Proc. of the 8th International Symposium on Geotechnical Safety and Risk (ISGSR)
Edited by Jinsong Huang, D.V. Griffiths, Shui-Hua Jiang, Anna Giacomini and Richard Kelly
 Published by Research Publishing, Singapore.
doi: 10.3850/978-981-18-5182-7_11-001-cd

Uncertainty Quantification of Landslide Susceptibility Mapping Considering Landslide Boundary Geometry

Sahand Khabiri[1] and Yichuan Zhu*[2]

[1]Department of Civil and Environmental Engineering, Temple University, Philadelphia, USA.
E-mail: sahandkhabiri@temple.edu
[2]Department of Civil and Environmental Engineering, Temple University, Philadelphia, USA.
E-mail: yichuan.zhu@temple.edu

Abstract: Landslide Susceptibility Mapping (LSM) plays an important role in identifying, characterizing, and managing landslide disasters. The uncertainty of LSM is known associated with several geo-variables, including geology, land use, topography, vegetation cover, among others. However, the effect of the varying geometry of mapped landslide boundaries on the accurate characterization of landslide events has not been fully investigated in the literature. In this study, we performed a sensitivity test of a landslide susceptibility model, based on the varying complexity of mapping boundaries, including triangular, circle, square, and irregular shapes of mapping polygons. We trained a Bayesian network model to evaluate the posterior landslide probability for the Western Oregon area using model variables including slope, elevation, curvature, relief, land use/cover, soil types, and precipitation, and the model data consists of landslides inventories, established according to different complexity levels of mapping polygons. Results show that the circular and irregular polygons generated model predictions with higher precision and accuracy compared to those associated with triangular and square polygon shapes. The results of this analysis can serve as quantitative guidance for the faithful characterization of future landslide events and set the basis for uncertainty quantification of other participating sources in landslide susceptibility analysis.

Keywords: Landslide assessment; landslide susceptibility mapping; Bayesian network; uncertainty quantification.

1 Introduction

Landslides occur frequently throughout theUnited States that until 2020, several individual landslides inventories among 10 states have shown a total of 310,392 landslides, in which 57,975 have occurred in Oregon state that is represented as one of the most impacted areas in terms of both high frequency and severity of mapped landslides (Burns and Madin 2009;Mirus et al. 2020).

To identify the landslide prone areas, scholars have developed methods to assess the spatial probability of landslides occurrence that often recognized as Landslide Susceptibility Mapping (LSM)(Varnes 1984; Guzzetti et al. 1999). Methods of LSM include geomorphological mapping (Van Westen et al. 2003), heuristic modeling (Ruff and Czurda 2008), statistical correlating (Hong et al. 2017; Wang et al. 2016), machine learning techniques (Ayalew and Yamagishi 2005; Marjanović et al. 2011), physically-based methods (Chen et al. 2015), among others. One of the machine learning methods is the Bayesian Network (BN) that can represent the conditional dependence samong participating variables through the probabilistic graphic modeling. The ability of BN in representing uncertain knowledge and its capability in updating models in light of incoming evidence attracted scholars for constructing LSM models (Song et al. 2012; Oommen et al. 2018; Wang et al. 2019; Lee et al. 2020; Khalaj et al. 2020).

Similar to any modeling activities, however, there is a certain amount of uncertainty inherent in every LSM model, which can lead to the varying degree of accuracy and precision of model performance (Feizizadeh, and Blaschke 2014; Pourghasemi et al. 2020a). The type of uncertainty can be both aleatoric and epistemic, presented as incompleteness of data populations, and limitations in the techniques for developing the susceptibility map (Ardizzone et al., 2002). One source of uncertainty can be originated from different geometries of polygons used for mapping the landslides. Specifically, Huang et al. (2022)have recently investigated the effectiveness of landslide boundaries to measure the uncertainty of LSM by considering three forms of mapping geometrics comprised of point, circle, and irregular shapes. Results show that landslide mapping by points can generate corresponding results having a higher uncertainty and lower accuracy than those of polygon shapes.

In this study, we assess the uncertainty of posterior probabilities of landslide occurrence by using BN method. The effects of varying geometry complexity of mapping boundaries on the model estimation have been investigated. Results of this analysis can serve as quantitative guidance for the faithful characterization of future landslide events and set the basis for uncertainty quantification of other participating sources in landslide susceptibility analysis.

2 Study Area and Dataset

The west side of Oregon, close to the coastal region, is selected as the study area that situates between a longitude of121°80′Wto 124°42′W and a latitude of 42°N to 46°24′N, encompassing an area of 48672 km^2 (Fig. 1). The Statewide Landslide Information Database for Oregon (SLIDO) is a geo referenced openly available database from which we collected the data of landslide events. We first selected 2,781 landslides characterized by slopes between 20° and 40°, landslide movement of slide or combined slide and rotation, and identified with high confidence in terms of landslide occurrence. We then exclude parts of the datasets in which landslide occurred close to the border that resulted incompleteness of the datasets. Finally,359 landslides we reselected for the present study. The range of the area of the selected landslides is between 103 m^2 and 68888 m^2, with a mean value of 18408 m^2. For the negative samples used for training the BN model, we randomly selected the same amount of non-landslide points compared to landslide points as introduced above.

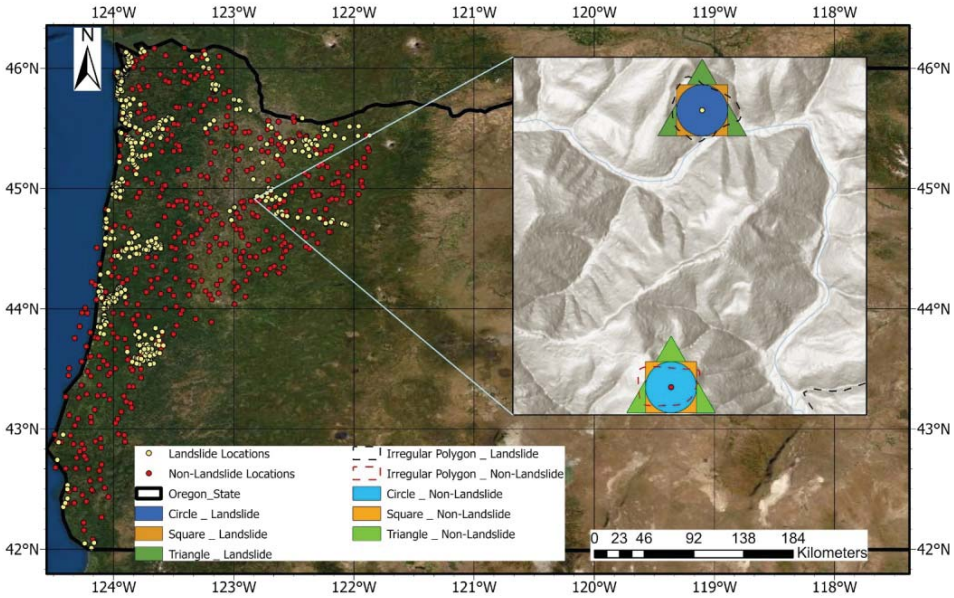

Figure1. Study area including landslide inventory map and mapped landslide and Non-Landslide boundaries: Landslide centroid (yellow points), Non-landslide centroid (red points), mapped irregular polygon landslide (black dashed line), circle landslide boundary (dark blue), square landslide boundary (dark orange), triangle landslide boundary (dark green),irregular polygon Non-landslide (red dashed line), circle landslide boundary (light blue), square landslide boundary (light orange), triangle landslide boundary (light green).

3 Methodology

The pre-processing of the datasets consists of drawing landslide boundaries, identifying, and categorizing attributes that can impact the landslide occurrence. The mean value of all mapped landslide areas equals to 18408 m^{2}which was used as the benchmark to determine the buffer size of both regular and irregular polygons. For identifying the impact attributes, one should be noted that the landslide occurrence can be triggered by various geo-environmental factors, such as topography, geomorphology, hydrology, geology, surface cover, as well as external induced factors such as earthquake and rainfall (Guzzetti et al. 1999; Huang et al. 2021). In this study, we have selected eight attributes where elevation, curvature, slope, roughness index, and relief amplitude were extracted from Digital Elevation Model (DEM) with a spatial resolution of 10m, the land-use with a spatial resolution of 10m, and soil texture map together with annual rainfall with a spatial resolution of 0.5°. All continuous numerical attributes were converted to categorical classes, including elevation (30 classes) ranging from 4.5ft to 1443ftwith 48 ft interval scale, curvature (7 classes) ranging from -1.694 to 1.176 with 0.41 interval scale, slope (7 classes) ranging from 0.24° to 35.24°with 0.41 interval scale, relief (8 classes) ranging from 2 to 402 with50 interval scale, roughness (6 classes) ranging from 0.23 to 0.83 with0.1 interval scale, and rainfall (3 classes) ranging from 106.326 to 271.326 mm with 55 interval scale. Soil texture and land-use possess original categorical values.The presence/absence of landslide was used as target variables labeled by binary values 1 and 0, respectively. All available landslide datasets were randomly partitioned into 85% to 15% as training and test data, respectively.

The BN model used by this study is a graphical model defined as a pair of$\langle G, \theta \rangle$. The first component shows the nodes exist in the directed acyclic graph (DAG), while the second component shows the conditional interdependency and connection among the nodes. Bayesian network modeling consists of three steps: structure

learning, parameter learning, and Inference. Structure learning is the first step which estimates the structure of BN through methods such as score-based functions, constraint-based functions, or hybrid approaches (Beretta et al. 2018).Among the proposed BN structures, Naïve Bayes and Tree-Augmented Naive Bayes (TAN) are the two most widely used ones due to their simplicity and computational efficiency(Langley et al. 1992; Friedman et al. 1997). Eq. (1) and (2) present the functional definition of Naïve Bayes and Tree-Augmented Naive Bayes (TAN), respectively. Compared to Naïve Bayes method, TAN provides more flexibility that allows for inter-connections between child nodes. Thus, we used TAN as the BN structure for this study. In Eq. (2), $P_{TAN}(Y)$ is the prior probability of class node (parent node of all attributes) in TAN; $P_{TAN}(x_{root}|Y)$is the conditional probability of root node (class node is its only parent and all the edges outward from it) given class node in TAN; $P_{TAN}(x_i|Y, x_{Parent})$is the conditional probability of *i-th* attribute given its parent and class node in TAN.

$$P_{BN}(x_1, \dots, x_n, Y) = P_{BN}(Y) \prod_{i=1}^{n} P_{BN}(x_i|Y) \quad (1)$$

$$P_{TAN}(x_{root}, x_2 \dots, x_n, Y) = P_{TAN}(Y) P_{TAN}(x_{root}|Y) \prod_{i=2}^{n} P_{TAN}(x_i|Y, x_{Parent}) \quad (2)$$

4 Result and Discussion

4.1Posterior probability of landslide under different geometry shapes

In this study, we used Bayesian Dirichlet equivalent uniform (BDeu) score for structure learning, and Bayesian estimation for parameter learning for constructing the BN model (Heckerman et al. 1995; Lee et al. 1968). After the model was trained, the posterior conditional probability of landslide given each attribute was computed separately using Eq.(2).The results of posterior probability of landslide versus participating attribute categories are plotted on Figure 2. It shows the influence of geometry complexity of mapping boundaries as well as different classes of attributes on landslide occurrence. Elevation, curvature, land-use, rainfall, and soil textures are generally more important environmental factors, affecting the landslide occurrence due to higher variation of posterior probabilities through their corresponding classes. Figure 2also shows that the landslide occurrence is highly sensitive to: elevation when the elevation varies within the range of 4.5ft to 340.5ft(categories are from 1 to 7); curvature varies between -0.464 to -0.054(category is equal to 4); slope varies from 10.24° to 15.24°(category is equal to 3); lands where is covered with trees (category is equal 2);soil texture composed from sandy clay loam (category 2), and annual rainfall within the range of 216 to 271mm which corresponds to category 3. The two remaining attributes roughness and relief do not show a significant effect on landslide occurrence and the posterior probabilities fluctuate smoothly within the range of 0.45 to 0.55.

The effects of boundary complexity on posterior probabilities are significant for attributes elevation, curvature, slope, relief, roughness, and land-use (Fig. 2a-f), while not for attributes soil types and rainfall. It may be resulted from the latter two attributes are varied more uniform in space, at the same time were mapped with lower resolutions compared to others.

4.2 Model validation

For validation purposes, different statistical indices are considered to assess and compare the accuracy of different geometry types. ACC (Accuracy); TP (True Positive); FP (False Positive); FN (False Negative); TN (True Negative); SST (True Positive Ratio); 1 – SFP (False Positive Ratio); PPV (Positive Predictive Value); NPV (Negative Predictive Value); AUC (Area under the ROC curve) are computed to evaluate the reliability, uncertainty, and accuracy of different geometries. The comparison is conducted between the test and prediction datasets where the latter is generated from the trained BN model.

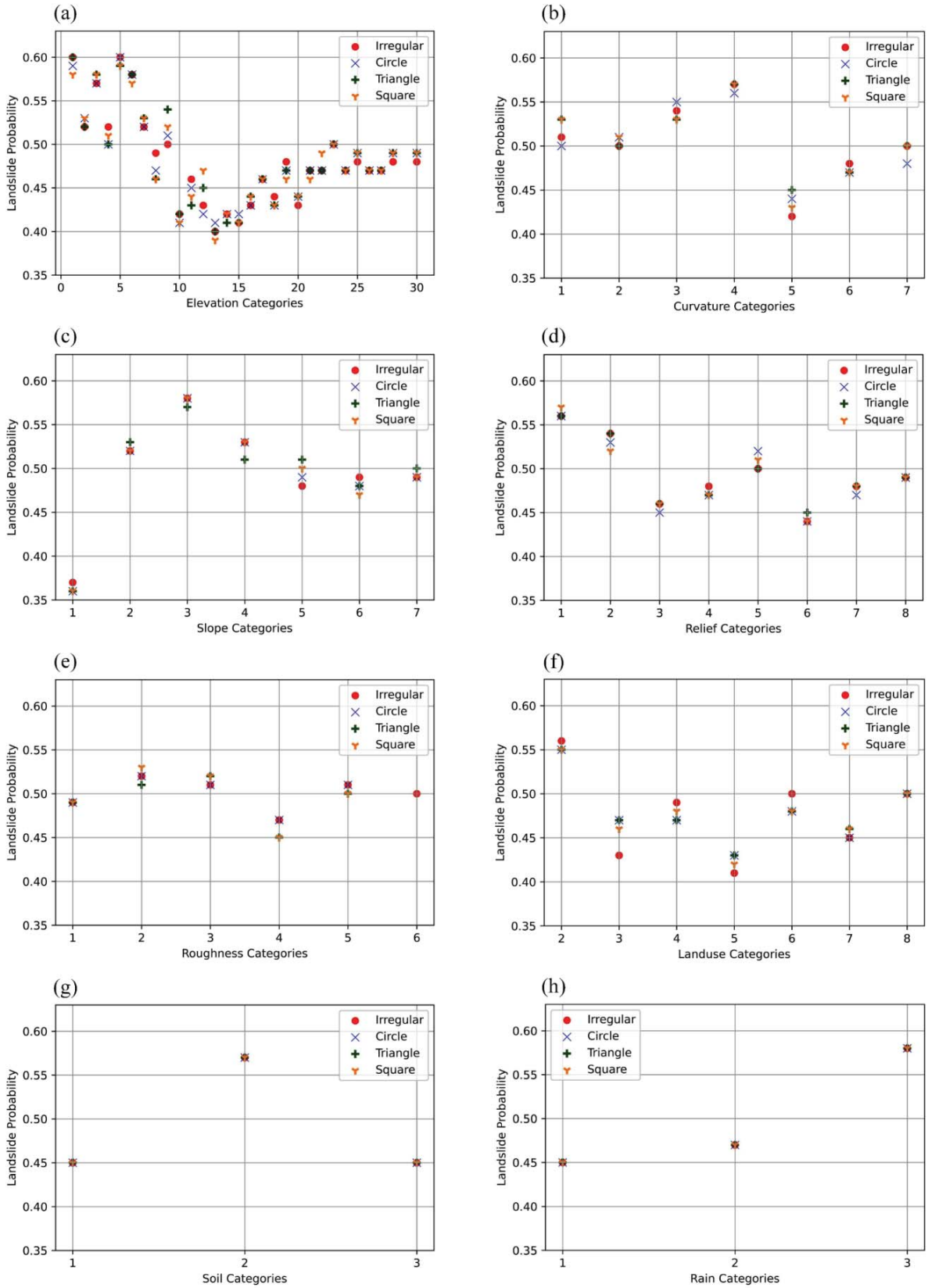

Figure2. Posterior probability of Landslide given (a) Elevation(b) Curvature (c) Slope (d) Relief (e) Roughness (f) Landuse (g) Soil texture (h) Annual Rainfall rate.

Table 1 summarizes the results according to four types of geometry shapes. In this study, AUC is used to determine the uncertainty and prediction performance of LSM, while the ACC, PPV, and NPV were calculated to evaluate the accuracy of LSM modeling under different condition. Kappa was computed to show the reliability of models. According to Table 1, a threshold equal to 50 percent is assumed as a benchmark to compare the results of all statistical indices except Kappa. AUC values computed for irregular and circle polygons are higher

than those of triangular and square polygons. This indicates the former are more reliable buffer shapes for BN model development.

Table1. Statistical indexes for model validation

Index	Threshold	Irregular Polygon	Circle	Triangle	Square
Kappa	10%	0.98902	0.98868	0.98902	0.98919
	50%	0.98645	0.98508	0.98593	0.98525
	90%	0.98748	0.98696	0.98748	0.98765
ACC	50%	0.7407	0.8148	0.7685	0.8055
TP	50%	46	49	46	48
FP	50%	20	15	17	15
FN	50%	8	5	8	6
TN	50%	34	39	37	39
SST	50%	0.85	0.90	0.85	0.88
1-SFP	50%	0.37	0.27	0.31	0.27
PPV	50%	0.69	0.76	0.73	0.76
NPV	50%	0.80	0.88	0.82	0.86
AUC	0-100%	0.867	0.87	0.83	0.84

Figure 3 presents the receiver operating characteristic (ROC) curves of four types of geometry shapes. The ROC curves depicttrue positive rate (sensitivity) versus false positive rate (1-specificity) and result with successively chosen thresholds started from 0% (1, 1) to 100% (0, 0). Curves close to the left corner are indicative of higher AUC values which further suggests higher accuracy in model performance. Results show that the irregular and circular polygons produced better prediction performance than those oftriangular and square polygons due to higher AUC index values as shown in Table 1.

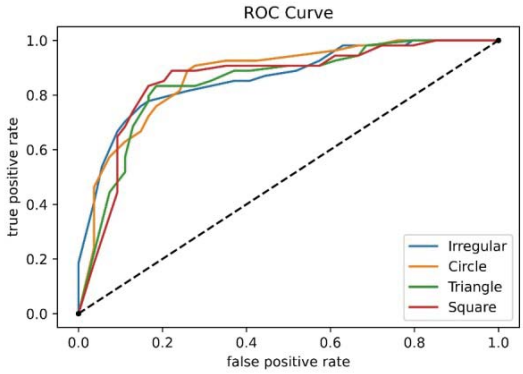

Figure3.ROC curves of irregular mapped landslide boundary (blue), circle mapped landslide boundary (orange), triangle mapped landslide boundary (green), square mapped landslide boundary (red)

5 Conclusion

In this study, we evaluated the effects of mapping boundaries on LSM by comparing the results yielded from BN model predictions. The main findings of this study can be summarized as the following:

1. The high influencing attributes contributing the LSM are elevation, curvature, slope, relief, roughness, and land-use, while soil types and rain fall attributes present insignificant impact on LSM development.

2. The occurrence of landslide is highly sensitive when elevation ranging from 4.5ft to 340.5ft, curvature ranging from -0.464 to -0.054, slope ranging from 10.24° to 15.24°; lands where is covered with trees, soil texture composed from sandy clay loam, and annual rainfall ranging from 216 to 271 mm.

3. Overall, results show that circular and irregular polygons generated higher accuracy and lower uncertainty compared to those associated with triangular and square polygon shapes.

References

Ardizzone, F., Cardinali, M., Carrara, A., Guzzetti, F., and Reichenbach, P. (2002). Uncertainty and errors in landslide mapping and landslide hazard assessment. *Natural Hazards and Earth System Sciences*, 2(1-2), 3-14.

Ayalew, L., and Yamagishi, H. (2005). The application of GIS-based logistic regression for landslide susceptibility mapping in the Kakuda-Yahiko Mountains, Central Japan. *Geomorphology*, 65(1-2), 15-31.

Beretta, S., Castelli, M., Gonçalves, I., Henriques, R., andRamazzotti, D. (2018). Learning the structure of Bayesian Networks: A quantitative assessment of the effect of different algorithmic schemes. *Complexity*, 1-12.

Burns WJ, andMadin, IP. (2009). Protocol for inventory mapping of landslide deposits from light detection and ranging (lidar) imagery. *DOGAMI Special Paper* 42, 36 pp.

Chen, H.X., Zhang, L.M., Gao, L., Zhu, H., and Zhang, S. (2015). Presenting regional shallow landslide movement on three-dimensional digital terrain. *Engineering Geol*ogy, 195, 122-134.

Feizizadeh, B., Jankowski, P., andBlaschke, T. (2014). A GIS based spatially-explicit sensitivity and uncertainty analysis approach for multi-criteria decision analysis. *Computers andGeosciences*, 64, 81-95.

Feizizadeh, B andBlaschke, T (2014) An uncertainty and sensitivity analysis approach for GIS-based multicriteria landslide susceptibility mapping. *International Journal of Geographical Information Science*, 28(3), 610-638.

Friedman, N., Geiger, D., andGoldszmidt, M. (1997). Bayesian network classifiers. *Machine learning*, 29(2), 131-163.

Guzzetti, F., Carrara, A., Cardinali, M., and Reichenbach, P., (1999). Landslide hazard evaluation: a review of current techniques and their application in a multi-scale study, Central Italy. *Geomorphology* 31(1), 181-216.

Heckerman, D., Geiger, D., and Chickering, D. M. (1995). Learning Bayesian networks: The combination of knowledge and statistical data. *Machine learning*, 20(3), 197-243.

Hong, H., Liu, J., Bui, D. T., Pradhan, B., Acharya, T. D., Pham, B. T., ... and Ahmad, B. B. (2018). Landslide susceptibility mapping using J48 Decision Tree with AdaBoost, Bagging and Rotation Forest ensembles in the Guangchang area (China). *Catena*, 163, 399-413.

Huang, F., Ye, Z., Jiang, S. H., Huang, J., Chang, Z., and Chen, J. (2021). Uncertainty study of landslide susceptibility prediction considering the different attribute interval numbers of environmental factors and different data-based models. *Catena*, 202, 105250.

Huang, F., Yan, J., Fan, X., Yao, C., Huang, J., Chen, W., and Hong, H. (2022). Uncertainty pattern in landslide susceptibility prediction modelling: Effects of different landslide boundaries and spatial shape expressions. *Geoscience Frontiers*, 13(2), 101317.

Khalaj, S., BahooToroody, F., Abaei, M. M., BahooToroody, A., De Carlo, F., andAbbassi, R. (2020). A methodology for uncertainty analysis of landslides triggered by an earthquake. *Computers and Geotechnics*, 117, 103262.

Kim, H. G., Lee, D. K., Park, C., Ahn, Y., Kil, S. H., Sung, S., andBiging, G. S. (2018). Estimating landslide susceptibility areas considering the uncertainty inherent in modeling methods. *Stochastic environmental research and risk assessment*, 32(11), 2987-3019.

Langley, P., Iba, W., and Thompson, K. (1992). An analysis of Bayesian classifiers. *Proceedings of the Tenth National Conference on Artificial Intelligence*,223-228.

Lee, T. C., Judge, G. G., and Zellner, A. (1968). Maximum likelihood and Bayesian estimation of transition probabilities. *Journal of the American Statistical Association*, 63(324), 1162-1179.

Lee, S., Lee, M. J., Jung, H. S., and Lee, S. (2020). Landslide susceptibility mapping using Naïve Bayes and Bayesian network models in Umyeonsan, Korea. *Geocarto international*, 35(15), 1665-1679.

Marjanović, M., Kovačević, M., Bajat, B., andVoženílek, V. (2011). Landslide susceptibility assessment using SVM machine learning algorithm. *Engineering Geology*, 123(3), 225-234.

Mirus, B. B., Jones, E. S., Baum, R. L., Godt, J. W., Slaughter, S., Crawford, M. M., ... and McCoy, K. M. (2020). Landslides across the USA: occurrence, susceptibility, and data limitations. *Landslides*, 17(10), 2271-2285.

Oommen, T., Cobin, P. F., Gierke, J. S., andSajinkumar, K. S. (2018). Significance of variable selection and scaling issues for probabilistic modeling of rainfall-induced landslide susceptibility. *Spatial Information Research*, 26(1), 21-31.

Pourghasemi, H. R., Gayen, A., Edalat, M., Zarafshar, M., andTiefenbacher, J. P. (2020a). Is multi-hazard mapping effective in assessing natural hazards and integrated watershed management? *Geoscience Frontiers*, 11(4), 1203-1217.

Qin, C. Z., Bao, L. L., Zhu, A. X., Wang, R. X., and Hu, X. M. (2013). Uncertainty due to DEM error in landslide susceptibility mapping. *International Journal of Geographical Information Science*, 27(7), 1364-1380.

Rahali, H. (2019). Improving the reliability of landslide susceptibility mapping through spatial uncertainty analysis: A case study of Al Hoceima, Northern Morocco. *Geocarto International*, 34(1), 43-77.

Ruff, M., and Czurda, K., (2008). Landslide susceptibility analysis with a heuristic approach in the Eastern Alps (Vorarlberg, Austria). *Geomorphology* 94(3–4), 314-324.

Song, Y., Gong, J., Gao, S., Wang, D., Cui, T., Li, Y., and Wei, B. (2012). Susceptibility assessment of earthquake-induced landslides using Bayesian network: A case study in Beichuan, China. *Computers and Geosciences*, 42, 189-199.

Van Westen, C.J., Rengers, N., and Soeters, R. (2003). Use of geomorphological information in indirect landslide susceptibility assessment. *Natural Hazards* 30(3), 399-419.

Varnes, D.J. and the IAEG Commission on Landslides and other Mass-Movements, (1984). Landslide hazard zonation: a review of principles and practice.*The UNESCO Press, Paris*.

Wang, Q., Li, W., Wu, Y., Pei, Y., andXie, P. (2016). Application of statistical index and index of entropy methods to landslide susceptibility assessment in Gongliu (Xinjiang, China). *Environmental Earth Sciences*, 75(7), 1-13.

Wang, H. J., Xiao, T., Li, X. Y., Zhang, L. L., and Zhang, L. M. (2019). A novel physically-based model for updating landslide susceptibility. *Engineering Geology*, 251, 71-80.