input layers” and are connected to the “hidden layers” through a neural connection which holds the weights of the hidden layers.

The first steps in computing the *FR* model for both the FPI and FFPI was to determine the positive ratio (*Ratio +*) and the prediction ratio (*PR*). These tasks were carried out in Microsoft Excel. After obtaining the values, the next step consisted of reclassifying each flood or flash-flood conditioning variable based on the *Ratio +* values, computing them using the raster calculator in ArcGIS, and multiplying them with the *PR* values.

The FPI-*FR* and FFPI-*FR* models were classified into four hazard classes by using the Natural Breaks classification method in ArcGIS, as follows: low, average, high and very high. The resulting hazard map for the FPI-*FR* model shows that the hazard class.

The computation of the MLP model consisted in training the neural network with the flood locations for the FPI and with the torrential areas for the FFPI. This task was performed in Weka. The MLP model for each index was trained using 1000 maximum training epochs and 30 validation thresholds.

The resulting weights of each variable were used to compute both indexes. The variable importance for the Multilayer Perceptron was determined using the sensitivity analysis, which generates/computes the importance of the factors used in the neural network. The sensitivity analysis is generated automatically by the software after each run. The indices were computed in the same manner as for the *FR* model. As for the *FR* model, the FPI and FFPI MLP model was classified into four hazard classes by using the Natural Breaks classification method.

The role of hybrid models is to develop more accurate methods and reduce the potential disadvantages of the more traditional methods. The MLP-*FR* hybrid model was classified, as with the previous models, into four hazard classes by using the Natural Breaks method.

The flood and flash-flood hazard map show the areas which overlap the high and highest values of both indices.

The performance evaluation of a model using the ROC (receiver operating characteristic) curve is a widely used method in research. The ROC curve represents a 2D plot which indicates the performance of a classifying system as the value of the discrimination cut-off is changed with respect

to the predictor variable. The AUC (area under the curve) model represents a way to evaluate the testing ability in order to discriminate the true values. The ROC and AUC curves are made up from the sensitivity and specificity axes [33].

The methodology developed in this study has been applied on the river catchment, known in Romania as one of the most affected catchments by these types of natural hazards. The methods used can be applied at a national level or on different river catchments, considering the increase in intensity of the climatic events and anthropic activities which nevertheless have a direct impact on the generation of floods and flash-floods.

References

1. Hennessey, R.; Pittman, J.; Morand, A.; Douglas, A. Co-benefits of integrating climate change adaptation and mitigation in the Canadian energy sector. *Energy Policy* **2017**, *111*, 214–221.
2. Dobler, C.; Bürger, G.; Stötter, J. Assessment of climate change impacts on flood hazard potential in the Alpine Lech watershed. *J. Hydrol.* **2012**, *460–461*, 29–39.
3. Zeleňáková, M.; Purcz, P.; Blišfan, P.; Vranayová, Z.; Hlavatá, H.; Diaconu, D.C.; Portela, M.M. Trends in Precipitation and Temperatures in Eastern Slovakia (1962–2014). *Water* **2018**, *10*, 727.
4. Pradhan, B. Flood susceptible mapping and risk area delineation using logistic regression, GIS and remote sensing. *J. Spat. Hydrol.* **2010**, *9*.
5. Giustarini, L.; Chini, M.; Hostache, R.; Pappenberger, F.; Matgen, P. Flood Hazard Mapping Combining Hydrodynamic Modeling and Multi Annual Remote Sensing data. *Remote Sens.* **2015**, *7*, 14200–14226.
6. Gigović, L.; Pamučar, D.; Bajić, Z.; Drobnjak, S. Application of GIS-Interval Rough AHP Methodology for Flood Hazard Mapping in Urban Areas. *Water* **2017**, *9*, 360.
7. Balogun, A.-L.; Matori, A.-N.; Hamid-Mosaku, A.I. A fuzzy multi-criteria decision support system for evaluating subsea oil pipeline routing criteria in East Malaysia. *Environ. Earth Sci.* **2015**, *74*, 4875–4884.
8. Tehrany, M.S.; Pradhan, B.; Jebur, M.N. 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.* **2013**, *504*, 69–79.
9. Mukerji Aditya; Chatterjee Chandranath; Raghuwanshi Narendra Singh Flood Forecasting Using ANN, Neuro-Fuzzy, and Neuro-GA Models. *J. Hydrol. Eng.* **2009**, *14*, 647–652.
10. Pulvirenti, L.; Pierdicca, N.; Chini, M.; Guerriero, L. An algorithm for operational flood mapping from Synthetic Aperture Radar (SAR) data using fuzzy logic. *Nat. Hazards Earth Syst. Sci.* **2011**, *11*, 529–540.
11. Tehrany, M.S.; Pradhan, B.; Jebur, M.N. Flood susceptibility mapping using a novel ensemble weights-of-evidence and support vector machine models in GIS. *J. Hydrol.* **2014**, *512*, 332–343.
12. Campolo, M.; Soldati, A.; Andreussi, P. Artificial neural network approach to flood forecasting in the River Arno. *Hydrol. Sci. J.* **2003**, *48*, 381–398.
13. An artificial neural network model for flood simulation using GIS: Johor River Basin, Malaysia | SpringerLink Available online: <https://link.springer.com/article/10.1007/s12665-011-1504-z> (accessed on Sep 21, 2019).
14. Tiwari, M.K.; Chatterjee, C. Uncertainty assessment and ensemble flood forecasting using bootstrap based artificial neural networks (BANNs). *J. Hydrol.* **2010**, *382*, 20–33.
15. Rahmati, O.; Pourghasemi, H.R.; Zeinivand, H. Flood susceptibility mapping using frequency ratio and weights-of-evidence models in the Golastan Province, Iran. *Geocarto Int.* **2016**, *31*, 42–70.
16. Nandi, A.; Mandal, A.; Wilson, M.; Smith, D. Flood hazard mapping in Jamaica using principal component analysis and logistic regression. *Environ. Earth Sci.* **2016**, *75*, 465.
17. Panagos, P.; Borrelli, P.; Poesen, J.; Ballabio, C.; Lugato, E.; Meusburger, K.; Montanarella, L.; Alewell, C. The new assessment of soil loss by water erosion in Europe. *Environ. Sci. Policy* **2015**, *54*, 438–447.
18. Batista, G.; Prati, R.; Monard, M.-C. A Study of the Behavior of Several Methods for Balancing machine Learning Training Data. *SIGKDD Explor.* **2004**, *6*, 20–29.
19. Shirzadi, A.; Shahabi, H.; Chapi, K.; Tien Bui, D.; Pham, B.; Shahedi, K.; Ahmad, B.B. A comparative study between popular statistical and machine learning methods for simulating volume of landslides. *Catena* **2017**, *157*, 213–226.
20. Cao, C.; Xu, P.; Wang, Y.; Chen, J.; Zheng, L.; Niu, C. Flash Flood Hazard Susceptibility Mapping Using Frequency Ratio and Statistical Index Methods in Coalmine Subsidence Areas. *Sustainability* **2016**, *8*, 948.

21. Li, X.; Yan, D.; Wang, K.; Weng, B.; Qin, T.; Liu, S. Flood Risk Assessment of Global Watersheds Based on Multiple Machine Learning Models. *Water* **2019**, *11*, 1654.
22. Mardani, M.; Mardani, H.; De Simone, L.; Varas, S.; Kita, N.; Saito, T. Integration of Machine Learning and Open Access Geospatial Data for Land Cover Mapping. *Remote Sens.* **2019**, *11*, 1907.
23. He, Q.; Xu, Z.; Li, S.; Li, R.; Zhang, S.; Wang, N.; Pham, B.T.; Chen, W. Novel Entropy and Rotation Forest-Based Credal Decision Tree Classifier for Landslide Susceptibility Modeling. *Entropy* **2019**, *21*, 106.
24. Khan, H.; Shafique, M.; Khan, M.A.; Bacha, M.A.; Shah, S.U.; Calligaris, C. Landslide susceptibility assessment using Frequency Ratio, a case study of northern Pakistan. *Egypt. J. Remote Sens. Space Sci.* **2019**, *22*, 11–24.
25. Kumar Samanta, R.; Bhunia, G.; Shit, P.; Pourghasemi, H.R. Flood susceptibility mapping using geospatial frequency ratio technique: a case study of Subarnarekha River Basin, India. *Model. Earth Syst. Environ.* **2018**.
26. Cao, C.; Xu, P.; Wang, Y.; Chen, J.; Zheng, L.; Niu, C. Flash Flood Hazard Susceptibility Mapping Using Frequency Ratio and Statistical Index Methods in Coalmine Subsidence Areas. *Sustainability* **2016**, *8*, 948.
27. Yalcin, A.; Reis, S.; Aydinoglu, A.C.; Yomralioglu, T. A GIS-based comparative study of frequency ratio, analytical hierarchy process, bivariate statistics and logistics regression methods for landslide susceptibility mapping in Trabzon, NE Turkey. *CATENA* **2011**, *85*, 274–287.
28. Peponi, A.; Morgado, P.; Trindade, J. Combining Artificial Neural Networks and GIS Fundamentals for Coastal Erosion Prediction Modeling. *Sustainability* **2019**, *11*, 975.
29. Gómez, H.; Kavzoglu, T. Assessment of shallow landslide susceptibility using artificial neural networks in Jabonosa River Basin, Venezuela. *Eng. Geol.* **2005**, *78*, 11–27.
30. Shiruru, K. An Introduction to Artificial Neural Network. *Int. J. Adv. Res. Innov. Ideas Educ.* **2016**, *1*, 27–30.
31. Taravat, A.; Rajaei, M.; Emadodin, I.; Hasheminejad, H.; Mousavian, R.; Biniyaz, E. A Spaceborne Multisensory, Multitemporal Approach to Monitor Water Level and Storage Variations of Lakes. *Water* **2016**, *8*, 478.
32. Sánchez-Reolid, R.; García, A.S.; Vicente-Querol, M.A.; Fernández-Aguilar, L.; López, M.T.; Fernández-Caballero, A.; González, P. Artificial Neural Networks to Assess Emotional States from Brain-Computer Interface. *Electronics* **2018**, *7*, 384.
33. Yang, S.; Berdine, G. The receiver operating characteristic (ROC) curve. *Southwest Respir. Crit. Care Chron.* **2017**, *5*, 34–36.



© 2019 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons by Attribution (CC-BY) license (<http://creativecommons.org/licenses/by/4.0/>).