Flood Risk Area Mapping with Logistic Regression: A Case Study of Phuntsholing City in Bhutan

Thongley Thongley[1]**, Chaiwiwat Vansarochana**[2] **and Vasker Sharma**[1]

[1]Department of Civil Engineering and Surveying, Jigme Namgyel Engineering College, Royal University of Bhutan

Email: thongley21@gmail.com, vaskarsharma@jnec.edu.bt

[2]Faculty of Agriculture, Natural Resources and Environment, Naresuan University, Phitsanulok, Thailand, 65000

Email: chaiwiwatV@gmail.com

ABSTRACT

Flood is one of the most frequent natural disasters in the different parts of the world. Phuntsholing city which is located in southern Bhutan also experiences several floods every year. Phuntsholing city is a commercial hub of Bhutan which is bordered with the Indian state of West Bengal. The Phuntsholing city cooperation has a plan to expand its area due to the rapid increase in population. However, some of the areas are prone to flood during the monsoon season. Therefore, this study aimed at developing a flood risk map at Phuntsholing using the logistic regression model. The factors used for this study area are elevation, terrain slope, soil type, lithology, distance from streams, flow accumulation, topographic wetness, index normalized difference vegetation index. The collinearity among these factors was checked using tolerance (TOL) and variance inflation factors (VIF) which should be more than 0.1 for TOL and less than 5 for VIF. All the factors have more than 0.1 for TOL and less than 5 for VIF. During the initial field survey, a total of 26 flood points were identified which is divided into training datasets (70%) and validation datasets (30%). The training dataset is used to train the selected factors while the validation datasets are used to validate the generated map. The generated flood map is divided into five classes namely: very low(22.71%), low(24.25), moderate(18.91%), high(23.95%), and very high(10.18%). The success rate of the flood map is 0.895 (89.5%) and the prediction rate is 0.9622 (96.22%) which is accurate enough for engineers and urban planners for future references.

1. INTRODUCTION

Flood is one of the most frequent natural disasters in the world. As per the International disaster database Centre for Research on the Epidemiology of Disasters (CRED), the flood claimed the life of 2,383,739 people, injured 124,4492, and 77,004,216 became homeless from the span of 1900-2021 in Asia alone (EM-DAT, 2021). The flood affected nearly 99 million people every year between 2000 and 2008 around the world (McClean, 2010). Urban development needs appropriate studies to minimize the damages caused by the flood.

Flood risk identification, event planning, and preparedness are crucial for flood risk reduction. One of the most common methods for risk reduction is preparing a flood risk map in the most vulnerable area. There are numerous methods to prepare a flood risk map which starts from simple statistical calculation to advance machine learning techniques. This study uses logistic regression is used to preparation of the flood risk map. The logistic regression requires a dichotomous dependent variable and several independent variables. The dichotomous variable is prepared during flood inventory while the independent variables are prepared from several datasets.

The report published by the Intergovernmental Panel on Climate Change (IPCC) in 2014 forecasted that Bhutan will experience a 5% decrease in rainfall during the dry season, and an 11% increase during the monsoon (Acharya, 2017). This may be true because the southern parts of Bhutan receive torrential rainfall during the monsoon season causing numerous floods. Bhutan is also vulnerable to glacial lake outburst floods (GLOFs) due to the huge amount of perpetual snow and ice creating many glacial lakes. There are around 2,674 glacial lakes are covering an area of about 107 sq.km in northern Bhutan threatening the low-lying places from the GLOF(Mool et al., 2001). Phuntsholing has vulnerable to flood due to its location in the low-lying terrain and contains numerous rivers such as Torsa river, Singyechu, and many other numerous petty rivers. The flood left 400 families homeless and 43 people went missing in August 2000 at the Pasakha area of Phuntsholing city(Tshering & Sithey, 2008). Asian Development Bank (ADB) has drafted a document about the expansion of Phuntsholing city by 66hectare which cost around $53m (ADB, 2018). However, it is very important to look into the possibility of flood events in the future at the time of town planning.

Therefore, this study aimed at developing a flood risk map of Phuntsholing city in Bhutan using the logistic regression model. This study is expected to guide urban planners for the future planning of the town, and flood risk management

2 Study Area

Phuntsholing (**Figure 1**) is located strategically on the Indo-Bhutan border at 89°23'E longitude and 26°52'N latitude. Phuntsholing is the second largest city of Bhutan and it is the main commercial hub of Bhutan with the majority of goods imports and export transit through Phuntsholing. Phuntsholing city is located at an elevation of about 293m above mean sea level characterized by a tropical monsoon climate. Its average annual precipitation is about 2953mm with a long rainy season and short dry winter season. The highest temperature recorded at Phuntsholing is about 40degree celsius.

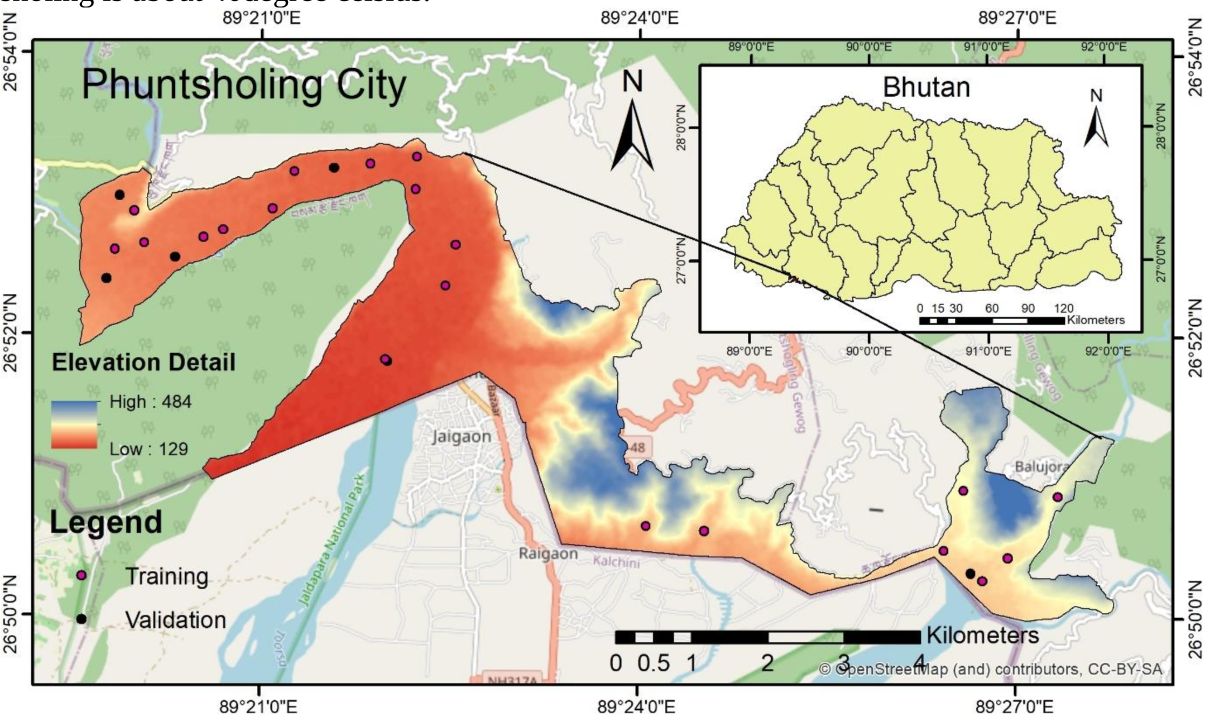

Figure 1. Phuntsholing City in Bhutan

3. Materials and methods

3.1 Flood Inventory

The flood inventory was done through google earth and field visits. A total of 26 flood points and an equal number of non-flood points were identified which is divided into a training dataset (70%) and validation dataset (30%) as shown in **Figure 1**. The training datasets were used to train the factors while the validation datasets were used to validate the generated map.

3.2 Dataset collection and preparation of factors.

The datasets were collected from ALOS PALSAR DEM, FAO soil map, Geological map of Bhutan, and sentinel 2 data. The elevation map, slope gradient, flow accumulation, and topographic wetness index were derived from ALOS PALSAR DEM while soil data was derived from the Food and Agriculture Organization of the United Nations (FAO). Similarly, the lithological map was derived from the geological map of Bhutan, and the normalized difference vegetation index was derived from the sentinel 2 image. All the factors were resampled to 12.5m spatial resolution.

3.3 Multi-collinearity diagnosis

Logistic regression is sensitive to collinearity among the factors and it is important to check the collinearity among the factors. It is important to examine the collinearity among the independent factors. Multicollinearity is caused by the high correlation between the independent factors (Tehrany, Jones, & Shabani, 2019). The tolerance (TOL) and variance inflation factor (VIF) are commonly used to check multi-collinearity. The TOL and VIF are calculated using Equations 1 and 2.

$$TOL = 1 - R^2 \tag{1}$$

$$VIF = \frac{1}{TOL} \tag{2}$$

The non-collinearity should have a tolerance of more than 0.1 and a VIF of less than 5 (Tien Bui et al., 2019).

3.4 Logistic Regression Model

Logistic regression is a multivariate statistical method and it is widely used in flood and landslide risk analysis. The logistic regression requires several independent variables on a single dichotomous outcome variable (Mind'je et al., 2019). The single dichotomous variable contains binary points (1 and 0) whereby 1 indicates flood point and an equal number of 0 indicates non-flood point. The logistic regression evaluates the correlation between the flood event and the influencing factors (Tehrany et al., 2019).

The probability (p) of the flood in logistic regression is calculated using Equation 3.

$$P = \frac{1}{1 + e^{-z}} \tag{3}$$

where p is the probability of flooding which ranges between 0 to 1 on an S-shaped curve. Z represents a linear combination and it is using Equation 4.

$$Z = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_nx_n \tag{4}$$

where b_0 is the intercept of the model, b_i (i = 0, 1, 2, …, n) represents the coefficients of the logistic regression model, and x_i (i = 0, 1, 2, …, n) denotes the independent variables (Lee & Sambath, 2006).

3.5 Accuracy assessment for the flood risk map

The accuracy assessment is an important step to check the reliability and efficiency of the map (Tehrany et al., 2019). The accuracy assessment of the generated flood map was done using the Area Under the Curve (AUC) of the Receiver Operating Characteristics (ROC) curve. The AUC is calculated using Equation 7. The ROC curve is constructed False Positive Rate (FPR) on the x-axis and True Positive Rate (TPR) on the y-axis (Thongley & Vansarochana, 2021b). The AUC is interpreted as excellent (0.9-1.0), very good (0.8-0.9), good (0.7-0.8), moderate (0.6-0.7), and poor (0.5-0.6) (Thongley & Vansarochana, 2021a). The sensitivity and specificity are calculated using the equation 5 and 6.

$$\text{TPR} = \frac{TP}{TP + FN} \quad (5)$$

$$\text{FPR} = \frac{FP}{\text{FP} + \text{TN}} \quad (6)$$

$$AUC = \frac{\sum TP + \sum TN}{TP + TN + FP + FN} \quad (7)$$

Where TP is true positive, TN is true negative, FP is false positive, FN is false negative

4. RESULT AND DISCUSSION

4.1 Multi-collinearity diagnosis and flood risk mapping

The multi-collinearity was checked using TOL and VIF. The factors were considered non-collinear if TOL is more 0.1 and VIF is less than 5. Table 1 shows the coefficient, TOL and VIF for individual factors. It is noticed that all the independent factors are free of collinearity.

Table 1. Coefficient of factors and the value of tolerance (TOL) and Variance Inflation Factor (VIF)

Factors	Coefficient	Collinearity Statistics	
		TOL > 0.1	**VIF <5**
Dist from river	-0.005	0.462	2.164
Elevation	-0.270	0.373	2.683
Flow Accumulation	0.001	0.884	1.132
Lithology	0.139	0.798	1.254
Slope gradient	0.044	0.609	1.642
Soil	0.555	0.875	1.143
TWI	-0.040	0.781	1.281
Constant	4.595		

The coefficients from Table 1 were used to calculate linear combination Z(Equation 3) which will ultimately be used to calculate the flood probability P (Equation 3) and generate a flood risk map.

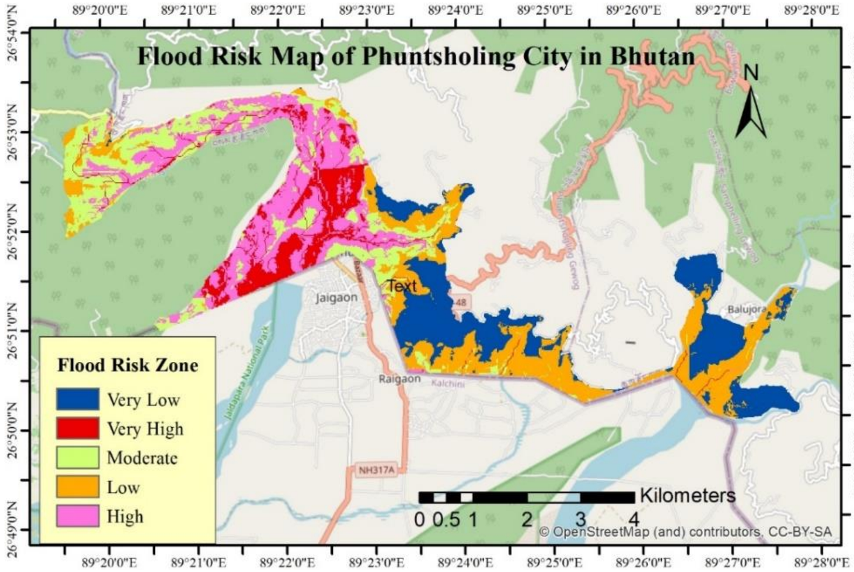

Figure 2. Flood Risk Map of Phuntsholing city in Bhutan.

The generated flood map **(Figure 2)** is divided into five classes based on the jenk classification due to its clear breakpoint between the classes (Toshiro, 2002). The five classes constitute very low (22.71%), low (24.25), moderate (18.91%), high (23.95%), and very high (10.18%).

4.2 Validation result of the flood risk map

The success rate shows the reliability of the model using trained data while the prediction rate shows the future prediction capability of the flood risk map (Thongley & Vansarochana, 2021b). The success rate of the flood risk map is 0.895 which falls under very good category while the prediction rate is 0.9622 which falls under the excellent category for the flood prediction **(Figure 3)**.

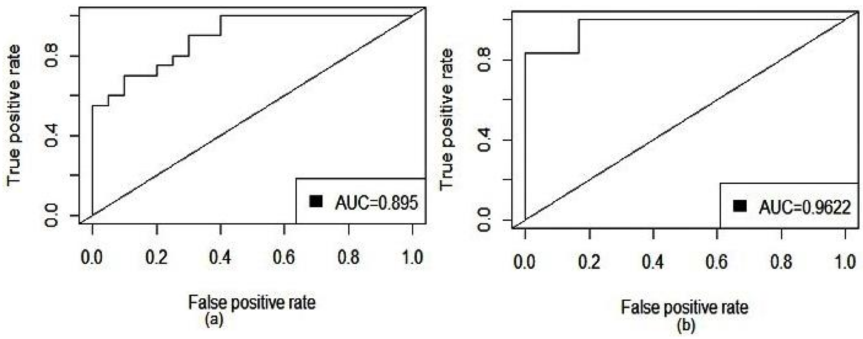

Figure 3. ROC Curve for validation (a) Success rate curve, (b) Prediction rate curve

5. CONCLUSION

The flood risk map is a preliminary step for minimizing major disasters due to the flood. The logistic regression is a simple and effective model for the preparation of the flood risk map in conjunction with the remote sensing data and geographic information techniques.

The generated map is a reliable map for future use due to its good performance with a success rate of 0.895 and excellent prediction rate of 0.9622. The final output shows a perfectly good result while comparing with the past flood area. The natural break classification shows the very high-risk area constitutes 10.18% of the total area. This study recommends not to plan any future developmental activities in the very high flood risk area which is 10.18%.

Since the success rate of the flood map is very good, the planners, engineers, and politicians can use it for future planning purposes while the researchers can also apply the logistic regression model for their future studies. It also recommends future researchers to study the highest volume of river water in the specified time and its return period for the prevention of possible loss of lives.

6. REFERENCES

Acharya, G. 2017, Flash floods the dangerous new normal in Bhutan. Retrieved from https://www.thethirdpole.net/en/climate/bhutan-flash-floods/

ADB. 2018. *Proposed Loan and Grant Kingdom of Bhutan: Phuentsholing Township Development Project* Asian Development Bank.

EM-DAT. 2021. *The International disaster database.* .

Lee, S., & Sambath, T. 2006. Landslide susceptibility mapping in the Damrei Romel area, Cambodia using frequency ratio and logistic regression models. *Environmental Geology, 50*(6), 847-855.

McClean, D. 2010. World disasters report 2010: focus on urban risk *World Disasters Report 2010: Focus on urban risk*: International Federation of Red Cross and Red Crescent Societies (IFRC).

Mind'je, R., Li, L., Amanambu, A. C., Nahayo, L., Nsengiyumva, J. B., Gasirabo, A., & Mindje, M. 2019. Flood susceptibility modeling and hazard perception in Rwanda. *International journal of disaster risk reduction, 38*, 101211.

Mool, P. K., Wangda, D., Bajracharya, S. R., Kunzang, K., Gurung, D. R., & Joshi, S. P. 2001. Inventory of glaciers, glacial lakes and glacial lake outburst floods. Monitoring and early warning systems in the Hindu Kush-Himalayan Region: Bhutan. *Inventory of glaciers, glacial lakes and glacial lake outburst floods. Monitoring and early warning systems in the Hindu Kush-Himalayan Region: Bhutan.*

Tehrany, M. S., Jones, S., & Shabani, F. 2019. Identifying the essential flood conditioning factors for flood prone area mapping using machine learning techniques. *Catena, 175*, 174-192.

Thongley, T., & Vansarochana, C. 2021a. Landslide Identification and Zonation Using the Index of Entropy Technique at Ossey Watershed Area in Bhutan. *Applied Environmental Research, 43*(1), 102-115.

Thongley, T., & Vansarochana, C. 2021b. Spatial zonation of landslide prone area using information value in the geologically fragile region of Samdrup Jongkhar-Tashigang national highway in Bhutan. *Environment and Natural Resources Journal, 19*(2), 122-131.

Tien Bui, D., Khosravi, K., Shahabi, H., Daggupati, P., Adamowski, J. F., Melesse, A. M., . . . Bahrami, S. 2019. Flood spatial modeling in northern Iran using remote sensing and gis: A comparison between evidential belief functions and its ensemble with a multivariate logistic regression model. *Remote Sensing, 11*(13), 1589.

Toshiro, O. 2002. Classification Methods for Spatial Data Representation. *Center for Advanced Spatial Analysis. University College London. London*.

Tshering, D., & Sithey, G. 2008. Climate change and health in Bhutan. *Thimpu: Royal Society for Protection of Nature*.