7-2024

# Is there a space in landslide susceptibility modelling: A case study of Valtellina Valley, Northern Italy

Min Naing KHANT
*Singapore Management University*, mnkhant.2020@scis.smu.edu.sg

Mei Yi Victoria Grace ANN
*Singapore Management University*, victoriaann.2021@scis.smu.edu.sg

Tin Seong KAM
*Singapore Management University*, tskam@smu.edu.sg



## Citation

KHANT, Min Naing; ANN, Mei Yi Victoria Grace; and KAM, Tin Seong. Is there a space in landslide susceptibility modelling: A case study of Valtellina Valley, Northern Italy. (2024). *Proceedings of the 24th International Conference on Computational Science and Its Applications, Hanoi, Vietnam, July 1-4*. 221-238.

**Available at:** https://ink.library.smu.edu.sg/sis_research/9350

# Is There a Space in Landslide Susceptibility Modelling: A Case Study of Valtellina Valley, Northern Italy

Khant Min Naing(✉), Victoria Grace Ann, and Tin Seong Kam

Singapore Management University, 80 Stamford Road, Singapore 178902, Singapore
{mnkhant.2020,victoriaann.2021,tskam}@scis.smu.edu.sg

**Abstract.** Landslides pose significant and ever-threatening risks to human life and infrastructure worldwide. Landslide susceptibility modelling is an emerging field of research seeking to determine contributing factors of these events. Yet, previous studies rarely explored the spatial variation of different landslide factors. Hence, this study aims to demonstrate the potential contribution of spatial non-stationarity in landslide susceptibility modelling using Global Logistic Regression (GLR) and Geographically Weighted Logistic Regression (GWLR). The second objective of this study is to demonstrate the important role of data preparation, data sampling, variable sensing, and variable selections in landslide susceptibility modelling. Using Valtellina Valley in Northern Italy as the study area, our study shows that by incorporating spatial heterogeneity and modelling spatial relationships, the measures of Goodness-of-fit of GWLR outperform the traditional GLR. Furthermore, the model outputs of GWLR reveal statistically significant factors contributing to landslides and the spatial variation of these factors in the form of coefficient maps and a landslide susceptibility map.

**Keywords:** Landslide Susceptibility · Geographically Weighted Logistic Regression · Logistic Regression · Explanatory Modelling

## 1 Introduction

Landslides refer to the geomorphic phenomenon of slope failure and mass movement in mountainous regions due to eroding and depositing sediment. They are an ever-present threat to critical infrastructure and urban communities worldwide expanding costly damage and massive displacement hampering urban development [1, 2]. Within the past five decades landslide events have increased tenfold [3]. By and large, they are driven by many triggers and conditioning factors [4] including historical evolutions in topography or environment, changes in weather patterns, vegetation cover and river networks and other man-made stimuli.

Extensive engineering prevention works such as concrete surfacing, terracing [5] and slope modification [6] are typical and successful landslide risk mitigations for more gradually sloped areas. However, implementing these works on steeper terrains incurs greater costs due to the complexity of these modifications. Slope geometry alterations,


O. Gervasi et al. (Eds.): ICCSA 2024, LNCS 14813, pp. 221–238, 2024.
https://doi.org/10.1007/978-3-031-64605-8_16

complex drainage systems installations and reinforcing internal structures for a single high-risk site easily cost millions [6]. Landslide susceptibility maps and assessments can serve as decision-support tools for urban planners and engineers by tiering areas based on their susceptibility levels. This information can then be used to identify highest-risk areas to be prioritised for formulation of preventive measures and site suitability assessments for new developments [7].

This paper aims to demonstrate the potential use of Geographically Weighted Logistic Regression (GWLR) to determine factors influencing landslide susceptibility. GWLR is an extension of Global Logistic Regression (GLR) and offers a significant advantage in binary classification over GLR by addressing spatial non-stationarity present in the global models. Landslide occurrences are closely linked to environmental characteristics and render spatial heterogeneity, thus explanatory variables are unlikely to have an equal contribution to landslide events [8]. Applying local techniques like GWLR to datasets with inherent spatial heterogeneity enables to capture local variations based on the location and proximity of the data samples. Such local variations tend to be lost or averaged out in the global parameter estimates. As a result, GWLR produced more spatially adaptive and accurate outcomes. This paper discusses how GWLR, compared to GLR, is more appropriate for analysing the contribution of each factor to landslides in different parts of the study area.

This paper consists of seven sections, namely literature review, study design, study area and data overview, results, discussion, and conclusion. The literature review presents an overview of existing landslide modelling techniques and factors used in landslide susceptibility modelling. The study design section explains the research questions, research methodology, and the models explored in the study. The study area and data overview sections account for Valtellina's geographical background and the preparation of its landslide inventory and susceptibility factors. The results section covers the experiment results and observations. Finally, the discussion and conclusion sections summarise the key research findings and interpretations, maps of landslide factor coefficient and landslide susceptibility and potential follow-ups on this research.

## 2 Literature Review

### 2.1 Landslide Susceptibility Modelling Approaches

Landslide susceptibility modelling is a rapidly evolving field of research given the urgency for landslide prediction and mitigation. Previous studies have investigated and established various methodology frameworks, each with different objectives and modelling techniques. Generally, these studies were focused on either explanatory landslide modelling or predictive landslide modelling, or both. Explanatory landslide studies seek to identify, illustrate, and explain critical factors contributing to landslide events within a study area and their level of influence on these events. Predictive modelling attempts to identify and predict potential landslide occurrences in the study area and to tier the study area into *high-risk* and *low-risk* zones for future preventive measures.

In terms of variation in modelling techniques, landslide susceptibility studies can be broadly classified into four groups: physically based, expert-based, statistical, and machine learning-based models. Physically based models are considered to have the highest utility [9] as historically accurate analytical outcomes and clear interpretation are extracted using physical and mechanical principles [9]. Models like the Shallow Landsliding Stability Model (SHALSTAB) use a deterministic approach to analyse slope failures under steady-state conditions relying on physical data such as rainfall, topography, local slope, and soil transmissivity [10]. However, physically based models are challenged with uncertainties mainly due to limited spatially differentiated geotechnical data that are not readily available [9, 11]. Consequently, these models are applied mostly to smaller-scale landslide risk assessments. Expert-based models are less data-intensive than physically based models mostly developed from expert knowledge of local interactions between landslide occurrences and their controlling factors [12]. Previous studies have used heuristic models by qualitative multicriteria analysis [13] and index methods [14]. These models have proven effective in geographical settings exhibiting high spatial variability. However, such expert models with a qualitative nature tend to be subjective and have limited reproducibility [9, 14, 15] and comparability [16] with other models or locations. Therefore, quantitative methods are more commonly used than qualitative approaches in scientific research.

Statistical landslide susceptibility modelling aims to estimate the relative spatial probability of spatial units with future landslide incidents such that the relationships between the dependent variable (landslides) and the independent variables (individual or combination of conditioning factors) can be identified and quantified. These models are widely adopted since less data input is required than physically based models and quantitative and objective results are provided which expert-based models lack [9, 17].

Machine learning (ML) models are emerging, suitable alternatives to statistical-driven models given their potential for high predictive accuracy [9]. They are favoured for solving non-linear associations between landslides and factors [8] and improving predictive performance across various classifiers. Recent studies have successfully applied machine learning techniques, such as random forest, convolutional neural network [2], deep learning and tree-based models [18] to identify landslide susceptibility with higher accuracy. Despite improved accuracies, many ML techniques suffer from limited interpretability and explainability due to model complexity [9] and hence are primarily used in predictive rather than explanatory modelling.

Overall, a literature review of previous studies has demonstrated that each model type has its unique strengths and limitations, and choosing an appropriate model depends largely on specific research objectives [9] or context of study area or datasets.

### 2.2 Factors Influencing Landslide Susceptibility

There are numerous causes and factors for landslides. While slope movement can be triggered by heavy rainfall, snowmelt events, earthquakes or volcanic activities, landslide susceptibility is usually attributed to several underlying conditions and factors [4] which weaken slope stability over time. A trigger event causes weakened slopes to undergo mass movements and landslides. Previous research has extensively explored factors including topographical, geological, hydrological, and environmental characteristics in landslide susceptibility modelling. However, many of these studies overlook anthropogenic, or human factors, creating a research gap in the role of human-environmental interaction factors in landslide susceptibility. Today's rapid urbanisation has expanded the spatial scope of human activities [19], often at the expense of ecosystems and the geological environment. Recent research also reveals a growing correlation between urbanisation and landslide risk [20], particularly concerning construction activities, mining and hill cutting [4]. The dynamic changes in land use land cover and infrastructure underscores the need for a more comprehensive approach that includes geo-environmental and anthropogenic factors in landslide susceptibility studies.

Furthermore, limited explanatory landslide susceptibility studies assess the significance of different susceptibility factors and their geomorphic impacts on model outcomes. Moreover, the growing interest in spatial properties, such as spatial heterogeneity and spatial non-stationarity, calls for more studies to explore, illustrate and explain how such spatial properties intercept with the influence of different landslide factors.

# 3 Study Design

## 3.1 Research Questions

Landslide susceptibility modelling using statistical techniques heavily relies on spatial input data, which exhibit inherent spatial characteristics. Given the First Law of Geography which states that "*everything is related to everything else, but near things are more related than distant things*" [21], traditional global models generalising study areas cannot capture the relationship among parameter attributes relative to geographical space. Fotheringham, Charlton and Brunson [22] highlighted how regression estimates can fluctuate across space and underscored the need for model calibration to accommodate spatial heterogeneity in spatial dataset.

Through a comparative analysis of GLR and GWLR models, this study examines whether calibrating global models into local derivatives accounts for spatial relationships and enhances model outcomes. The study also investigates the role of "space" in landslide susceptibility and the individual influence of landslide factors on susceptibility results. Three research questions guide this study – i) which landform properties and human-environment interaction factors affect landslide susceptibility; ii) whether the contribution of landslide factors varies across the study area and iii) whether landslide susceptibility is geospatially independent.

## 3.2 Research Methodology

The research methodology for this study is summarised in Fig. 1. First, the study will build and calibrate a GLR model. This model serves as a basis for understanding the general influence of geo-environmental and human factors on landslide occurrences. The study then builds and calibrates different GLR and GWLR models to analyse the local and spatially adaptive effect of different geo-environmental and human factors on landslide occurrences. Model calibrations are determined by a stepwise selection algorithm to improve model performance per iteration. By comparing the results from calibrated global and local models, the study attempts to develop a methodology framework for explanatory landslide susceptibility modelling.

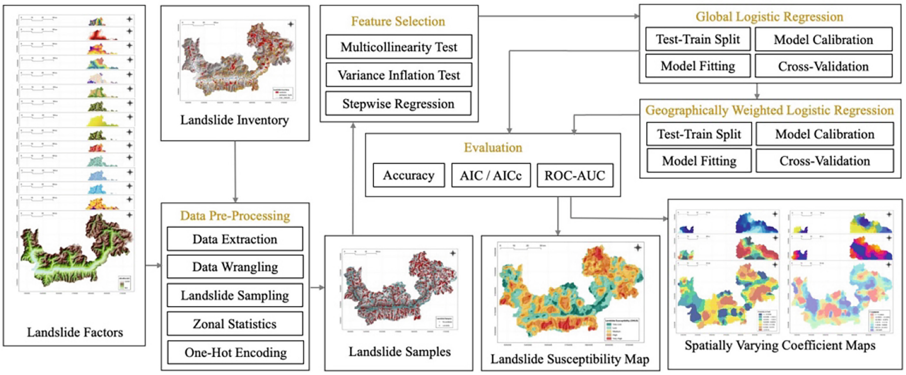



**Fig. 1.** Research Methodology Diagram.

### 3.3 Global Logistic Regression (GLR)

Global Logistic Regression (GLR) is a multivariate statistical method established as a common technique for binary classification. GLR is appropriate for landslide susceptibility modelling due to the binary nature of the response variable representing either the landslide presence (indicated as positive class or value of 1) or absence (indicated as negative class or value of 0). Moreover, the target variables can be either discrete or continuous and are not required to satisfy the normal distribution [23]. The GLR model first calculates the log-odds ratio of the event being a positive class by the probability of the event being a negative class. This is done by the linear combination of the independent variables [11, 24]. Next, this linear combination is transformed into a probability value by fitting a sigmoid function. The GLR model can be formulated as:

$$log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k \tag{1}$$

$$p(y = 1|x) = \frac{e^{(\beta_0+\beta_1 x_1+\cdots+\beta_n x_n)}}{1 + e^{(\beta_0+\beta_1 x_1+\cdots+\beta_n x_n)}} \tag{2}$$

where $x_1$, $x_2$, ..., $x_n$ are the independent variables. $\beta_n$ is the coefficient estimate of the independent variable $x_n$. The coefficient estimates of the GLR model are calculated using *maximum likelihood estimation* [23, 24] which maximises the likelihood of observing the given target results by iteratively updating the previously fitted coefficients until the optimal coefficient values are obtained.

### 3.4 Geographically Weighted Logistic Regression (GWLR)

Geographically Weighted Logistic Regression (GWLR) is a local regression framework proposed by Fotheringham, Charlton and Brunson [22]. This model is suited for spatial analysis as it extends traditional and global regression frameworks such as Ordinary Least Squares (OLS) and GLR by integrating spatial weights and generating local-level model statistics [25]. GLR model can be calibrated to GWLR as follows:

$$p_{(u_i,v_i)} = \frac{e^{\left(\beta_{0(u_i,v_i)} + \beta_{1(u_i,v_i)}x_{i1} + \cdots + \beta_{ik(u_i,v_i)}x_{ik}\right)}}{1 + e^{\left(\beta_{0(u_i,v_i)} + \beta_{1(u_i,v_i)}x_{i1} + \cdots + \beta_{ik(u_i,v_i)}x_{ik}\right)}} \tag{3}$$

where $(u_i, v_i)$ denotes the coordinates of i$^{\text{th}}$ sample in space. $\beta_k(u_i, v_i)$ is a realisation of the continuous function $\beta_k(u, v)$ at i$^{\text{th}}$ sample point. In this way, the coefficient estimates of independent variables are calculated at different sample points. As a result, the GWLR equation does not assume the coefficients to be spatially constant or random but varies within the sample space [16, 22]. In GWLR, spatial weighting quantifies spatial dependencies between variables into an $n \times n$ weight matrix, $W(u_i, v_i)$. The assigned weights vary according to the proximity, with observations nearer to $i$ are assigned higher weight compared to distant ones [22]. This geographic weighting allows nearby observations to exert more influence on the estimation of local regression coefficients.

There are three key elements in building a weight matrix: (i) The distance metric element calculates the distances between locations; (ii) The kernel function defines how weights decay with distance and captures spatially varying relationships; (iii) The bandwidth controls a location's neighbourhood influenced by a fixed distance or an adaptive distance determined by a fixed number of neighbours. The model is then calibrated by data points within the coverage of the bandwidth. GWLR generates location-specific realisations of model goodness-of-fit measures and parameter estimates. These local estimates of explanatory variables can be visualised as a surface to illustrate spatial variations in the relationship between landslide factors and susceptibility [25].

# 4 Study Area and Data Overview

## 4.1 Study Area

Located in the Central Alps of Northern Italy, Valtellina Valley extends between 515,000 m and 620,000 m in easting (9.24°E to 10.63°E in longitude), and between 5,050,000 m and 5,170,000 m in northing (46.00°N to 46.64°N in latitude) and covers about 3308 square kilometres. Its average mountain elevation ranges from 2,500 to 3,000 m with the bottom of the valley lying about 1000 to 1100 m above sea level. The East-West orientation of the valley is also attributed to the Periadriatic Line which imposes tectonic lineament upon the valley [26–28]. The region's susceptibility to disasters has been attributed to geological instability from tectonic and post-glacial conditions [28], soil type and moisture, and soil acclivity activating mass movements [29]. Tourism-related activities have indirectly remodified the landform's geomorphology and raised the susceptibility of landslides [28, 29] (Fig. 2).

## 4.2 Data Acquisition and Processing

The quality of landslide inventory plays a critical role in effectively assessing and predicting the likelihood of landslide occurrences [30]. Inventory of historic landslides can be constructed using traditional field surveys or remote sensing and satellite imageries.

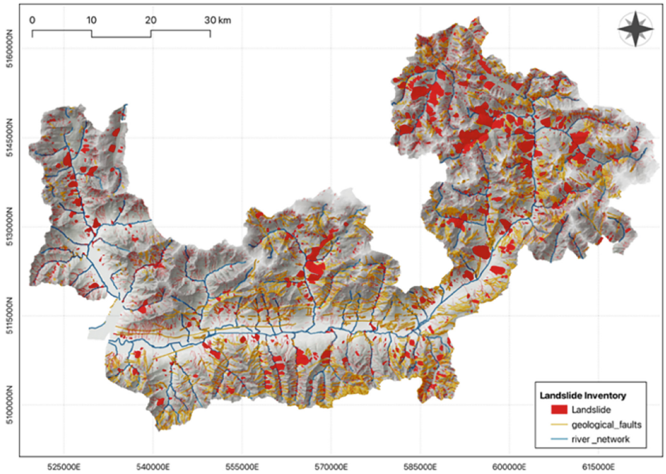



**Fig. 2.** Location map of study area.

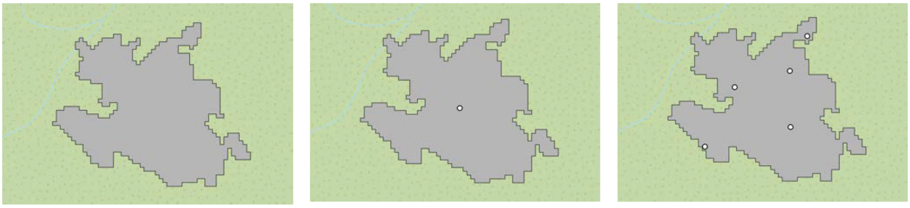
Remote sensing images in particular can capture high-resolution surface data and allow for more precise detection and delineation of landslide areas.

For this study, a total of 10,483 instances of translational, rotational shallow landslides and debris flows were gathered from the Inventory of Landslide Phenomena in Italy (IFFI). The inventory is maintained by The Italian Institute for Environmental Protection and Research (ISPRA) [31] through collaboration between regional and autonomous provinces building a comprehensive and quality national inventory corroborating the suitability for this study. The landslide inventory also include non-landslide samples developed by The Geomatics and Earth Observation laboratory (GIS-GEOLab @Politecnico di Milano) and complement landslide samples for training and testing sets of this study.

### 4.3 Landslide Data Sampling

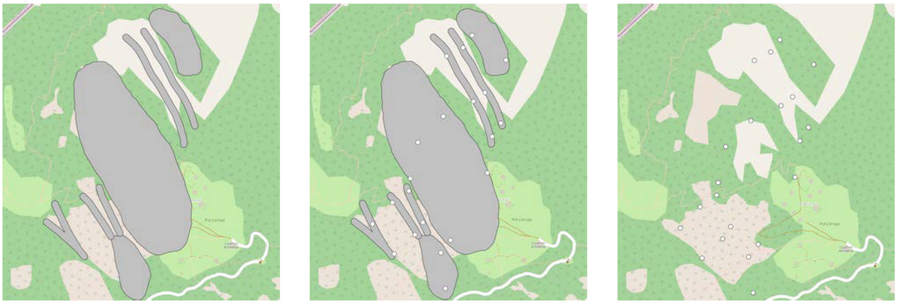
High-quality input data is heavily emphasised for landslide susceptibility models, but it is challenging to sample precise landslide initiation points. Hussin et al. [32] mitigated this with sampling strategies like the mass center, body mass, crown and scarp method. Gaidzik & Ramírez-Herrera [9] highlighted sampling rupture zones and undisturbed close vicinity. However, these strategies lack consistency between non-landslide and landslide sampling methods. Gaidzik & Ramírez-Herrera [12] and Regmi et al. [7] found landslide mass sampling and scarp samples more effective than mass centre sampling. In contrast, Steger and Kofler [9] argued that one-cell sampling is superior given its reduced uncertainties in identifying landslide initiation zones and boundaries (Fig. 3).

In this study, five randomly distributed random points within landslide boundaries were sampled in this study to mitigate centroid-based sampling bias and the uncertainties of body mass sampling (see Fig. 4). Random point sampling counters the biasedness of randomly sampling absence data causing the underrepresentation or overrepresentation of landslide absences that may reinforce misleading modelling results [9].



*Body Mass Sampling* *Mass Centre Sampling* *Random Point Sampling*

**Fig. 3.** Landslide sampling methods.



*Sample Polygons* *Random Point Samples in Sample Polygons* *Random Point Samples*

**Fig. 4.** Random point sampling of landslide inventory.

### 4.4 Complete and Quasi-Complete Separation

Complete and quasi-complete separation are common conditions preventing the convergence for maximum likelihood estimates in logistic regression by perfectly or partially separating the target variable from predictors [33].

A quasi-complete separation is observed for slope angle values in Fig. 5. Most landslide samples are found in steep-sloping areas (angles $>$ 20), while non-landslide samples predominantly occupy areas with gentler slopes (angles between 0 and 20). This data imbalance could bias predictions and any instance with a slope angle greater than 20 is classified as a landslide. The slope angle exhibits an IV value of 5.3575 significantly higher than the other features, whose IV values range from 0 to 0.5120, indicating that the slope angle feature has the greatest predictive capability.

To correct data imbalance and ensure the model is trained on a more representative sample, a slope angle threshold of up to 20 degrees is established to reduce the sample to areas with gentler slopes. The stratification ensures the relationship between slope angle and landslide occurrence is captured without the influence of the data balance in steeper slope areas, leading to a more reliable and robust model.

### 4.5 Landslide Susceptibility Factors Selection and Preparation

Careful consideration and acquisition of appropriate data for various landslide susceptibility factors is of equal significance to landslide inventory. To this end, Geoportale

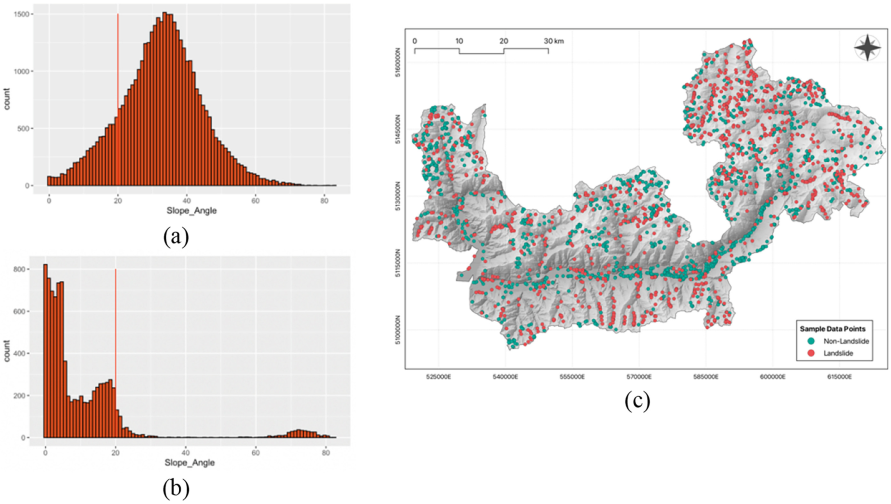



**Fig. 5.** Distribution of slope angle values in (a) landslide samples, (b) non-landslide samples and (c) slope-based stratified sample points for modelling GLR and GWLR.

della Lombardia has maintained a comprehensive collection of geospatial data tailored for the Lombardy Region [31]. Moreover, the digital lithology compiled by Bucci et al. [34] has furnished an extensive classification of geological formations in Italy at a scale of 1:100 000. A selection and preparation of 15 landslide susceptibility factors (See Table 1) have been undertaken for this study. These factors encompass a wide range of geological, topological, meteorological, hydrological, and human-related aspects.

### 4.6 Curse of Multicollinearity

Multicollinearity arises when two or more explanatory variables display moderate or high correlation, and it complicates the assessment of the importance and significance of each individual predictor. Nonetheless, multicollinearity can be detected constructing a correlation matrix of the explanatory variables. Pearson's correlation coefficient is commonly used to calculate and construct a correlation matrix. It quantifies the linear association between two variables, with its values ranging from $-1$ (indicating a completely negative correlation) to 1 (indicating a completely positive correlation). A value of 0 signifies no correlation. The maximum correlation coefficient is observed at 0.57 for *elevation* and *distance to settlement*, and *elevation* and *distance to road*. Other factors exhibit a range of low to moderate correlations. Overall, no factors among those selected for this study had a high positive correlation.

Subsequently, variance inflation factor (VIF) and tolerance (TOL) values are calculated to evaluate landslide variables. VIF is a statistical measure used in regression analysis to observe the increase in variance of the regression coefficient estimates due to multicollinearity. There are currently no established criteria for identifying the extent of VIF values that result in poorly estimated coefficients, but a frequent benchmark in

**Table 1.** Summary of landslide susceptibility factors used for modelling.

| Factor | Source | Scale | Type |
|---|---|---|---|
| **Topographic Factors** | | | |
| (1) Elevation | Geoportale della Lombardia, Derived from DEM | 15 × 15 m | Continuous |
| (2) Slope Angle | | 15 × 15 m | Continuous |
| (3) Aspect | | 15 × 15 m | Categorial |
| (4) Profile Curvature | | 15 × 15 m | Continuous |
| (5) Plan Curvature | | 15 × 15 m | Continuous |
| **Geological Factors** | | | |
| (6) Lithology | Bucci et al., 2021 | 1:100 000 | Categorical |
| (7) Distance to Faults | Geoportale della Lombardia | 15 × 15 m | Continuous |
| **Meteorological Factors** | | | |
| (8) Average Precipitation | ARPA Lombardia | 15 × 15 m | Continuous |
| **Hydrological Factors** | | | |
| (9) Distance to Streams | Geoportale della Lombardia, Derived from DEM | 15 × 15 m | Continuous |
| (10) Topographic Wetness Index | | 15 × 15 m | Continuous |
| (11) Steam Power Index | | 15 × 15 m | Continuous |
| (12) Sediment Transport Index | | 15 × 15 m | Continuous |
| **Anthropogenic Factors** | | | |
| (13) Distance to Settlements | OpenStreetMap | 15 × 15 m | Continuous |
| (14) Distance to Road Networks | Geoportale della Lombardia | 15 × 15 m | Continuous |
| (15) Land Use Land Cover | | 15 × 15 m | Categorial |

numerous regression studies is VIF $\geq 5$ [35]. VIF $\geq 5$ indicates significant multicollinearity and may necessitate further investigation or actions [8]. Three lithological categories – metamorphic, sedimentary, and unconsolidated – show high VIF values (7.85, 5.98 and 9.42 respectively), indicating significant multicollinearity. To avoid potential issues in subsequent modelling results, the variable with the highest VIF value, lithology (unconsolidated) is removed from the dataset. Post-removal, all VIF values are less than 2.27 and TOL values are above 0.44, indicating no further multicollinearity issues.

### 4.7 Stepwise Selection of Significant Landslide Factors

When using multiple logistic regression to model landslide susceptibility, it is important to evaluate each variable's significance level to avoid overcomplexity. In this regards, stepwise regression can be employed to iteratively select variables by adding or removing them based on likelihood ratio and $p$-value [11]. Variables are included based on their statistical significance, i.e., $p < 0.05$ and are removed otherwise, thereby selecting only meaningful predictors. The selection continues until no variables meet the inclusion or exclusion criteria. AIC can indicate the optimal complexity where selecting the model with the lowest AIC score will optimise the performance solely based on the reliance on *p*-values. The final 8 landslide factors – slope angle, profile curvature, plan curvature, lithology (plutonic), lithology (metamorphic), distance to roads, landuse (vegetation) and average precipitation – were and used for subsequent model training.

## 5 Results

### 5.1 Model Evaluation and Validation

Different GLR and GWLR models were fitted and calibrated using *GWmodel* package in the R Environment [36]. A total of 8 landslide factors from stepwise selection are the selected independent variables. An adaptive bandwidth of 76 and Gaussian kernel were used in the GWLR model to calibrate the model. Upon completion, three Goodness-of-Fit measures, namely deviance, AICc and pseudo- $R^2$ were used to compare the performance of GLR and GWLR models. Deviance assesses how well the fitted model compares to the null model, indicating goodness of fit. A higher deviance indicates a poorer fit than the "*best case*". Pseudo-$R^2$ serves as a versatile goodness-of-fit indicator for logistic regression models. AIC and corrected AIC (AICc). When the sample size is smaller, a higher penalty term is needed and corrected AIC (AICc) is a more reliable criterion. AIC and AICc can be used to rank models based on their model fit, and smaller values indicate a better model. Model diagnostic values of the GLR model are given in Table 2 and it reveals that GWLR outperforms GLR in all three measures of Goodness-of-Fit as there is an improvement in the indicators.

**Table 2.** Evaluation of GLR and GWLR Model.

| Performance Measures | Models | |
|---|---|---|
| | GLR | GWLR |
| Pseudo-$R^2$ | 0.402259 | 0.532152 |
| Deviance | 2291 | 1793.14 |
| AIC | 2309 | 2103.232 |
| AICc | 2309.047 | 2121.513 |

The coefficient estimates results, and statistical significance measures derived from the GLR model have been reported in Tables 3 and 4. The coefficient for slope angle and landuse (vegetation) is estimated to have positive relationships with landslide events at 0.2455, and 0.7871 respectively. On the other hand, two curvature measures - profile and plan, two lithology classes – plutonic and metamorphic as well as the distance to roads and average precipitation show a negative relationship, implying that an increase in these variables leads to a decrease in the probability of landslide occurrence. All the coefficient estimates from the GLR model show statistical significance. On the other hand, GWLR produces a coefficient estimate for every sample location, rather than a single fixed value.

The minimum, maximum, and quantile values of the coefficient estimates are reported in Table 4.

**Table 3.** Coefficient Estimate Results of GLR Model.

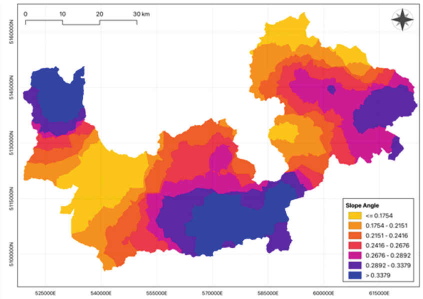
| Landslide Factors | Coefficient Estimates from GLR | | |
|---|---|---|---|
| | Coefficients | z value | Pr(>\|z\|) |
| Slope Angle | 0.2455 | 24.3025 | 0.0000 |
| Profile Curvature | −879.6932 | −15.3261 | 0.0000 |
| Plan Curvature | −732.9813 | −11.3090 | 0.0000 |
| Lithology (Plutonic) | −1.3382 | −4.2717 | 0.0000 |
| Lithology (Metamorphic) | −0.3410 | −2.8987 | 0.0037 |
| Distance to Roads | −0.0077 | −2.9310 | 0.0034 |
| Landuse (Vegetation) | 0.7871 | 7.003 | 0.0000 |
| Average Precipitation | −2.5301 | −2.1931 | 0.0283 |

**Table 4.** Coefficient Estimate Results of GWLR Model.

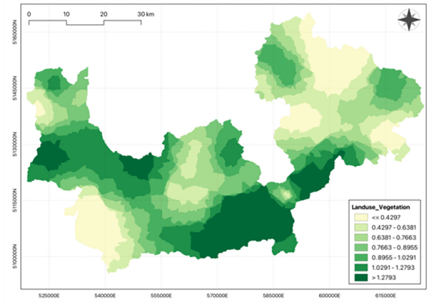
| Landslide Factors | Coefficient Estimates from GWLR | | | | |
|---|---|---|---|---|---|
| | Min | $1^{st}$ Quantile | Median | $3^{rd}$ Quantile | Max |
| Slope Angle | −0.096495 | 0.2094 | 0.25654 | 0.30969 | 0.4220 |
| Profile Curvature | −1325.6 | −1020.9 | −882.82 | −765.73 | −316.9567 |
| Plan Curvature | −1884.14 | −1007.3 | −664.35 | −443.95 | −329.3031 |
| Lithology (Plutonic) | −10.819 | −3.0798 | −1.5746 | −0.47993 | 1.8951 |
| Lithology (Metamorphic) | −1.7751 | −0.57195 | −0.27482 | −0.18384 | 2.1692 |
| Distance to Roads | −0.054247 | −0.0262 | −0.013533 | 0.00157 | 0.0324 |
| Landuse (Vegetation) | −0.22363 | 0.5102 | 0.8376 | 1.0731 | 3.7799 |
| Average Precipitation | −29.477 | −5.8893 | 2.9382 | 12.095 | 46.7696 |

## 6 Discussion

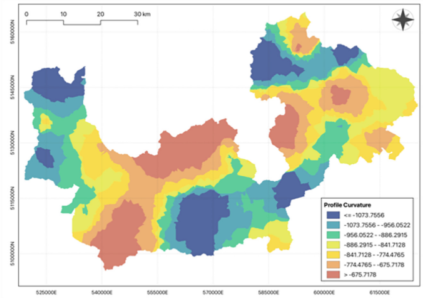
The study explored the influence and significance of space in landslide susceptibility modelling and detected spatial non-stationarity in the relationships between the landslide factors and the landslide locations. The GWLR model has the influence and significance of landslide explanatory factors that vary locally. Table 4 details such variations in numerical forms. Figure 6 provides a visual representation of the coefficient estimates for each variable on a planar surface, highlighting the spatially varying degrees of influence each variable exerts on the landslide susceptibility results.



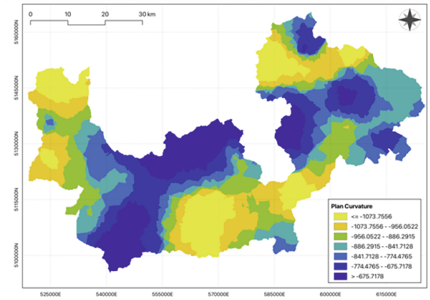

(a) Slope Angle



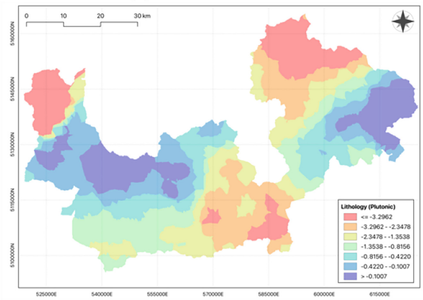

(b) Landuse (Vegetation)



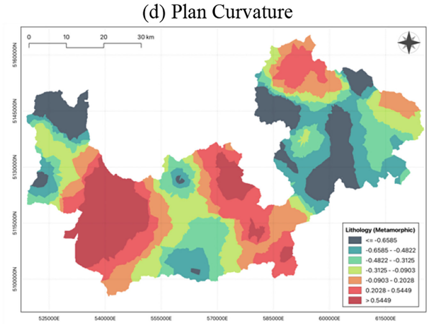

(c) Profile Curvature



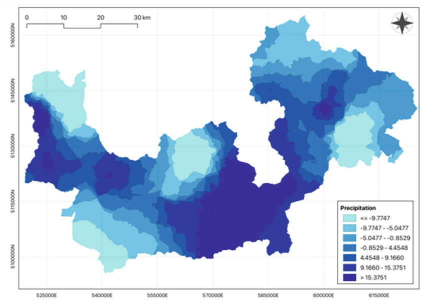

(d) Plan Curvature



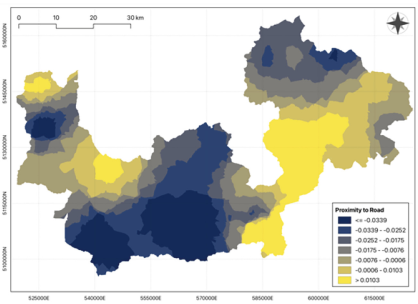

(e) Lithology (Plutonic)




(f) Lithology (Metamorphic)




(g) Average Precipitation




(h) Proximity to Road

**Fig. 6.** Coefficient estimate maps of GWLR showing spatially-varying estimates.

The results from GWLR reveal spatial variations in the coefficient estimates of different variables. The slope angle has small variations ranging from –0.096 to 0.42 due to stratified sampling. Except for minimum coefficient estimates, other estimate ranges reflect a positive correlation with landslide, indicating that a unit degree increase in slope angle contributes to a relative increase in landslide risk in different scales for most of the study area. The coefficient estimates for profile and plan curvature are larger than other variables (coefficient range: –1325.6 to –317.0 for profile curvature and –1884.1 to –329.3 for plan curvature). This may be attributed to the smaller scale of the values of these variables, and as a result, a slight change in the value leads to a significant change in landslide risk. Notably, the coefficient estimates remain negative across the study areas. It can be interpreted that upward convexness in profile curvature and sideward concaveness in plan curvature increase landslide risks.

Plutonic and metamorphic lithology variables exhibited variations ranging from negative to positive coefficient estimates with negative predominance across quantiles, indicating such lithology does not increase landslide risk in most areas. As for distance to roads, similar regions with positive and negative correlations are observed. The magnitude only ranges from –0.054 to 0.032, which may be attributed to smaller variations in proximity values of the sampled data. Nonetheless, relating the varying coefficient estimates to underlying geological structures or soil types could provide further insights into how these factors influence the changing relationships between predictors and landslide susceptibility across the study area.

Interestingly, landuse (vegetation) is positively correlated to landslide risk in most of the study area. Though this contradicts the initial hypothesis that urban and industrial landuse areas are more susceptible to landslides due to geomorphological modifications, a counterargument can be made that actual developments on the ground may largely differ from the planned landuse classification at a policy level, suggesting the need for alternative, empirical data quantifying landuse in future studies. Lastly, for average precipitation, except for the first quantile (coefficient of –0.22 and lower), a positive correlation is observed to landslide risk. On average, increases in precipitation slightly raised landslide risk (coefficient of 2.94), but there are also extreme cases observed in two ends – an increase in precipitation either increases (coefficient of –29.48) or decreases landslide risk (coefficient of 46.77) to a greater deal. The identical unit increase in precipitation thus leads to vastly different outcomes.

Figure 7 shows the landslide susceptibility map derived by using GWLR methods. Overall, there is variation in the susceptibility level estimated by the GWLR model across our study area. The distribution of susceptibility levels across the region is not uniform. A large proportion of *very high* susceptible areas are found in the southern part of the study area. The other areas observing *high* susceptibility include the northeastern edge of Valtellina Valley and the western peripheries. These are areas estimated to have a 0.8 probability likelihood of landslide events. Many areas with *high* risk are generally closely located to *very high-risk* areas, proving the existence of spatial autocorrelation. Expanding the GWLR model to surrounding regions can also enable the understanding of the changes in coefficient estimates and susceptibility patterns over larger land areas.

Specifically to this study area, it is worth further investigating how the northern region between Easting 585000 and 600000 coinciding with the area of highest coefficients of at least 0.0103 for proximity to roads in Fig. 6h has a *very low* risk of landslide which deviates from the general trend of the study area. However, given the lower coefficient values for proximity to roads compared to the other selected predictors, future research can focus on benchmarking the significance of such coefficient estimates.

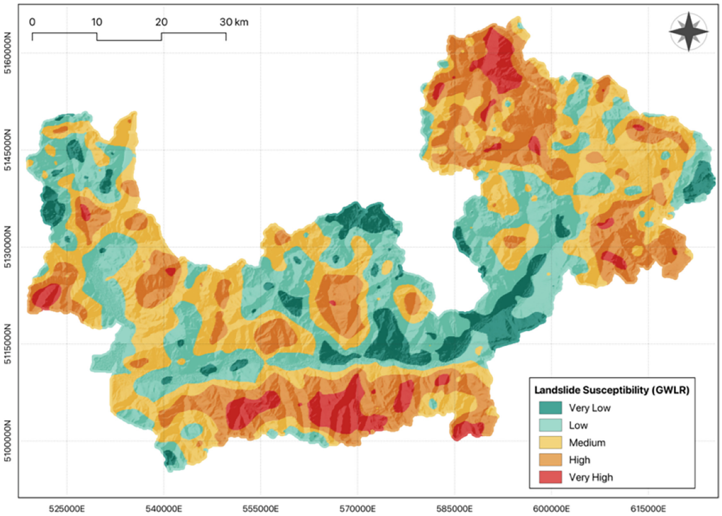



**Fig. 7.** Landslide susceptibility map produced from GWLR model results.

Overall, the comparison of results from GLR and GWLR models demonstrates the importance of considering spatial variability when developing landslide susceptibility models. By incorporating this spatial non-stationarity, more accurate, region-specific models can be created to better predict and mitigate landslide risks in different areas.

## 7 Conclusion

In conclusion, our study shows that the coefficient estimates from GLR model indicate that both landform properties (slope angle, curvature, and lithology) and human-environmental factors (distance to roads, landuse, precipitation) have statistically significant contributions to landslide susceptibility. Complementing this, the application of a geographically weighted regression model in this study reveals the presence of spatial heterogeneity as an inherent characteristic in landslide susceptibility mapping. The spatially varying relationships between landslide explanatory variables and landslide occurrence probability were visualized and interpreted through maps. The study discussed how the landslide events across different regions of the study area may be influenced by a distinct set of landslide explanatory variables, each contributing at varying levels of significance. The landslide susceptibility map produced by this study also indicate that landslide susceptibility is not geospatially independent. This is evidenced by the observation that nearby locations tend to exhibit similar levels of susceptibility, underscoring that landslide susceptibility is not a random process but a function of an underlying spatial process influenced by susceptibility factors. This finding underscores the importance of considering spatial dependencies when assessing landslide susceptibility.

This study has laid a methodology framework for landslide susceptibility modelling which is reproducible[1] and can be established as a baseline for further studies in other geographical areas. Landslide susceptibility map presented in Fig. 7 provides useful insights that can facilitate the surety in gazetting high landslide susceptibility areas and prevent urban development in such *high-risk* areas. Such measures will be helpful in the long term, especially with increasing demand for socio-economic infrastructure to support rural-urban migration demand. By referencing the landslide susceptibility maps produced using the methodological approach in this study, planners and decision-makers can make informed choices regarding land use planning and development to minimize landslide hazard risks and safeguard the communities living in susceptible areas.

**Acknowledgments.** We are grateful to Professor Chris Poskitt from Singapore Management University, School of Computing and Information Systems. We also thank anonymous peer-reviewers for their comments.

**Disclosure of Interests.** The authors have no competing interests to declare that are relevant to the content of this article.

## References

1. Aleotti, P., Chowdhury, R.: Landslide hazard assessment: summary review and new perspectives. Bull. Eng. Geol. Environ. **58**, 21–44 (1999)
2. Wang, H., Zhang, L., Yin, K., Luo, H., Li, J.: Landslide identification using machine learning. Geosci. Front. **12**, 351–364 (2021)
3. Cendrero, A., Forte, L.M., Remondo, J., Cuesta-Albertos, J.A.: Anthropocene geomorphic change. Climate or human activities? Earth's Future **8** (2020)
4. Froude, M.J., Petley, D.N.: Global fatal landslide occurrence from 2004 to 2016. Nat. Hazards Earth Syst. Sci. **18**, 2161–2181 (2018)
5. Shrestha, A.B., Ezee, G.C., Adhikary, R.P., Rai, S.K.: Resource manual on flash flood risk management; module 3 - structural measures (2012)
6. Popescu, M.E., Sasahara, K.: Engineering measures for landslide disaster mitigation. In: Sassa, K., Canuti, P. (eds.) Landslides – Disaster Risk Reduction, pp. 609–631. Springer, Berlin, Heidelberg (2009). https://doi.org/10.1007/978-3-540-69970-5_32
7. Regmi, N.R., Giardino, J.R., McDonald, E.V., Vitek, J.D.: A comparison of logistic regression-based models of susceptibility to landslides in western Colorado, USA. Landslides **11**, 247–262 (2013)
8. Achu, A.L., et al.: Machine-learning based landslide susceptibility modelling with emphasis on uncertainty analysis. Geosci. Front. **14**, 101657 (2023)
9. Steger, S., Kofler, C.: Statistical modeling of landslides: landslide susceptibility and beyond. In: Pourghasemi, H.R. and Gokceoglu, C. (eds.) Spatial Modeling in GIS and R for Earth and Environmental Sciences, pp. 519–546. Elsevier, Amsterdam, Netherlands (2019)

[1] Researchers interested in further investigation can access the complete code and methodology framework employed in this study. Please direct requests to the corresponding author.

10. Moore, I.D., O'Loughlin, E.M., Burch, G.J.: A contour-based topographic model for hydrological and ecological applications. Earth Surf. Process. Landf. **13**(4), 305–320 (1988)
11. Kuriakose, S.L., van Beek, L.P., van Westen, C.J.: Parameterizing a physically based shallow landslide model in a data poor region. Earth. Surf. Proc. Land. **34**, 867–881 (2009)
12. Gaidzik, K., Ramírez-Herrera, M.T.: The importance of input data on landslide susceptibility mapping. Sci. Rep. **11** (2021)
13. Castellanos Abella, E.A., Van Westen, C.J.: Qualitative landslide susceptibility assessment by multicriteria analysis: a case study from San Antonio del Sur, Guantánamo, Cuba. Geomorphology **94**, 453–466 (2008)
14. Ruff, M., Czurda, K.: Landslide susceptibility analysis with a heuristic approach in the eastern alps (Vorarlberg, Austria). Geomorphology **94**, 314–324 (2008)
15. van Westen, C.J., Rengers, N., Soeters, R.: Use of geomorphological information in indirect landslide susceptibility assessment. Nat. Hazards **30**, 399–419 (2003)
16. Boussouf, S., Fernández, T., Hart, A.B.: Landslide susceptibility mapping using maximum entropy (Maxent) and geographically weighted logistic regression (GWLR) models in the Río Aguas catchment (Almería, SE Spain). Nat. Hazards **117**, 207–235 (2023)
17. Guzzetti, F., Carrara, A., Cardinali, M., Reichenbach, P.: Landslide hazard evaluation: a review of current techniques and their application in a multi-scale study, Central Italy. Geomorphology **31**, 181–216 (1999)
18. Saha, S., et al.: Comparison between deep learning and tree-based machine learning approaches for landslide susceptibility mapping. Water **13**, 2664 (2021)
19. Wu, W., Guo, S., Shao, Z.: Landslide risk evaluation and its causative factors in typical mountain environment of China: a case study of Yunfu City. Ecol. Indic. **154** (2023)
20. Ozturk, U., Bozzolan, E., Holcombe, E.A., Shukla, R., Pianosi, F., Wagener, T.: How climate change and unplanned urban sprawl bring more landslides (2022)
21. Tobler, W.R.: A computer movie simulating urban growth in the Detroit Region. Econ. Geogr. **46**, 234 (1970)
22. Fotheringham, A.S., Brunsdon, C., Charlton, M.: Geographically Weighted Regression the Analysis of Spatially Varying Relationships. Wiley, Hoboken (2010)
23. Sun, X., Chen, J., Bao, Y., Han, X., Zhan, J., Peng, W.: Landslide susceptibility mapping using logistic regression analysis along the Jinsha River and its tributaries close to Derong and Deqin County, southwestern China. ISPRS Int. J. Geo Inf. **7**, 438 (2018)
24. Matsche, D.T.: A Geographically Weighted Regression Approach to Landslide Susceptibility Modeling. Theses and Dissertations Collection, University of Idaho Library Digital Collections (2017)
25. Matthews, S.A., Yang, T.-C.: Mapping the results of local statistics: using geographically weighted regression. Demogr. Res. **26**, 151–166 (2012)
26. Azzoni, A., Chiesa, S., Frassoni, A., Govi, M.: The valpola landslide. Eng. Geol. **33**, 59–70 (1992)
27. Camera, C., Apuani, T., Masetti, M.: Modeling the stability of terraced slopes: an approach from Valtellina (Northern Italy). Environ. Earth Sci. **74**, 855–868 (2015)
28. Alexander, D.: Valtellina landslide and flood emergency, Northern Italy, 1987. Disasters **12**, 212–222 (1988)
29. Luino, F., et al.: The role of soil type in triggering shallow landslides in the Alps (Lombardy, northern Italy). Land. **11**, 1125 (2022)
30. Lu, J., et al.: Investigation of landslide susceptibility decision mechanisms in different ensemble-based machine learning models with various types of factor data. Sustainability. **15**, 13563 (2023)
31. Xu, Q., Yordanov, V., Brovelli, M.A.: Landslide Influencing Factors for Landslide Susceptibility Mapping in Lombardy, Italy. Zenodo (2023)

32. Hussin, H.Y., et al.: Different landslide sampling strategies in a grid-based bi-variate statistical susceptibility model. Geomorphology **253**, 508–523 (2016)
33. Heinze, G., Schemper, M.: A solution to the problem of separation in logistic regression. Stat. Med. **21**, 2409–2419 (2002)
34. Bucci, F., et al.: A new digital lithological map of Italy at the 1:100 000 scale for geomechanical modelling. Earth Syst. Sci. Data. **14**, 4129–4151 (2022)
35. Li, Y., Huang, S., Li, J., Huang, J., Wang, W.: Spatial non-stationarity-based landslide susceptibility assessment using PCAMGWR model. Water **14**, 881 (2022)
36. Gollini, I., Lu, B., Charlton, M., Brunsdon, C., Harris, P.: GWMode: an r package for exploring spatial heterogeneity using geographically weighted models. J. Stat. Softw. **63** (2015)