Author Preprint
To cite this article: Puente-Sotomayor, F., Mustafa, A. & Teller, J. Landslide Susceptibility Mapping of Urban Areas: Logistic Regression and Sensitivity Analysis applied to Quito, Ecuador. Geoenviron Disasters 8, 19 (2021). https://doi.org/10.1186/s40677-021-00184-0

Landslide Susceptibility Mapping of Urban Areas: Logistic Regression and Sensitivity Analysis applied to Quito, Ecuador

Abstract

Although the Andean region is one of the most landslide-susceptible areas in the world, limited attention has been devoted to the topic in terms of research, risk reduction practice, and urban policy. Based on the collection of early landslide data for the Andean city of Quito, Ecuador, this article aims to explore the predictive power of a binary logistic regression model (LOGIT) to test secondary data and an official multicriteria evaluation model for landslide susceptibility in this urban area. Cell size resampling scenarios were explored as a parameter, as the inclusion of new "urban" factors. Furthermore, two types of sensitivity analysis (SA), univariate and Monte Carlo methods, were applied to improve the calibration of the LOGIT model. A Kolmogorov–Smirnov (K-S) test was included to measure the classification power of the models. Charts of the three SA methods helped to visualize the sensitivity of factors in the models. The Area Under the Curve (AUC) was a common metric for validation in this research. Among the ten factors included in the model to help explain landslide susceptibility in the context of Quito, results showed that population and street/road density, as novel "urban factors", have relevant predicting power for high landslide susceptibility in urban areas when adopting data standardization based on weights assigned by experts. The LOGIT was validated with an AUC of 0.79. Sensitivity analyses suggested that calibrations of the best-performance reference model would improve its AUC by up to 0.53%. Further experimentation regarding other methods of data pre-processing and a finer level of disaggregation of input data are suggested. In terms of policy design, the LOGIT model coefficient values suggest the need for deep analysis of the impacts of urban features, such as population, road density, building footprint, and floor area, at a household scale, on the generation of landslide susceptibility in Andean cities such as Quito. This would help improve the zoning for landslide risk reduction, considering the safety, social and economic impacts that this practice may produce.

Keywords: landslide susceptibility; Quito; LOGIT; sensitivity analysis; Kolmogorov-Smirnov test; Andean cities; urban factors

1. Introduction

1.1. Urban Landslides in the Andes

The Andes is a sub-region located in western South America near the Pacific Ocean, named after the presence of the Andean mountains. The Andean mountain range is among the cordilleras with the highest elevations in the world and part of the Pacific Ring of Fire, a global mountain system characterized by frequent volcanic eruptions and earthquakes (Blanchard-Boehm, 2004). The range crosses the territories of Colombia, Ecuador, Perú, Bolivia, Chile, and—with less significance for human settlements—Argentina and Venezuela. The Andes orography is of particular concern in terms of sustainable urban development because it has been subject to significant urbanization processes in recent decades, at an average of 20 m^2 per minute, particularly informal and diverse in typologies (Inostroza, 2017). This growth includes metropolises such as Bogotá, Santiago, and Lima. Other medium-size cities, such as Medellin, Quito, and Cali, and smaller cities, also must be considered in terms of urban population growth. This urbanization has been induced by country–city migration and natural growth, for which housing and urban development—mostly informal and contributing to urban poverty—is a challenge to planning and management at all governmental levels, which have had to shift their policy paradigm (Blanchard-Boehm, 2004; van Lindert, 2016). Furthermore, cities are subject to urban risk in the Andes, where one of the most frequent risks, with high accumulated impact, is landslides.

Susceptibility analysis of landslide risk (LRisk) has been broadly studied in case studies at regional scales and mostly covering rural areas, often involving natural conditions and, to a lesser extent, considering anthropic factors, such as road network and urban areas. However, urbanization is often treated only as a generic land-use category, without further detail. Against this background, LRisk is one of the main concerns for urban development in the Andes in light of physical and social factors. The geodynamics of this region make it prone to landslides. This condition is aggravated by climate change, in addition to the extreme events produced by El Niño climatic phenomena, which affect diverse locations of the region with an irregular time cycle. In some of these areas, urbanization has expanded rapidly in recent decades, as mentioned above. Cities in the Andes account for 70% of the population and the share of the urban population continues to grow rapidly. Unplanned urbanization is developing without any consideration of LRisks and governmental bodies have limited capacities to manage urban development (Comunidad Andina, 2017; D'Ercole, Hardy, Metzger, & Robert, 2009; UNISDR, 2018). Evidence regarding landslide-prone conditions in the region has been presented by Kirschbaum & Stanley (2018) and Sepúlveda & Petley (2015). These studies portray the concentration of landslide-susceptible areas in the irregular orography of Colombia, Ecuador, and Peru, with hundreds of fatalities. By comparison, fewer fatalities have occurred in neighboring countries in South America, with the exception of Brazil. Accordingly, disaster risk management (DRM) should be better integrated with land-use planning (LUP) for appropriate diagnostics and effective prevention of risks related to landslides.

1.2. Theoretical Background

1.2.1. Key definitions

A landslide is defined as: “the downslope movement of soil, rock, and organic materials under the effects of gravity” (Highland & Bobrowsky, 2008, p. 4). Its types include slides, falls, topples, flows, and lateral spreads, and combinations of these, whose causes can be geological, morphological, or anthropic (shaping of built or natural landscapes), which can be triggered by water, seismic, and volcanic activities (GEMMA, 2007; USGS, 2004). In complement, landslide disaster risk is the combination of natural hazard conditions, such as weak soil, intense precipitation, and earthquakes; vulnerability, such as soil cuts and fills, or structural weakness; and, exposure, such as construction LRisk-prone areas, as illustrated by Puente-Sotomayor, Egas, & Teller (2021).

This understanding of LRisk is directly related to landslide susceptibility, which beyond the definition of disaster risk as a social product, aims to identify the interaction between natural and built components, which is susceptible to landslides. By comparison, vulnerability—due to a closer relationship to anthropic action on land—can lead to analysis at minor scales. Anthropic vulnerability factors have less been taken into account in LSM; such in the case of road networks, specific urban land uses, and other human settlement features. The latter is a particular focus of attention for this study because extensive landslide disasters are produced in cities. However, few case studies address LSM in urban areas, such as reported by Bathrellos, Kalivas, & Skilodimou (2009); Dragićević, Lai, & Balram (2014); Klimeš & Rios Escobar (2010); Lara, Sepúlveda, Celis, Rebolledo, & Ceballos (2018); Lee, Baek, Jung, & Lee (2020). Furthermore, of these, few consider vulnerability-related factors at a fine level, such as population, urban street networks, and urban structures (buildings), as noted in Table 1.

Reichenbach, Rossi, Malamud, Mihir, & Guzzetti (2018) define landslide susceptibility as the probability of incidence in a determined terrain relying on specific factors, including climate. These authors distinguish susceptibility from threat/hazard or vulnerability analyses in that the former is analyzed at a large scale and the data is acquired and processed at an aggregate level. Reichenbach et al. (2018) conclude from a global review of LSM that the usual determinant factors are slope, geology, and aspect; of these, the first two have a higher influence on the prediction power of models. They also state that results may vary according to methodologies, model validation, landslide types, triggering factors, and researcher background. Other studies include precipitation, population density, and land use as significant factors (Hemasinghe, Rangali, Deshapriya, & Samarakoon, 2018; Sepúlveda & Petley, 2015).

1.2.2. A brief review on LSM

Reichenbach et al. (2018) classify landslide susceptibility assessment into five groups, namely: (i) geomorphological mapping; (ii) analysis of landslide inventories; (iii) heuristic or index-based approaches; (iv) process-based methods; and (v) statistical modelling methods. The present work combines the heuristic approach officially adopted by the municipality of Quito as a preliminary input.

Modelling approaches

A number of machine learning modelling techniques have been developed and applied in diverse locations globally to achieve the finest possible precision—each time with more sophistication—to provide better inputs into LRR policy and planning. Among the most used LSM techniques during the past two decades are multi-criteria evaluation (MCE), analytical hierarchical process (AHP), weighted linear combination (WLC), logistic regression (LR), data-driven frequency ratio (FR), random forest (RF), support vector machines (SVM), and artificial neural networks (ANN) (see Table 1). Most of the applied techniques have been proven to provide accurate results and differentiated advantages, sufficient for LSM practices and, therefore, LRR zoning policies. For instance, a comparison between LR, SVM, and RF applied to the Sihjhong watershed, Taiwan, found RF performed best, whereas LR ran faster (Chang, Merghadi, Yunus, Pham, & Dou, 2019), which can be useful for large datasets, as in the present work. Regardless of the method adopted, research on LSM for Andean cities and regions is generally very limited and a few application cases have been recently published (Puente-Sotomayor et al., 2021).

Modelling for LSM has prompted a discussion on the impact of different parameters of the process have on the accuracy of the produced results. In addition to the modelling technique used, these parameters include data preprocessing, scale or cell/pixel size of input factors, and the type and number of factors used for the modelling. Table 1 shows a brief comparison of the described parameters among previous studies.

Relevant LSM parameters

Among the most relevant parameters is the preprocessing of data. There are different techniques to make factors comparable. These include different methods of normalization or standardization. Terminology varies. This study adopted the source assignation of weights in a discrete scale, also called weight-encoding, and data discretization using a percentile scale. Although standardization of weighted data for landslide susceptibility mapping is still an open discussion (Ronchetti et al., 2013), it is considered a valid option whenever intervals between ordinal categories are considered equal, regardless of statistical limitations such as the limited number of categories and overestimation of statistical power (Norman, 2010; Pasta, 2009; Williams, 2019).

Regarding the factor parameters for an urban LSM, it must be noted that only five of the twenty reviewed studies relate to urban areas. However, even in these cases, the factors are similar to those applied in the other works, often covering regional scales and rarely involving cities, which the current work aims to analyze. Therefore, few previous studies include human/urban related factors, such as the buildings (see factors column in Table 1), population, and urban road networks, as applied to this work. It is relevant to note that, unless an specific factor approach or restraints on the availability of data are stated, the most considered factors are topography/digital terrain model (DTM) derivatives (primarily slope angle, aspect, elevation, and curvature), annual precipitation, geology (primarily lithology and land use/vegetation coverage), distance to roads, hydrology (primarily distance to drainage, density, and topographic wetness index (TWI), which also relates to topography) and distance to faults in the seismicity. Furthermore, beyond the possibility of including a large number of factors, this does not necessarily mean better performance of a model and the optimal number of factors in a LSM is still debatable (Filippo Catani, Lagomarsino, Segoni, & Tofani, 2013).

Another discussed parameter in the literature is the resolution at which the input data is set. Table 1 also includes this parameter for each revised work. Resolutions vary from 1 to 500 m cell sizes. Regardless of the restraints based on the availability, reliability of data, and the context itself, some studies have tested the sensitivity of this parameter with diverse approaches and results. For instance, Chang et al., (2019) concluded that the finest resolution of topographic data does not necessarily result in the best performance of a model. Regarding DTM derivatives, Pawluszek et al. (2018) found in an example case that the optimal resolution was 30 m, classified by SVM. Another case proved that the 50 m resolution contributed best to the performance of an RF technique (F. Catani, Lagomarsino, Segoni, & Tofani, 2013a). Additional points of view state that different land-surface factors also have different optimal scales, and that the applied modelling technique may also influence this parameterization. Therefore, multiscale approaches are recommended for better performance in complex terrain settings (Filippo Catani et al., 2013; Sîrbu, Drăguț, Oguchi, Hayakawa, & Micu, 2019). This has also been corroborated by Dragićević et al. (2014), who examined these types of complex and multi-scalar contexts from regional to municipal and local scales.

Once the data is pre-processed and made available for modelling, one common and simple theoretical model used is multi-criteria evaluation (MCE), which can be combined with sensitivity analysis, such as in

Feizizadeh & Blaschke (2014) and Orán Cáceres et al. (2010), and explained below. A complementary approach to MCE is binary logistic regression (LOGIT), which is among the most used statistical methods for landslide susceptibility mapping (Reichenbach et al., 2018). This type of model helps to test weighted models, which do not support the assessment of the probability of landslide occurrences (Lombardo & Mai, 2018). For this research, a LOGIT was applied, followed by an SA.

Table 1. Review on landslide susceptibility modelling (LSM) modelling techniques, factors and inputs resolution used.

				Topography / DTM derivatives								Rainfall / Climate				Geology / Land Cover						Roads		Hydrology					Seismicity			Anthropic, other		
Reference code	**Approach, Models, Techniques**	**Cell Size for resolution (m)**	**Year**	Elevation / height /scarp	Slope angle / gradient	Slope aspect / sunlight exposition / orientation	Curvatures	Topographic Position Index (TPI)	Topographic Roughness Index (TRI)	Slope / Flow Direction	Other DTM derivatives ***	Annual Precipitation	Intense Precipitations	Long Precipitations	Potential rain effect (PRE)	Lithology / Geology	Geo(morpho)logical Units	Land Use / Vegetation Cover	Soils / textures	Normalized difference vegetation index (NDVI)	Potential erosion index (PIX)	Distance to roads	Road Density	Distance to drainage / streams	River / Gully Density	Flow Accumulation**	Topographic Wetness Index (TWI)**	Stream Power Index (SPI)	Distance to faults	Ground Peak Acceleration	Seismic Intensity / other Tectonic Features	Artificial Intervention / structures density	Declared flows	Random Value / other
1	FR, MCE+AHP, OB	10	2020	✓	✓	✓						✓				✓	✓	✓				✓		✓					✓					
2	AHP, WLC - LSI	10	2020		✓	✓						✓				✓		✓				✓		✓					✓					
3^{+}	DL	10	2020	✓	✓	✓	✓				✓							✓	✓								✓	✓						
4	FR, IV, CF, LR	30	2020		✓	✓	✓					✓				✓		✓	✓	✓				✓	✓				✓					
5	LR, RF, SVM	5-30	2019	✓	✓	✓	✓	✓	✓		✓	✓				✓						✓		✓										
6	LR, RF	4-50	2019	✓	✓		✓	✓	✓	✓																								
7	AHP, FR, MCE	30	2019	✓	✓	✓						✓				✓		✓				✓		✓					✓					
8	JT, SI	15	2019		✓	✓	✓								✓						✓		✓		✓							✓		
9	SA – PBA, ML, FFNN, SVM	1-30	2018	✓	✓	✓	✓	✓		✓	✓																							
10^{+}	WLC	9	2018		✓	✓					✓					✓		✓							✓	✓						✓	✓	
11	IV, LG	30	2017	✓	✓	✓										✓		✓				✓		✓					✓					
12	AHP, WLC, MCQ	10	2015		✓	✓						✓				✓		✓	✓	✓		✓		✓					✓					

13	AHP, WLC, OWA	20	2014	✓	✓	✓						✓				✓		✓				✓		✓					✓					
14+	MCE, AHP, WLC	1, 10, 50	2014	✓	✓	✓																✓		✓			✓	✓						
15*	SA – LCVs, MURs, LCVs, RF	10-500	2013	✓	✓	✓	✓				✓		✓			✓		✓				✓		✓		✓	✓		✓					✓
16	ANN	20	2013	✓	✓	✓	✓									✓		✓									✓							
17	RF, SA	10, 20, 50, 100, 250, 500	2013	✓	✓	✓	✓																			✓	✓							
18	SA - ANN	50	2011		✓					✓	✓		✓			✓	✓	✓	✓			✓			✓				✓	✓				
19+	MCE, WLC	5	2010		✓											✓		✓																
20+	WeF, MuF	20	2009		✓							✓				✓		✓					✓		✓						✓			
Total				12	20	16	9	3	2	3	6	8	2	0	1	14	2	14	4	2	1	10	2	10	5	3	5	2	9	1	1	2	1	1

+ Correspond to LSM studies in urban areas
* This modelling presented 35 factors; a summarized version of 13 is presented in this table.
** For this classification, these factors have been grouped in the hydrology. However, they are also considered in the topography.
*** Other DTM derivatives include: openness, side exposure index, hillshade, flow direction, and roughness.

Notes:

1. Diversity of factors for modelling have been classified by authors, criteria may vary.
2. The selection of factors varies according to approaches. For instance, 7, 8, and 9 involve a wide range of topography-related factors.

References:

1. Gudiyangada Nachappa, Kienberger, Meena, Hölbling, & Blaschke (2020), 2. Psomiadis, Papazachariou, Soulis, Alexiou, & Charalampopoulos (2020), 3. Lee, Baek, Jung, & Lee (2020), 4. Wubalem (2020), 5. Chang et al. (2019), 6. Sîrbu, Drăguț, Oguchi, Hayakawa, & Micu (2019), 7. Meena, Ghorbanzadeh, & Blaschke (2019), 8. Ramos-Bernal, Vázquez-Jiménez, Sánchez Tizapa, & Arroyo Matus (2019), 9. Pawluszek, Borkowski, & Tarolli (2018), 10. Lara, Sepúlveda, Celis, Rebolledo, & Ceballos (2018), 11. G. liang Du, Zhang, Iqbal, Yang, & Yao (2017), 12. Shahabi & Hashim (2015), 13. Feizizadeh & Blaschke (2014), 14. Dragićević, Lai, & Balram (2014), 15. Catani, Lagomarsino, Segoni, & Tofani (2013b), 16. Pascale et al. (2013), 17. Catani, Lagomarsino, Segoni, & Tofani (2013a), 18. Melchiorre, Castellanos Abella, van Westen, & Matteucci (2011), 19. Klimeš & Rios Escobar (2010), 20. Bathrellos, Kalivas, & Skilodimou (2009).

Abbreviations:

AHP = Analytical Hierarchical Process, ANN = Artificial Neural Networks, CF = Certainty Factor, DL = Deep Learning, DTM = Digital Terrain Model, FFNN = Feed-Forward Neural Network, FR = Data-Driven Frequency Ratio, IV = Information Value, JT = Jackknife Test, LCVs = Landslide Conditioning Factors, LR = Logistic Regression, LSI = Landslide Susceptibility Index, MCE = Multi-Criteria Evaluation, ML = Maximum Likelihood, MuF = Multiple Factor Model, MURs = Mapping Unit Resolutions, NDVI = Normalized Difference Vegetation Index, OWA = Ordered Weighted Average, PBA = Pixel Based Approaches, RF = Random Forest, SA = Sensitivity Analysis, SI = Susceptibility Index, SVM = Support Vector Machine, WeF = Weight Factor Model, WLC = Weighted Linear Combination.

1.2.3. Sensitivity Analysis

A further step in evaluating LOGIT models is to apply a sensitivity analysis (SA) (Reichenbach et al., 2018). SA is applied to determine the contribution of input parameters to the accuracy of the model prediction appraised in its outputs (Abbaszadeh Shahri, Spross, Johansson, & Larsson, 2019; Poelmans & Van Rompaey, 2010). The objective of sensitivity analysis is to help adjust the calibration of the studied parameters involved in the LSM model to improve its predicting/classification power. Among different methodologies, two stand out: the simple, univariate, or "one-at-a-time" (OAT) method, and the stochastic/random-selection method, also called "Monte Carlo", whose applications vary according to the needs of the field of practice (Bouyer, 2009). Details of both methods are explained in sections 2.7, 2.9, 3.3 and 5.4.

2. Methods

2.1. Study area and its landslide risk-reduction policy background

Quito, the capital of Ecuador, is the most populated of two existing metropolitan districts (regions) in the country, with 2,781,641 inhabitants projected for 2020 (INEC, 2016). The jurisdiction of the Metropolitan District of Quito (DMQ) covers 4235.2 km^2, of which 10% is urban land with 286,412 housing units (MDMQ, 2015). As one of the Andean mountain cities, Quito has suffered from multiple natural threats, including landslides, volcano eruptions, floods, and earthquakes. Hence, exposure to risks has been further exacerbated, given the fast population growth and the uncontrolled urbanization process. Accordingly, Quito has collected geodata relating to landslide disaster events during the past two decades. This has strengthened the city's management capacities and its approach to preventive policies and actions (Rebotier, 2016), in addition to preparedness and response. Most recently, resiliency has been adopted as an urban policy, to the point of being institutionalized, with the creation of a Resiliency Department and the design of the city's resiliency strategy (Alcaldia del Distrito Metropolitano de Quito, 2018; MDMQ, 2017)

Landslide risk reduction (LRR) policies in Quito have a history of approximately one decade. They began with landslide-related land-use zoning as part of the local plan, in 2011 (Puente-Sotomayor, Villamarin, & Cevallos, 2018). Previously, building regulations included generic risk prevention measures, such as setbacks from ravines, slope borders, and rivers (Concejo del Distrito Metropolitano de Quito, 2003). For lahar-prone areas, a transfer of responsibility from government to users was used, applying a notarized responsibility to the owner for building on high-risk areas, prior to city approval (Concejo del Distrito Metropolitano de Quito, 2011). Since 2014, this has no longer been allowed in national laws, which assign criminal liability of the generated risks to any official that approves subdivisions or projects in risk zones (Asamblea Nacional del Ecuador, 2014). During the past decade, the landslide preventive/reductive approach was materialized by establishing the LRisk zone (ZR) category in the local LUP. Construction was strictly banned in ZR areas. This zoning policy intuitively and imprecisely combined slopes (at the 1:5000 scale), soil stability (at the 1:25,000 scale), and field inspections as the only inputs in 2011. The application of this regulation triggered around 40 complaints per year from users, who claimed they were affected socially, through the violation of their housing rights; and, economically, due to previous investments and rent expectations related to the property labeled at-risk. In 2013, a reform to this ordinance relaxed the policy by returning to owners the right to build on ZRs, who provided geotechnical studies that justified their projects. This revealed limitations of the technical capacity of users and officials, and the problem with defining "mitigation", which subsequently evidenced the poor accessibility to geotechnical risk relief for low socio-economic strata. A new reform in 2015 cancelled the ZR land-use category and converted it to an overlay map, which, in practice, did not change the policy (Puente-Sotomayor et al., 2018). By 2015, the first landslide susceptibility studies were produced. The outputs of these studies were expected to improve the ZR policy. However, they have not yet been articulated with the LUP. These studies' outputs have been labeled as official data and were used as part of the input data for this research.

A brief comparison between the ZR layer, for which the limits have not yet changed, the existing landslide susceptibility study (FUNEPSA et al., 2015), and a landslide events database from 2005 to 2017 (see **Error! Reference source not found.**), provided by the Metropolitan Emergency Operations Committee of Quito (COE-M), reveals inconsistencies between the policy, research, and facts. Only 8% of recorded landslide events are contained in ZR polygons, 25% of ZR do not match with the high and very-high susceptibility areas, and 81% of high and very-high susceptibility areas are not covered by the ZR polygons. This presumably means that vast areas should be considered as landslide-prone, while other areas, although smaller in proportion, probably do not need a protection policy (Puente-Sotomayor et al., 2018).

2.2. Purpose, objectives and scope

The main purpose of this study was to construct a reliable Landslide Susceptibility Mapping (LSM) production that can support LRR policies for urban Quito (see study area section 2.1). This represents an advance from previous actions regarding landslide preventive/reductive zoning in this city. For this purpose, specific objectives included examining the incidence of urban-related factors in LRisk susceptibility; providing quality inputs for better ZR delimitation, considering the implications for safety, housing rights, and the economy; and generating improvements and further discussion and action in the local LSM processes, by exploring different parameters such as modelling approaches, scale, data generation, and factor identification.

The practical scope of this study was to derive a landslide susceptibility map based on a binary logistic regression model combined with a sensitivity analysis (SA) to provide optimal calibration options for factor coefficients used in the model. Developing an evidence-based LSM that considers SA is an indispensable input for developing an urban policy backed by a socio-political consensus (Orán Cáceres, Gómez Delgado, & Bosque Sendra, 2010). The context of application will be the urban core of the DMQ, including its surrounding peri-urbanized areas, as described below.

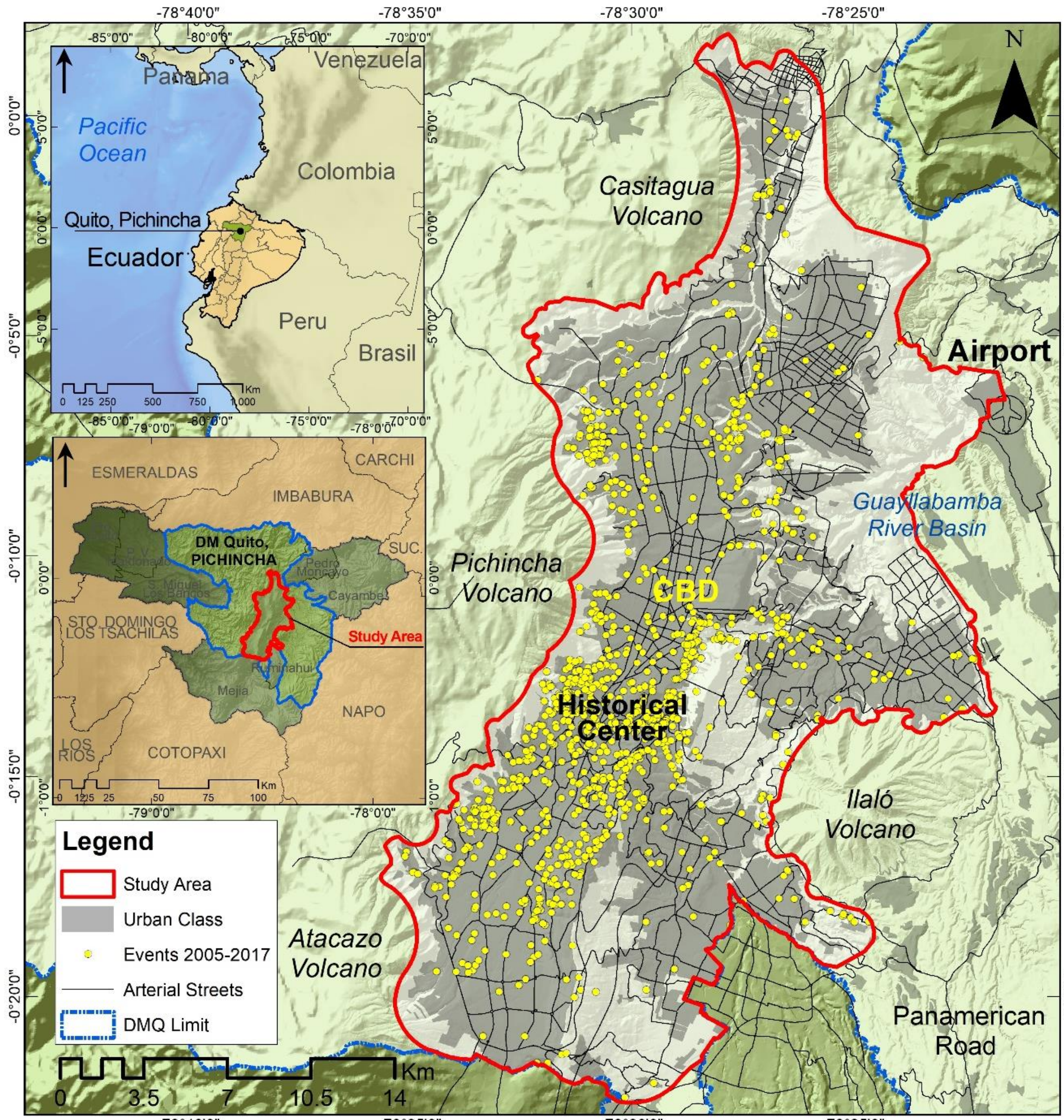


Figure 1. Study Area in the Metropolitan District of Quito, displaying the 2005-2017 period landslide events. Data Source: MDMQ, IGM Ecuador. *Legend: Study Area (red line), Urban Class (gray polygon), Landslide Events during the 2005-2017 period (yellow dots), Arterial Streets network (black continuous lines) and the Metropolitan District of Quito jurisdictional limit (dashed blue line).*

This research builds on data collected during a LRisk analysis produced by the municipality in 2015. This study delivered a theoretical weighted multi-criteria theoretical model including six factors that were surveyed and processed. These are slope, intense precipitations, soil stability after former large landslides, lithology, land use/vegetation coverage, and seismic intensity. Each factor had partial weights/susceptibilities proposed by local experts in the fields of geotechnics, meteorology, geography, disaster management, and seismology. The results of this model portrayed a landslide susceptibility map for Quito and its satellite "conurbated" areas (an approximate total area of 610 km^2) using the map algebra GIS tool to sum the partial weights as shown in **Figure 2** (FUNEPSA et al., 2015).

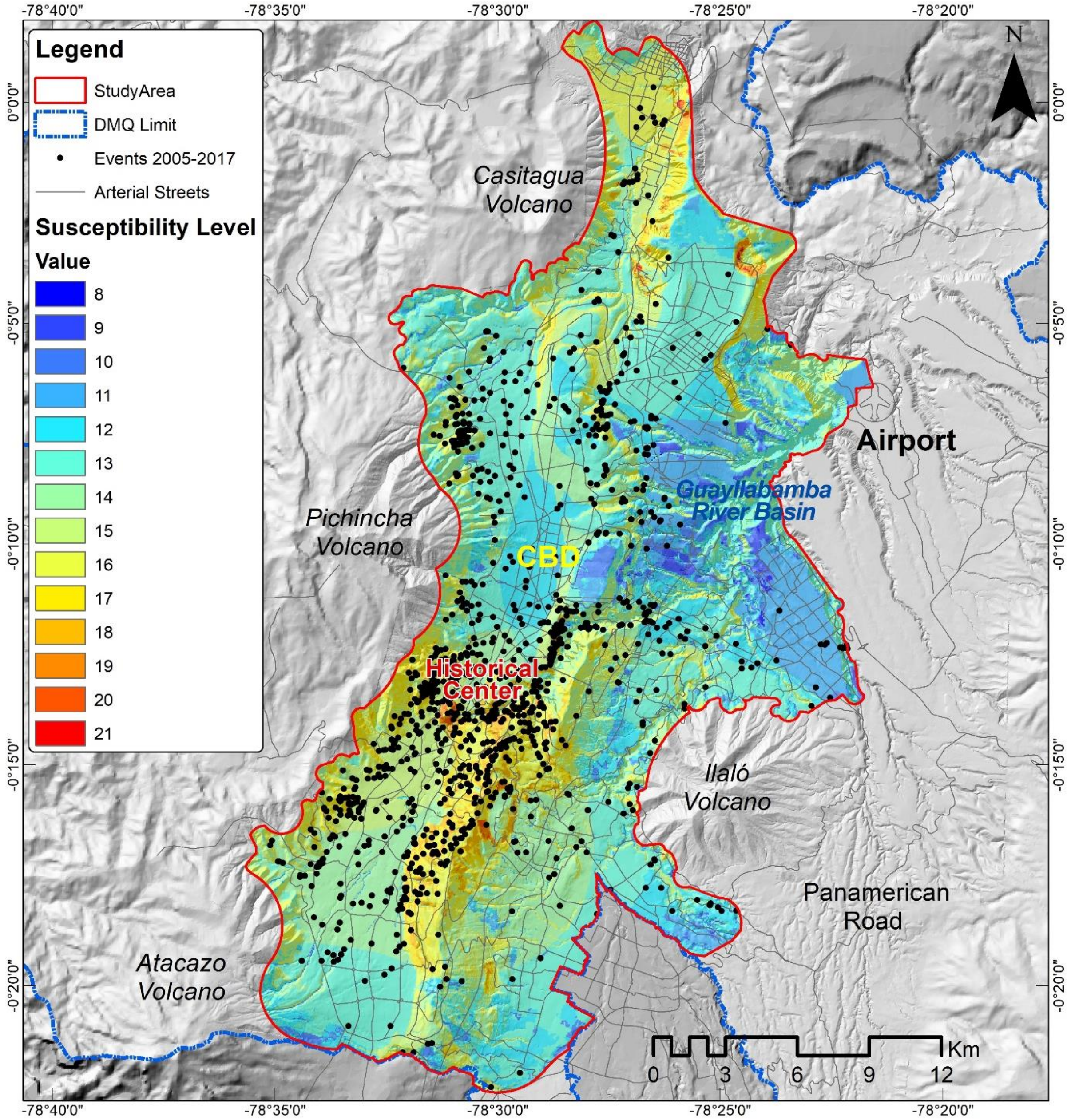


Figure 2. Landslide susceptibility map for Quito including 2005–2017 event spots. Data Source: Quito Municipality MDMQ. *Legend: Study Area (red line), Urban Class (gray polygon), Landslide Events during the 2005-2017 period (black dots), Arterial Streets network (black continuous lines), the Metropolitan District of Quito jurisdictional limit (dashed blue line) and eight to 21 landslide susceptibility levels (blue-yellow-red color spectrum).*

2.3. Inputs and Preprocessing

Initially, our study proposed to develop a binary logistic regression model on the basis of six factors identified by these experts, plus other related to the Quito urban settings, which we aimed to experiment with. A dataset of landslide events that occurred from 2005 to 2017 was therefore collected from the COE-M of Quito. This database includes around 1400 events, including rotational and translational landslides, flows, and topples, all considered generically as landslides (USGS, 2004). A minor limitation is that the dataset suffers from some underreporting, and unbalanced and unstructured elements.

From the data of the six initial factors, the first LOGIT was applied. Then, four additional factors were included in two steps to test the model. As a first addition, population, provided by the National Institute of Statistics and Census (INEC), and floor area, provided by the Quito municipality (MDMQ), were included to construct a second LOGIT. Then, road density and building footprint area, also provided by the MDMQ, were added to run the final LOGIT. All four additional factors were pre-processed and adapted for this research work. As explained in section 1.2.2, considering the urban context of Quito, where all of the landslide events were registered, this research aimed to determine the incidence of these factors on the results, whose content was more relevant to the urban context, i.e., buildings, streets, and population. Details of all of the ten factors included in this study are provided in Table 2.

Table 2. Input data for landslide susceptibility mapping in Quito

Code	Content	Disaggregation level	Specifications, type	Data Source	Year
bin	*Landslide Events*	Point	Binary	COE-M Quito	2005-2017
1geo	*Lithology*	50 m	Pre-discretized from categorical to weights (weighted classes)	FUNEPSA et al., 2015	2015
2cov	*Land Use / Vegetation Cover*	50 m	Pre-discretized from categorical to weights	FUNEPSA et al., 2015	2015
3sei	*Seismicity*	50 m	Pre-discretized from categorical to weights	FUNEPSA et al., 2015	2015
4pre	*Intense Precipitations* (in 24 hours)	50 m	Pre-discretized from continuous to weights	FUNEPSA et al., 2015	2015
5sta	*Soil Stability* (from former landslides)	50 m	Pre-discretized from categorical to weights	FUNEPSA et al., 2015	2015
6slo	*Slope*	50 m	Pre-discretized from continuous to weights	FUNEPSA et al., 2015	2015
7pop	*Population*	Block Scale	Continuous, discretized by natural breaks to weights	INEC	2010
8roa	*Streets (Roads) Density*	Street segment	Continuous, discretized by natural breaks to weights	STHV – MDMQ	2016
9bui	*Floor Area*	Building Scale	Continuous, discretized by natural breaks to weights	STHV – MDMQ	2017
10gro	*Building Footprint Area*	Building Scale	Continuous, discretized by natural breaks to weights	STHV – MDMQ	2017

NOTE: COE-M=Metropolitan Emergency Operations Committee of Quito, FUNEPSA et al., (2015) is the report containing the official surveyed data for Quito Municipality - Disaster Risk Management Division, INEC=Ecuadorian National Institute of Statistics and Census, STHV-MDMQ=Secretariat of Territory, Habitat and Housing of Quito Municipality.

2.4. Resolution

The dependent factor, landslide events (binary), and the ten independent, explanatory factors were pre-processed in raster files, with a cell size (disaggregation level) of 50 m. This was the resolution at which the lithology, land coverage, seismicity, precipitations, soil stability, and slope were previously provided by the municipality surveyors. Complementarily, the additional four "urban" factors were converted from a scale at which the detail of buildings, blocks, and streets was legible (1:1000 approximately), which was logically consistent with the 50 m cell size of the other six factors. Therefore, all of the datasets were standardized to this resolution. In this regard, the theoretical background review implied that, although scale may determine the modelling performance, performance is also dependent on the context and complexity of the process. For this study, although the secondary source input data was pre-processed at a 50 m cell size, an aggregation process through resampling GIS techniques (nearest neighbor mode) were applied to suit the complete datasets at resolutions of 100, 200, and 500 m before the application of the LR modelling for each resolution. Subsequently, the results provide a rationale to retain the original scale.

2.5. Standardization

To manage a standard scale of factor values before applying the LOGIT, the following process was undertaken. The binary layer of landslide events records one of two categories for each area unit or cell: true (or one), when one or more landslide events occurred in it; or, false (or zero), when no landslides occurred in it. Two standardization types proceeded. The first six factors provided by the municipality were previously converted from either classified continuous data or categorical data to weights in a discrete 1-to-4 scale. The additional four factors, all continuous, were classified using natural breaks in four classes, to correspond to the 1-to-4 scale. The details of this conversion are shown in Table 3. The complementary data, i.e., the additional four factors that were collected in vector format, were then converted into raster TIFFs to fit with the remainder of the dataset (initial six factors).

By looking at the data categories and their assigned weights, the contribution from the local experts in generating the datasets is notable. This can be seen, in particular, for factors such as lithology, land-use/vegetation coverage, and stability, whose information is specifically related to the Quito context. Furthermore, It is important to mention that, for the case of the slope factor, the class "greater than 50°" is weighted as 3 and not 4, as one would suppose. This is due to the fact that most of the local geology type, the "*cangahua*", found at this slope range and in most of the study area, is less susceptible to landslides than flatter slope ranges and has reported much less events, according to the local data mining experts. Regarding the rainfall, the meteorology surveyors stated that the Quito region is not affected by long-term persistent rainfall, which triggers landslides in other parts of the world, such as Central America or Southeast Asia, or in territories affected by El Niño phenomena extreme events, which are not regular in terms of time cycles and occur every 25 or 30 years. Instead, in Quito, landslides are more likely triggered by intense precipitation, which is why this variable was included instead of other climate-related factors, such as annual precipitation rates. Inclusive, long precipitation was not considered in the reviewed scientific articles, as shown in Table 1.

A second standardization process for the ten explanatory factors was applied and tested. It was based on a percentile discretization. The aim of applying percentile discretization was to have a finer value than the 1-to-4 scale. This also helped to correct a distortion existing in the provided data, produced by a marginal portion of outliers that widened the absolute value range of the dataset, which is an advantage of this discretization method (Grzenda, 2020). This distortion occurred particularly with the data of the floor area and building footprint area. Table 4 shows how the ten factors were discretized through percentiles. The categorical data weights were assigned to their corresponding percentile of the 1-to-100 scale, considering in it three equal segments (assuming intervals between weights as equal units in a discrete scale). For the remaining continuous data factors, the new values were simply the corresponding percentile.

Table 3. Conversion of factor classes or categories to weights (weights encoding) of input factors for LSM for Quito. Source: MDMQ, INEC, adapted by authors.

Code	**Weights (Partial Susceptibilities):**	**1**	**2**	**3**	**4**
1geo	*Slope* (degrees):	• 0° - 10°	• 10.1° - 25°	• 25.1° - 35° • > 50°	• 35.1° - 50°
2cov	*Lithology* (categories):	• Cotopaxi Lahars: steep ledges and slopes. Slopes and canyons or deep throats of ravines and rivers • Pululahua Domes: fractured dacites from the volcano, but they appear compact and with resistant weathering	• Quito Lake Deposits • Colluvial Mass Movements • Colluvial Conglomerates • Colluvials • Casitagua Volcanics • Undifferenced Volcanic Lahars • Pichincha Volcanics • Cangahua formations: compacted ashes, pyroclastic, lava	• Cangahua formations: undifferenced ashes-lapillistone with destroyed surfaces, strongly bisected in hills with rounded tips	• Alluvial terraces • Pululahua pyroclastic flows • Guayllabamba, San Miguel and Pisque Formations: sequences of volcanic sands, pyroclastic flows, silts, lahars; at the base: fluvial-lake sequence, occasionally very crumbly
3sei	*Land Use / Vegetation Cover* (categories):	• North Andes mountain bushes • Always-green North Andean high-mountain forests • Mountain pasture • High-mountain and mountain-moor grassland	• Reservoirs • Inter-Andean dry bushes • Inter-Andean dry forest • Eucalyptus forests • River vegetation from xerophytic mountain floor • Inter-Andean mountain Saxicola Vegetation	• Airport • Short-cycle crops • Cropped grass • Natural grass	• Quarries • Buildings • Eroded soils
4pre	*Soil Stability* (categories):	• Stabilized	• Latent	• Reactivated • Colluvial	• Active
5sta	*Intense Precipitations* in 24 hours (stations types: a. Maximum (mm) / n>10 Rt=100 years, b. Average rain (mm) / n<10 years):	• < 75.4 (a) • < 50 (b)	• 76.4–91.88 (a) • 51–90 (b)	• 91.88–107.34 (a) • 91–130 (b)	• 107.34–122.79 (a) • 137–175 (b)
6slo	*Seismic Intensity* (European Macro-seismic Scale, ordinal):	Not applicable	• EMS VII	• EMS VIII	Not applicable
7pop	*Population* (Inhabitants):	• 0–1.81	• 1.82–5.56	• 5.57–12.47	• 12.48–31.86
8roa	*Road Density* (m/Ha):	• 0–7.50	• 7.51–18.16	• 18.17–31.99	• 32–100.70
9bui	*Floor Area* (m^2):	• 0–4180.33	• 4180.34 – 35,950.76	• 35,950.77 – 104,508.01	• 104,508.02–213,196.34
10gro	*Building Footprint Area* (m^2):	• 0–3344.26	• 3344.27–34,278.63	• 34,278.64 – 104,508.01	• 104,508.02–213,196.34

NOTE: The data collected for intense precipitations was standardized by assigning partial susceptibilities (weights) to data provided by two types of meteorological stations: a. those with records of more than 10 years; and, b. those with records of less than 10 years. For the meteorology notation consider: n = period in years; Rt = return time. Further details on the calculation to complete the intense precipitations dataset are available in the FUNEPSA et al. report (2015)

Table 4. Conversion table of categorical data from weights to percentile normalization, and continuous data to percentiles.

Categorical Data Factors:	*Lithology* *Land Use/Vegetation Coverage* *Seismic intensities* *Intense Precipitations* *Soil Stability* after former landslide events	**Weights (Partial Susceptibilities) *:**	1	2	3	4
		Percentile Values:	1	33	67	100
Continuous Data Factors:	*Slope* *Population* *Road Density* *Floor Area* *Building Footprint Area*	**Weights (Partial Susceptibilities) **:**	1	2	3	4
		Percentile Values:	Corresponding percentile (from 1 to 100)			

* Assigned according to FUNEPSA et al. (2015)
** Classified by natural breaks, except for *slope*, classified according to FUNEPSA et al. (2015).

2.6. Logistic regression

Following the preparation of the data, the LOGIT proceeded. In regard to the sampling method, two sets of elements (the binary sample of cells) with equal number of items were then selected to test the LOGIT. The first set had cells that registered the occurrence of landslides (an average of 1.29 events per cell), in total, 1139 cells with a true/one value. The second set had cells that did not register landslides, i.e., false/zero value. Considering the 1139 true values, an equal number of false value cells were randomly chosen from more than 222,000 remaining equivalent value cells in the study area.

A generalized linear model regression function in the programming platform (MATLAB_R2018b) was then applied to obtain the values of the coefficients for all ten factors and the intercept of the function. With these values, the logistic regression (Equation 1) was applied to obtain the landslide susceptibility values for all cells for the study area. These values provide the probability of occurrence of a landslide, varying from 0 (null probability) to 1 (absolute occurrence). These values helped generate the reference landslide susceptibility map. This process was first undertaken for the six initial factors; second, with the addition of population and floor area as new factors; and finally, with the addition of road density and building footprint area. The area under the receiving operating characteristic (ROC) curve (AUC) was the performance indicator chosen for the LOGIT model, considering it is common in evaluation of the prediction accuracy of models for natural hazards (Abbaszadeh Shahri et al., 2019; Wang, Feng, Li, Ren, & Du, 2020).

Equation 1. Logistic regression function for landslide susceptibility mapping

$$ls = \frac{1}{1 + e^{-(b_0 + b_1 x_1 + b_2 x_2 + \ldots.. + b_n x_n)}}$$

where:

ls	=	landslide susceptibility: the probability of occurrence of a landslide (between 0 and 1)
e	=	the mathematical constant *e* (2.71828)
b_0	=	the intercept of the logistic function
b_n	=	the coefficient of factor x_n
x_n	=	the factor number n

2.7. Sensitivity Analysis – univariate method

After the generation of a referential susceptibility map and the validation of the LOGIT model that generated it using the AUC/ROC value, SA was performed to test the sensitivity of the model outputs to changes in one, many, or all selected parameters. For this research, the selected referential metric was the AUC value as the output for all of the generated simulations, as applied to SA by Poelmans & Van Rompaey (2010). Sensitivity analyses were performed using two methods.

The first was the simple, univariate, or OAT method, which is simple to apply and assess. It consisted in changing one "free" parameter of the model at a time to generate variations of the model, within a defined range and with a defined interval for the changes. In this case, while one factor changes its coefficient, the others remain unaffected (fixed parameters) and remain as the references. For the model used in this research, the parameters changed were each of the ten coefficients generated by the LOGIT model. A set of multipliers ranging from 0.1 to 20 with an interval of 0.1 modified each of the coefficients of the ten factors, one at a time, to generate a total of 2000 susceptibility maps, from which AUC values (outputs) were generated and plotted. The AUC values higher than the reference AUC value (the first generated) indicates that their corresponding models are better calibrated than the reference. This was reproduced for the weights-encoded and the percentile-discretized models.

2.8. Kolmogorov–Smirnov test for sensitivity

A two-sample Kolmogorov–Smirnov (K-S) test was applied to the univariate test results for both weights and percentile-based discretization methods as another means to determine the sensitivity of the factors. As a metric, the D-statistic (also called the KS-statistic) values are provided, indicating the D-critical value, as calculated using Equation 2. These provide a more insightful picture than the *p*-values of the same test, which considered an alpha value of 0.05 and were also calculated. The K-S tests were tabulated using the empirical distribution functions of two samples—ones and zeros—from each resulting map derived from the changes of the simple/univariate method, i.e., the distribution of the cell values corresponding to event occurrence cells (1139 observations/elements) compared to a distribution of a randomly selected similar number of non-occurrence cells.

Equation 2. Calculation of the D-critical value for a two-sample K-S test.

$$D_{\alpha} = c(\alpha)\sqrt{\frac{n_1 + n_2}{n_1 n_2}}$$

where:

$D\alpha$	=	D-critical value of the K-S test at an alpha value α
α	=	Alpha value determined for the K-S test (0.05 for this case)
$c(\alpha)$	=	constant based on α (1.36 for this case)
n_1	=	first sample size (1139 for this case)
n_2	=	second sample size (1139 for this case)

2.9. Sensitivity Analysis—Monte Carlo method

A second method to test the sensitivity based on factors used random variations for all of the factors, from one to all at a time, also within a defined range and with a defined interval. This is also called the Monte Carlo or stochastic method. For this research, multipliers of one or more coefficients at a time ranged from 0.1 to 5 with an interval of 0.1. The number of simulations for this random selection of possibilities was set to 8000, which may vary in replications of this study, according to the computer's processing capacities. Once again, AUC values (outputs) were generated and those higher than the reference AUC value indicated that their corresponding models with their modified coefficients' values had a better performance calibration than the reference itself (Bouyer, 2009). To better illustrate this, a table of random simulation calibrations is provided in the results, plus a chart showing the two best predictor factors.

This test was applied only to the results of the simple sensitivity analysis due to limitations in computer processing capacities and simplicity of visualization in charts, which provided for better communication of results.

2.10. Methodology summary and software used

To summarize, Table 5 shows all of the methodology and specific tools applied for this research.

Table 5. Methodology summary and applied tools

	6-Factor Model	8-Factor Model	10-Factor Model	
	Weights-encoding	Weights-encoding	Weights-encoding	Percentile-discretization
LOGIT and referential landslide susceptibility map and AUC value	✓	✓	✓	✓
Resolution tests of 50, 100, 200 and 500 m cell size, through LOGIT and AUC value			✓	
Sensitivity Analysis – Univariate method, variations plot and AUC values higher than reference			✓	✓
K-S test			✓	✓
Sensitivity Analysis – Monte Carlo Method, visualization of 2 best predictors and AUC values higher than reference			✓	✓

Regarding the software packages used to process data for this research work, GIS software (ArcMap 10.3) was applied to produce all maps using the integration, transformation, and geoprocessing tools, and conversion of shapefiles into raster TIFF files to make them suitable for calculation for the LOGIT model and the SA. A programming platform for matrix analysis (MATLAB_R2018b) was used for the SA, for which, particularly the generalized linear model regression, (glmfit function) and the AUC value computation (perfcurve function), were applied. For the two-sample K-S test, the function used was kstest2. The subsequent outputs were TIFF files for mapping and CSV datasets to produce graphs and charts in a spreadsheet package (Microsoft Excel) and a presentation/illustration package (PowerPoint). Tables were adjusted to a word processing package (Microsoft Word) format. The input TIFF files and the programming platform (MATLAB_R2018b) code are provided as Supplementary Material to this research work.

3. Results

4.1. LOGIT results by adding urban factors and a standardization variant

The first results portray the changes regarding the addition of factors, departing from the six-factor initial LOGIT model, which corresponds to the map shown in **Figure 2**, which delivered weight scores from eight (lowest landslide susceptibility) to 21 (highest landslide susceptibility). This considered six factors: lithology, land use/vegetation coverage, seismic intensities, intense precipitations, soil stability after large events, and slope. Subsequently, the LOGIT was tested with eight factors and finally with 10 factors. The eight-factor model included two more factors: population and floor area, while retaining the weights encoding (continuous factors classified by natural breaks). The 10-factor model included two more factors: road density and building footprint area, also weights-encoded. As new factors were added, the coefficients with the highest values changed their relative descending order (from highest to lowest value). For the 10-factor model, a variation by percentile-discretization of factor values was applied, which delivered a fourth set of results. The four models' results, including their corresponding AUC value, can be seen in Table 6. The initial MCE official map and the 6, 8, and 10-factor reference maps (weights and percentile standardizations), can be seen in **Figure 3**. The last two maps (d and e) in this figure were used for SA.

Table 6. Output values from LOGIT modelling for landslide susceptibility in Quito

Code	**Factor**	**6-Factor Model**		**8-Factor Model**		**10-Factor Model**			
		Weights-encoding		Weights-encoding		Percentile-discretization			
		Coefficient (β value)	Descending Order	Coefficient (β value)	Descending Order	Coefficient (β value)	Descending Order	Coefficient (β value)	Descending Order
0int	Intercept	−0.5281		−10.7830		−4.1375		−2.4317	
1geo	*Lithology*	0.3756	3	0.2550	5	0.1905	5	0.0160	2
2cov	*Land use / vegetation coverage*	0.8483	1	0.4250	4	0.0125	7	0.0122	3
3sei	*Seismic Intensity*	−0.1628	6	−0.0910	7	−0.2004	9	−0.0110	10
4pre	*Intense Precipitations*	0.6528	2	0.4500	2	0.3943	3	0.0238	1
5sta	*Stability after large events*	0.0247	5	0.1160	6	−0.1526	8	0.0047	5
6slo	*Slope*	0.3756	4	0.4450	3	0.3896	4	−0.00045	7
7pop	*Population*	-	-	0.6840	1	0.5348	2	0.0034	6
8roa	*Road Density*	-	-	-	-	0.6101	1	0.0052	4
9bui	*Floor Area*	-	-	−0.1280	8	0.0566	6	−0.0040	9
10gro	*Building Footprint Area*	-	-	-	-	−0.2364	10	−0.0038	8
	AUC value	**0.755**		**0.784**		**0.7928**		**0.7417**	

NOTE: Descending Order columns refer to the relative order position that explanatory factor coefficients have among their group in the model by sorting them from the highest to the lowest value.

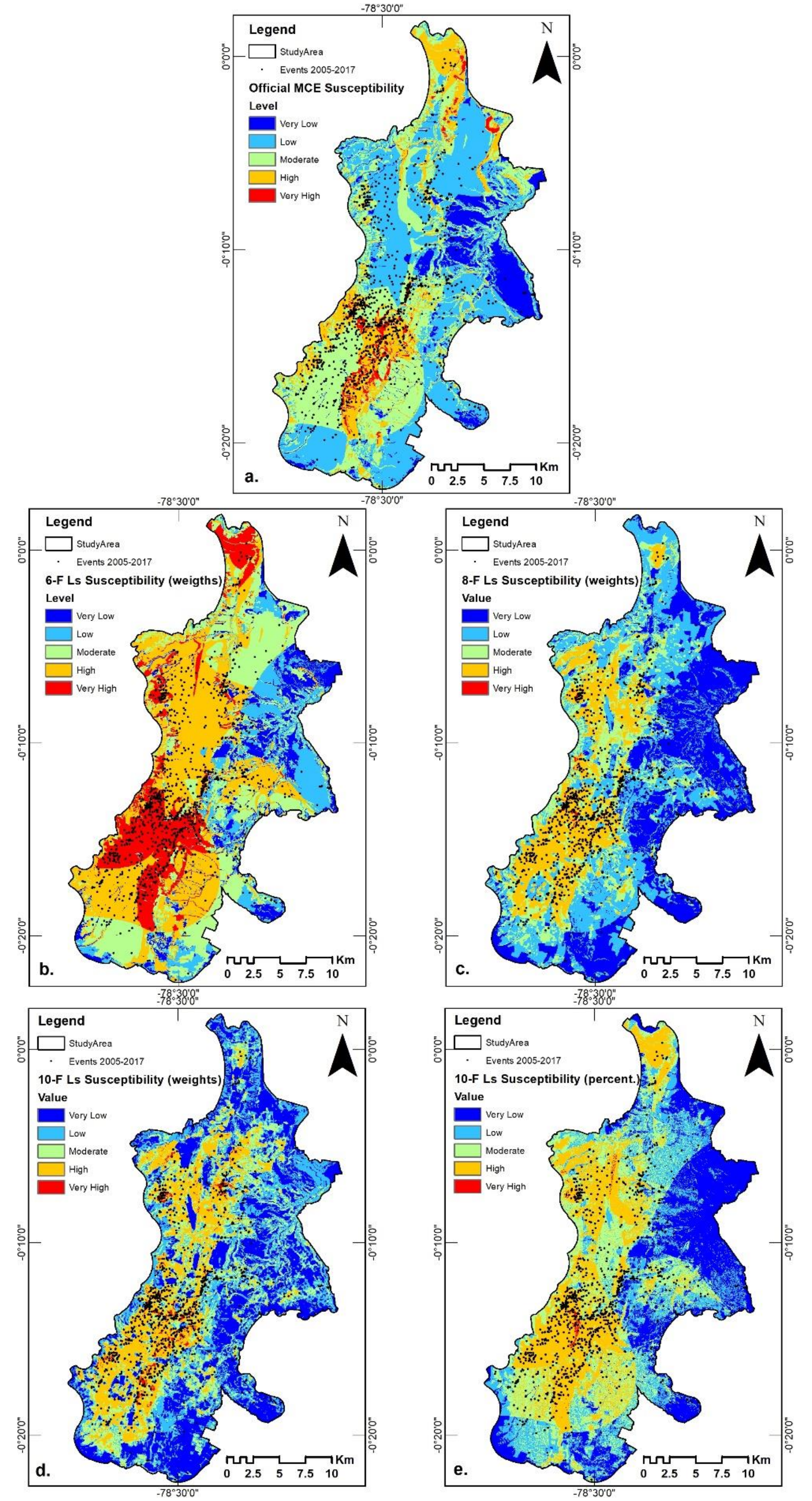


Figure 3. Landslide susceptibility maps for Quito: (a) MCE official map. From LOGIT modelling results: (b) 6-factor with weights encoding (c) 8-factor with weights encoding (d) 10-factor with weights encoding (e) 10-factor with percentile discretization. *Legend: All maps have been classified in 5 categories by geometrical interval method in GIS (ArcMap 10.3). Other features: Study Area (black line), Landslide Events during the 2005-2017 period (black dots), and five classes of susceptibility levels (blue-yellow-red color spectrum).*

4.2. Results based on resolution inputs

A previous approach in addition to testing SA based on factors, was to consider the impact of the variation in the input data cell size on the final result, using the AUC value as a metric of validation. To test this sensitivity, the minimum resolution (50 m) provided by the sources as input data was resampled to 100, 200, and 500 m. The selected standardization method for this test was the weighed-encoding approach because it performed with a higher AUC value than the percentile-discretized model (see Table 6). To illustrate changes based on resolution inputs, Table 7 shows the descriptive statistics of AUC values after simulating 100 LSM units for each of the cited cell sizes. It can be noted that the highest average AUC value corresponds to the 500 m cell case. Nonetheless, the model corresponding to this cell size is the least stable, as indicated by its standard deviation and range, which are higher than the cases of the other cell sizes. In contrast, the original 50 m cell size provided more stable behavior while maintaining a reasonable and reliable AUC value before testing other parameters for SA.

Table 7. Descriptive statistics of AUC values of LSM with LR for Quito for 100 simulations for cell sizes of 50, 100, 200, and 500 m.

Cell size (m)	**50**	**100**	**200**	**500**
Average AUC value	**0.7909**	0.7833	0.8012	**0.8277**
Maximum AUC value	0.7965	0.7892	0.8133	0.9155
Minimum AUC value	0.7740	0.7699	0.7795	0.6034
Standard deviation	**0.0052**	0.0057	0.0062	**0.0578**
Range in 100 simulations	**0.0226**	0.0194	0.0338	**0.3121**

3.3. Univariate SA results

With the 10-factor model, executed by both standardization methods (weights and percentiles), an SA based on factors was performed. For the weights-encoding method, the univariate method produced susceptibility maps whose AUC values (the metric of the sensitivity) were plotted, as seen in **Figure 4**. From this analysis, and from the range and interval set to produce the coefficient variations, there were 241 AUC values higher than the reference AUC value (0.7928) out of 2000 simulations. When observing this chart, the AUC improvement is slightly higher than the reference. From the 2000 simulations, the highest AUC value was 0.7943, which is almost 0.2% higher than the reference. The coefficients that have the stronger impacts on results, when the variations were applied, belong to the population, slope, and road density factors.

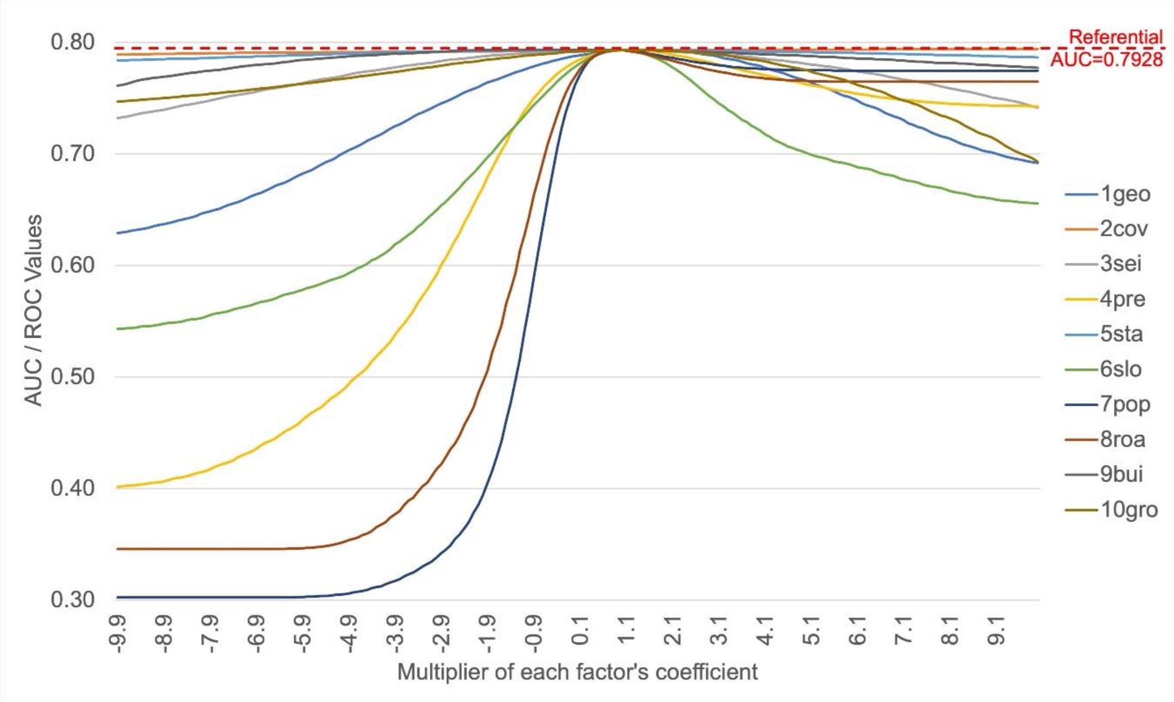
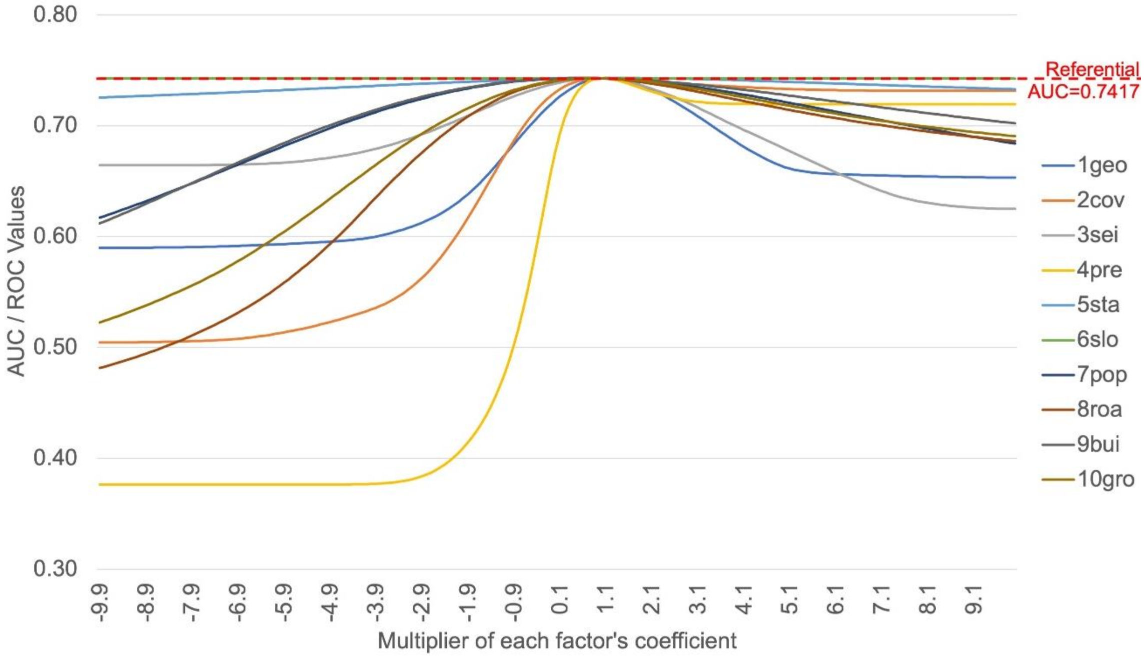


Figure 4. Univariate sensitivity analysis for LOGIT model (weights-encoding) showing AUC values as a metric of the impact of coefficient variations on the model performance. *Legend: Lines represent factors: 1geo=lithology (mid-blue), 2cov=land use/vegetation coverage (orange), 3sei=seismicity (light gray), 4pre=precipitations (yellow), 5sta=soil stability (light blue), 6slo=slope (green), 7pop=population (dark blue), 8roa=road density (brown), 9bui=floor area (dark gray) and, 10gro=building footprint area (ochre). X-axis show the Multipliers of each factor's coefficient and Y-axis show the AUC/ROC Values. The referential AUC value is shown in the dashed red line.*

The same test was applied to the percentile-discretized case. The results delivered 34 AUC values higher than the reference AUC value (0.7417) out of 2000 programmed simulations. The maximum improvement reached an AUC value of 0.7419 with a marginal improvement of 0.03%. The plotting of all AUC values derived from this univariate method variations' susceptibility maps/datasets can be seen in **Figure 5**. Precipitations, land-use cover, and road density stand out as the most sensitive factors within the defined range of variations. It must be noted that road density is among the most sensitive factors for both standardization methods, although it is not the most sensitive in either of them.



4.

5. ***Figure 5. Univariate sensitivity analysis for LOGIT model (percentile discretization) showing AUC values as a metric of the impact of coefficient variations on the model performance.*** *Legend: Lines represent factors: 1geo=lithology (mid-blue), 2cov=land use/vegetation coverage (orange), 3sei=seismicity (light gray), 4pre=precipitations (yellow), 5sta=soil stability (light blue), 6slo=slope (green), 7pop=population (dark blue), 8roa=road density (brown), 9bui=floor area (dark gray) and, 10gro=building footprint area (ochre). X-axis show the Multipliers of each factor's coefficient and Y-axis show the AUC/ROC Values. The referential AUC value is shown in the dashed red line.*

5.3. K-S test results

As an alternative means to measure sensitivity, the K-S test was applied to the same simulations undertaken for the univariate SA, including range and interval variations. Regarding the K-S test applied to the weights-encoding method, the *p*-value (at an alpha value of 0.05) showed that 13 resulting landslide susceptibility map samples, out of the 2000 simulated, were not significant. These 13 cases corresponded to variations in the coefficients of three factors: road density, intense precipitations, and population (see **Figure 6**). For the percentile-discretized case, with the same alpha value, number of simulations, and range of variation, the resulting *p*-values showed that the results of their samples were non-significant only for two maps, with both samples corresponding to the intense precipitations factor coefficient variations. These results are illustrated in **Figure 7**.

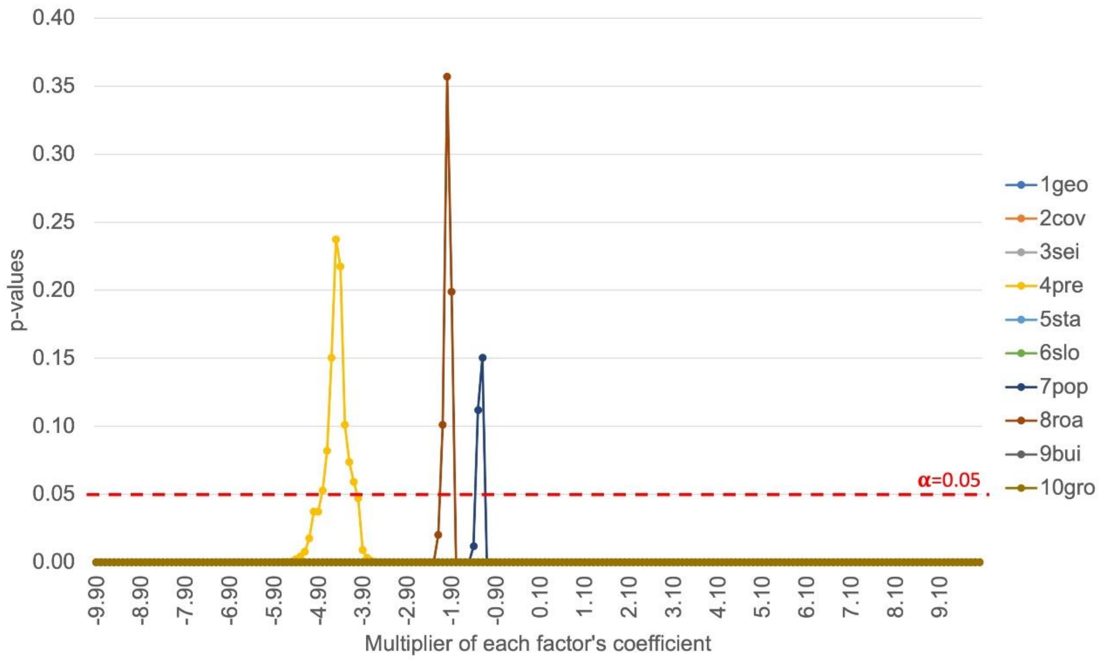


Figure 6. This shows the p-values of the K-S test (two-sample) applied to 2000 landslide susceptibility datasets resulting from the univariate susceptibility analysis of the LOGIT model by weights discretization. *Legend: The alpha value to determine the significance level was 0.05, which is marked with the dashed red line. The values (dots) above this level (precipitations, population, and road density) represent the non-significant samples from 13 simulations, whereas the remaining 1987 are considered to be at significance levels. The building footprint factor dots (ochre color) cover most of the dots representing these remaining simulations, which are very close to a p-value of zero. Lines with dots represent factors: 1geo=lithology (mid-blue), 2cov=land use/vegetation coverage (orange), 3sei=seismicity (light gray), 4pre=precipitations (yellow), 5sta=soil stability (light blue), 6slo=slope (green), 7pop=population (dark blue), 8roa=road density (brown), 9bui=floor area (dark gray) and, 10gro=building footprint area (ochre). X-axis show the Multipliers of each factor's coefficient and Y-axis show the p-values.*

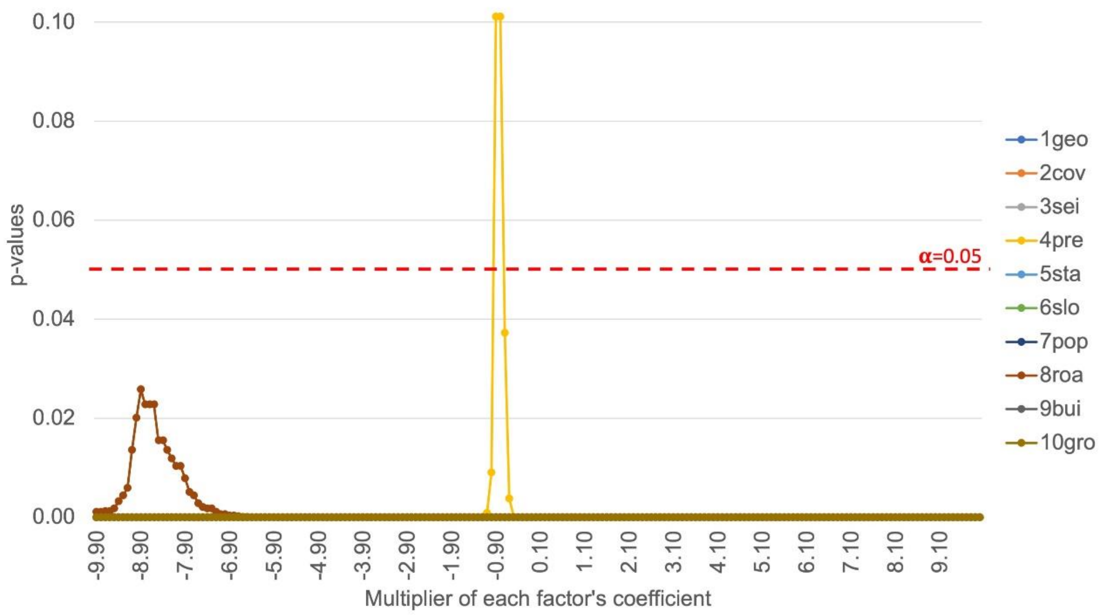


Figure 7. This shows the p-values of the K-S test (two-sample) applied to 2000 landslide susceptibility datasets resulting from the univariate susceptibility analysis of the LOGIT model by percentile discretization. *Legend: The alpha value to determine the significance level was 0.05, which is marked with the dashed red line. The values (dots) above this level (precipitations) represent the non-significant samples from two simulations, whereas the remaining 1998 are considered to be at significance levels. The building footprint factor dots (ochre color) cover most of the dots representing these remaining*

simulations, which are very close to a p-value of zero. Lines with dots represent factors: 1geo=lithology (mid-blue), 2cov=land use/vegetation coverage (orange), 3sei=seismicity (light gray), 4pre=precipitations (yellow), 5sta=soil stability (light blue), 6slo=slope (green), 7pop=population (dark blue), 8roa=road density (brown), 9bui=floor area (dark gray) and, 10gro=building footprint area (ochre). X-axis show the Multipliers of each factor's coefficient and Y-axis show the p-values.

As mentioned in the methodology, the presentation of the values and charts of the D-statistic of the K-S test, also called the KS statistic, is more insightful as a measure of sensitivity. To illustrate these values and its correspondence with the *p*-values described above, the D-critical value (Dα) was calculated, which for both discretization methods was 0.057. The 13 non-significant samples, according to their *p*-values from the weights-discretization method cases, and the two non-significant samples, according to the *p*-values from the percentile-discretization method cases, can be identified in **Figure 8**, corresponding to variations in the coefficients of the factors: road density, intense precipitations, and population. The high sensitivity of these factors is notable within the studied range of variation. The same can be appreciated in **Figure 9** for the percentile-discretized method cases. In this figure, two simulations, which are the non-significant samples, both corresponding to the intense precipitations factor coefficient variation, present D-statistic values below the critical Dα = 0.057. Similarly, in addition to this factor, road density and building footprint can also be appreciated as highly sensitive in **Figure 9**.

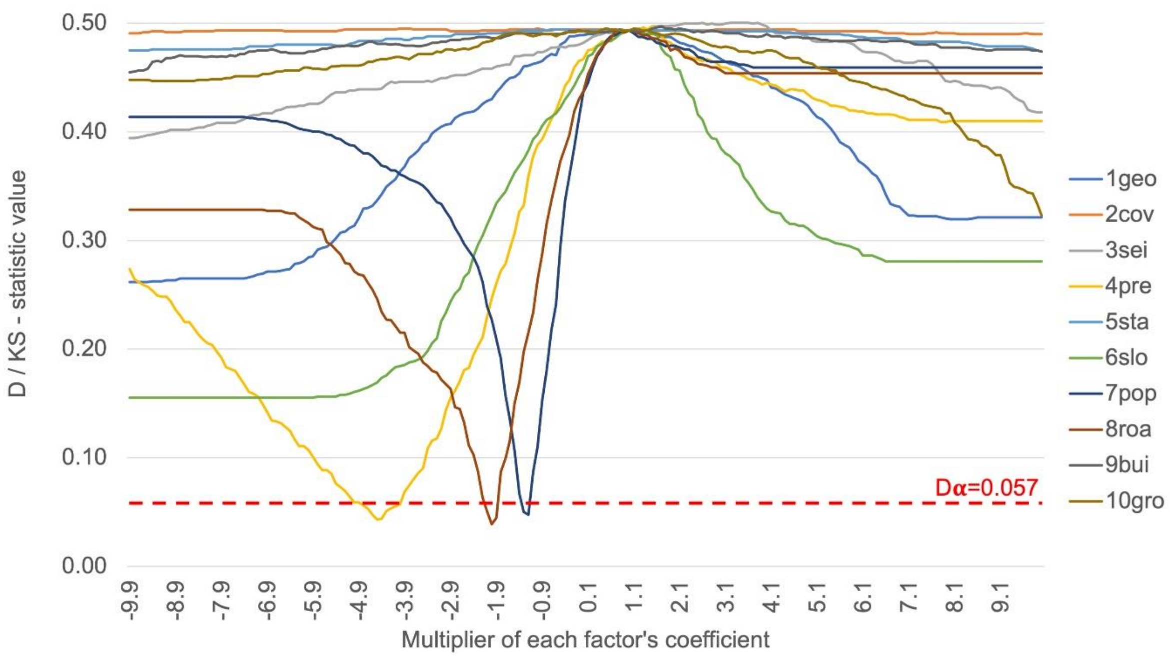


Figure 8. This shows the D-statistic values of the two-sample K-S test for univariate sensitivity analysis of 2000 simulations, by weights discretization. *Legend: The critical value (Dα), related to an alpha value of 0.05, is marked with the dashed red line. The values below this level (precipitations, population, and road density) represent the non-significant samples from 13 simulations, whereas the remaining 1987, represented above the line, are considered to be at significance levels. Lines represent factors: 1geo=lithology (mid-blue), 2cov=land use/vegetation coverage (orange), 3sei=seismicity (light gray), 4pre=precipitations (yellow), 5sta=soil stability (light blue), 6slo=slope (green), 7pop=population (dark blue), 8roa=road density (brown), 9bui=floor area (dark gray) and, 10gro=building footprint area (ochre). X-axis show the Multipliers of each factor's coefficient and Y-axis show the D/KS values.*

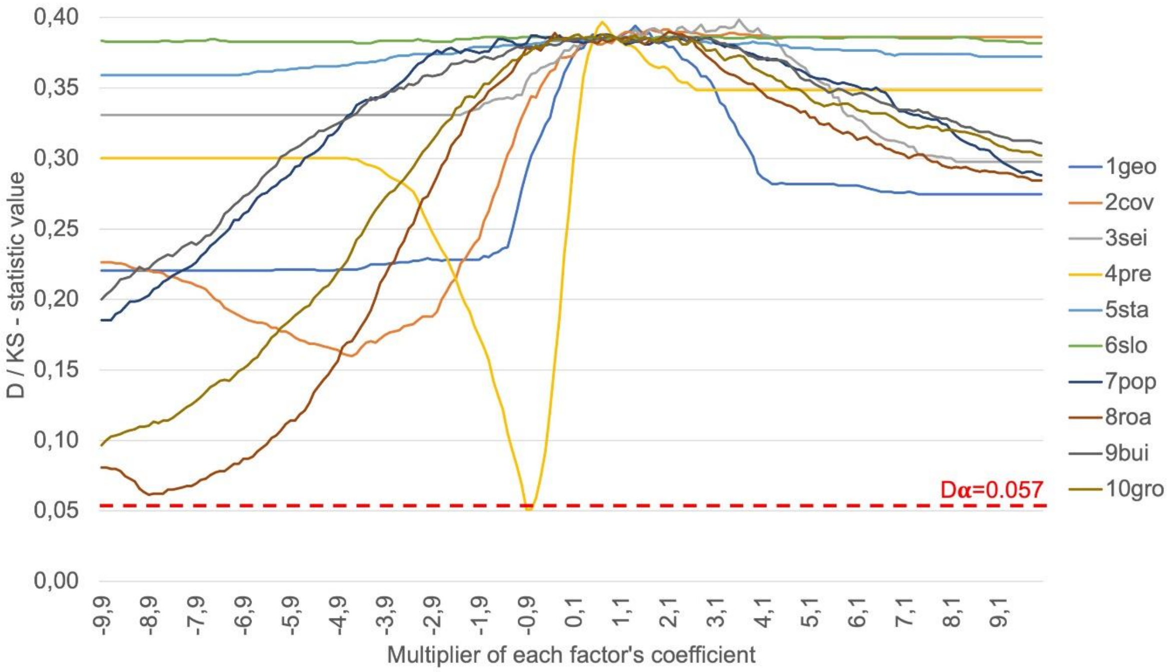


Figure 9. This shows the D-statistic values of the two-sample K-S test for univariate sensitivity analysis of 2000 simulations, by percentile discretization. *Legend: The critical value (Dα), related to an alpha value of 0.05, is marked with the dashed red line. The values below this level (precipitations) represent the non-significant samples from two simulations, whereas the remaining 1998, represented above the line, are considered to be at significance levels. Lines represent factors: 1geo=lithology (mid-blue), 2cov=land use/vegetation coverage (orange), 3sei=seismicity (light gray), 4pre=precipitations (yellow), 5sta=soil stability (light blue), 6slo=slope (green), 7pop=population (dark blue), 8roa=road density (brown), 9bui=floor area (dark gray) and, 10gro=building footprint area (ochre). X-axis show the Multipliers of each factor's coefficient and Y-axis show the D/KS values.*

5.4. Monte Carlo SA results

The second SA method was the random/stochastic method, also called Monte Carlo. The AUC value was, as previously discussed, used as a metric to test the sensitivity for both weights and percentiles standardization methods. For each, 8000 simulations were produced. A difference can be seen between the weights and the percentile standardization methods' results. The weights-encoding model generated 350 AUC values (out of the 8000) that were higher than the reference of 0.7928, with the highest AUC = 0.7970, an improvement of 0.53% compared to the reference; whereas the percentile-discretized model generated 4440 values (out of the 8000) that were higher than the referential 0.7417, with a maximum AUC = 0.7968, an improvement of 7.43% compared to the reference. For the percentile-discretization case, a set of the highest 17, out of the 4440 possible combinations to change the reference coefficients, are presented in Table 8.

Table 8. Randomly selected combination of multipliers of coefficients of the LOGIT model (percentile-discretized) to calibrate it for optimal results in defining the landslide susceptibility map.

AUC	***Slope***	***Land Use / Veg Cover***	***Seismic Intensity***	***Intense Precipitations***	***Soil Stability***	***Slope***	***Population***	***Road Density***	***Floor Area***	***Building Footprint***
0.74170**	**0.01600***	**0.01220***	**-0.01100***	**0.02380***	**0.00470***	**-0.00045***	**0.00340***	**0.00520***	**-0.00400***	**-0.00380***
0.79682	1	1	1	1.3	1	1	1	1	1	1
0.79682	1	1	1	1.3	1	1	1	1	1	1
0.79676	1	1	1	1.4	1	1	1	1	1	1
0.79676	1	1	1	1.4	1	1	1	1	1	1
0.79676	1	1	1	1.4	1	1	1	1	1	1
0.79671	1	1.5	1.4	1.3	1	1	1	1	1.9	1
0.79667	1	1	1	1.2	1	1	1	1	1	1
0.79667	1	1	1	1.2	1	1	1	1	1	1
0.79662	1	1	1	1.4	3.9	1	1	1	1	1
0.79655	1	1	1	1.5	3.1	1	1	1	0.3	1
0.79653	1	1	1	1.1	1	1	1	1	1	1
0.79653	1	1	1	1.1	1	1	1	1	1	1
0.79651	3.4	4.2	1.6	3.9	0.1	4.1	3.1	3.7	2.9	1.1
0.79651	1	1	1	1.1	1.8	1	1	1	1	1
0.79651	1	1	1	1.1	1.8	1	1	1	1	1
0.79646	0.9	1	1	1	1	1	1	1	1	1
0.79646	0.9	1	1	1	1	1	1	1	1	1

* The reference coefficients for each factor provided from the former LOGIT model application (percentile-discretization)

** The reference AUC values provided after the former LOGIT model application (percentile-discretization)

*To provide a graphic example, the two factors with the highest values of coefficients (see Table 6) were selected to compare their AUC values with combinations from the variations of both coefficients, while the other eight coefficients maintained the reference values (ceteris paribus). They are illustrated in charts to support the identification of the coordinates (combination) that performs with the highest AUC value. For the weights-encoding model these factors were population and road density, and the random outputs selected 12 combinations; however, none of these were higher than the reference (see **Figure 10**). For the percentile-discretization model, the factors were lithology and precipitations, and the random outputs presented 18 combinations. In this last case, all combinations achieved higher AUC values than the reference (0.7417), as shown in*

Figure 11.

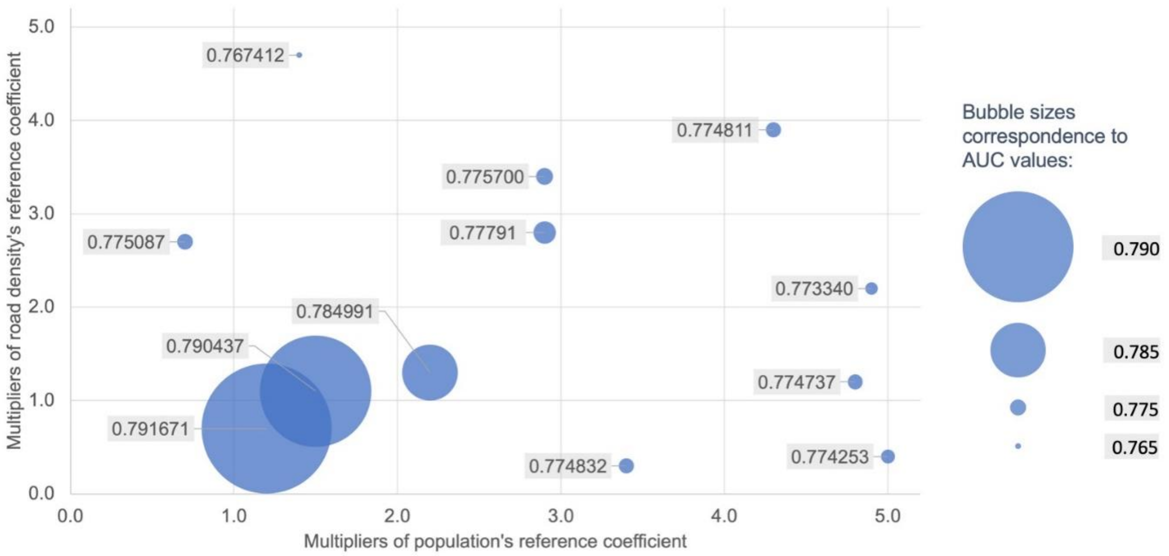


Figure 10. AUC values resulting from random combinations of population and road density coefficient variations of the reference outputs of the weights-encoding LOGIT model. *Legend: Blue bubble sizes represent the AUC values, whose area has been magnified according to the artifice (ROC*10)^200. X-axis show the multipliers of the population factor's reference coefficients and Y-axis show the multipliers of the road density factor's reference coefficients.*

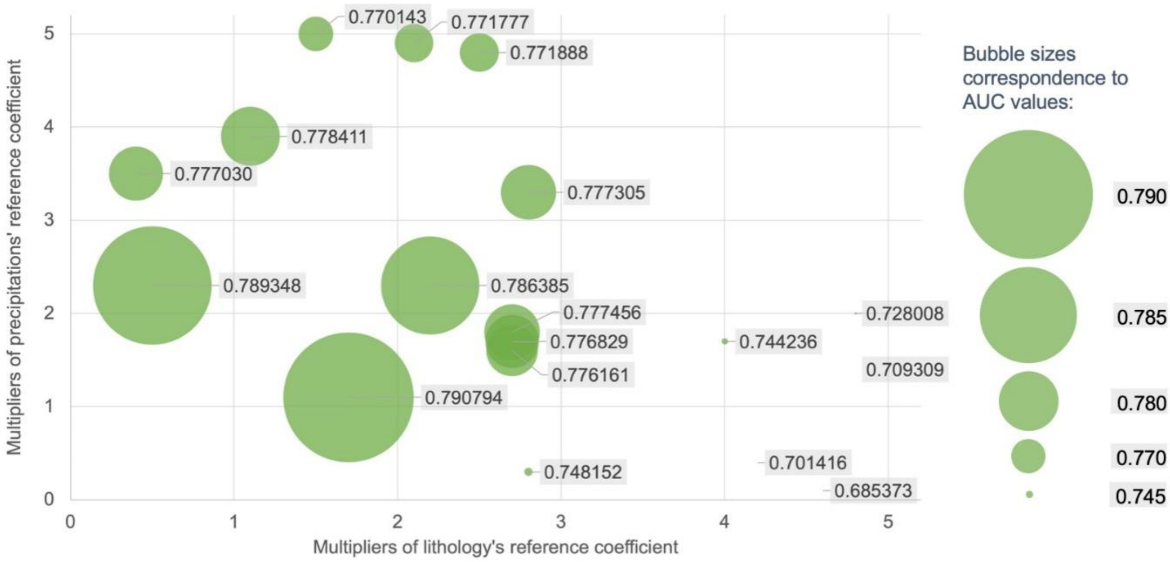


Figure 11. AUC values resulting from random combinations of lithology and precipitations coefficient variations of the reference outputs from the percentile-discretization LOGIT model. *Legend: Green bubble sizes represent the AUC values, whose area has been magnified according to the artifice (ROC*100)^100. X-axis show the multipliers of the lithology factor's reference coefficients and Y-axis show the multipliers of the precipitations factor's reference coefficients.*

6. Discussion

Prior to discussion of the particular component results and outcome of this study, consideration should be given to the broad set of possible combinations of methodologies that a global LSM process, such as that presented, may adopt. Regardless of the modelling technique, it is relevant to consider the manner in which the data is generated and preprocessed as a parameter. In this regard, Van Dessel, van Rompaey, & Szilassi (2011) stress the impact of the quality of the input datasets on the resulting coefficients of logistic regression models applied to landslide susceptibility analysis. In this regard, when revising the quality of the data corresponding to soil stability and seismic intensities, it is notable that the level of detail is poor. This might limit the accuracy and performance of the models used in this study. Currently, micro-zoning seismicity studies are being surveyed as part of a long-term project. In the future, this could help obtain precise results in LSM studies, in addition to better quality and newer data, which appears to be a promising development for Quito.

This research implemented a pseudo-quantitative method, which considered weights-encoded categorical data and discretized continuous data as inputs, previously assigned by official local experts. Data encoding is driven by needs and expertise (Saltelli et al., 2019) of local DRM professionals and scholars. In this experiment, local knowledge was useful in gathering reliable data, regardless of potential distortions in the results, which can be enhanced and complemented by the statistical analysis of the empirical data itself (Grzenda, 2020). In this regard, other studies have encoded data based on the frequency ratio of events of a specific class (Bui, Tsangaratos, Nguyen, Liem, & Trinh, 2020). For Quito, problems of unbalanced, unstructured, and underreported landslide events data are expected to be overcome in the future, thus improving weights encoding. Nevertheless, the encoded data was processed using a theorical multicriteria assessment for LSM, as undertaken in other research (Leoni et al., 2009; Lombardo & Mai, 2018), which this study tested with a LOGIT model and SA, as described above.

For the current research, complementary data preparation was performed by scaling the factors' values based on both weight and percentile methods. In this regard, there are differences that should not be overlooked when running the LOGIT. For the first method (weights), the discretization scaling to the additional four "urban" factors enhanced the ROC/AUC value progressively, from 0.7550 to 0.7840, by adding population and floor area; and then to 0.7928 by adding road density and building footprint area. Nonetheless, when discretizing all ten factors to a percentile scale, the ROC/AUC value dropped to 0.7417, which is still acceptable in terms of the classification power of the model, but is not as reliable as the higher former value. Therefore, the discretization from a 1-to-4 weights scale to percentiles of the nominal factors did not provide accuracy. This may be because, for the 1-to-4 discreet weights scale, the only possible percentile values assigned were 1, 33, 67, and 100, which affected the performance results of the model. In contrast, a broader scale, such as that of the percentiles, normally provides better parameterization possibilities and, therefore, global performance, than the cost-sensitive LOGIT model (Zhang, Ray, Priestley, & Tan, 2020).

In addition to using data provided by the municipality of Quito, this research included more factors than the official landslide susceptibility study. These factors were population, road density, floor area, and building footprint area, which we consider helped to characterize the urban category from the land use/vegetation coverage factor of the former study. This can be seen by observing the progressive change in the order of coefficients and AUC values in Table 6.

In relation to the urban variables, it can be expected that population, which appeared as an important predictor after the LOGIT application, may be related to building footprint area and floor area. Nonetheless, the latter factors were not found to be relevant predictors. This might be related to the fact that the largest floor area volumes are concentrated on the center-north of the city, an area where self-built and informal construction is low, and buildings are often medium-rise with appropriate construction techniques, soil management, and artificial drainage. A further step of this study could include factors to assess LSM at building scales, using a vulnerability and uncertainty quantification approach, and considering the heterogeneity of urban fabrics, such as those undertaken by Kaynia et al. (2008) and Du et al. (2013). This research addition may enhance the data quality in surveying building conditions and soil management, and data collection at a large urban scale, as for the case of Quito.

Another complementary remark regarding the results in terms of policy implications is that the relevance of the road density factor as a predictor does not necessarily mean that streets and roads per se increase LRisk. In fact, heterogeneity can have an important effect on the model (Wang et al., 2020). Because roads are potential boosters of urban development, it is necessary to examine in detail how vulnerability is produced at the household scale, as mentioned above.

7. Conclusions

This article presented the application of a binary logistic regression model with the objective of deriving a landslide susceptibility map for the urban area of Quito, Ecuador. A landslide events database covering the 2005–2017 period was used as the dependent binary factor. Ten explanatory factors were tested*:* lithology, land use / vegetation coverage, seismic intensities, intense precipitations, soil stability based on previous events, slope, population, road density, floor area, and building footprint area. Adding the last four factors—considered "urban"—was found to result in better performance of the model (AUC = 0.7928) compared to the model operating only with the first six factors (AUC = 0.7550).

Two data standardization methods were applied: weights encoding and percentile discretization. After operating the LOGIT model, the weighted-encoding method delivered an AUC of 0.7928, whereas the percentile-discretized model obtained 0.7417. Regarding the resulting coefficients of the explanatory factors, the weights-encoding method provided more stable values than the percentile-discretized approach, whose performance delivered greater oscillation in the coefficient values, after several simulations. The instability in the second method may be due to large differences in the percentiles' classification, particularly for the categorical data. According to the results of the weights-encoding method, which was the most stable model, the factors of road density, population, intense precipitations, and slope, in that order, improved the prediction/classification power. The percentiles-discretization method (least stable model) showed that intense precipitations, lithology, land use/vegetation coverage, and road density, in that order, provided the best prediction improvement.

Concerns about the resolution were addressed by testing changes of this parameter for cell sizes of 50 (the originally generated input), 100, 200, and 500 m. Results showed that, even when the highest values are achieved with smaller resolutions, this behavior is not stable when producing several simulations. In contrast, the 50 m cell size model remains stable, whereas its performance according to the AUC value (0.7928) demonstrates a high classification power. Hence, this resolution was chosen for the sensitivity analysis.

Univariate, Monte Carlo and K-S tests were applied to measure the sensitivity of factors (both standardization methods). AUC was used to measure the performance for the first two tests and the K-statistic for the final test. From the univariate SA results, it can be observed that the slope, road density, intense precipitations, and population factor curves resulted in the widest variations. This was the case for the weights-encoded method (see **Figure 4**). In the case of the percentile-discretized method (see **Figure 5**), the most sensitive factor curves were those of intense precipitations, land use/vegetation coverage, lithology, and road density, within the studied simulation range. Moreover, 241 out of 2000 simulations provided better calibration of the weights-encoded model, with a 0.2% improvement of the AUC value; and 34 out of 2000 simulations provided better calibration of the percentile-discretization model, but with only a marginal 0.03% improvement of the AUC value.

Regarding the Monte Carlo SA for the weights-encoding model, 350 out of 8000 simulations showed higher AUC values than the reference value, improving on it by up to 0.53%. This differs substantially from the weights-normalized model, which in the Monte Carlo application only resulted in 4440 AUC values that were higher than the reference, improving on it by up to 7.4%. This means that the calibration of the percentile-normalized model's coefficients can still be adjusted to improve predictability. Nonetheless, it does not achieve better performance than the weights-encoding model.

Finally, a two-sample K-S test was used to measure sensitivity, using the D-statistic as a metric, and using the same univariate SA simulations. In contrast to the univariate and Monte Carlo SAs, the K-S test indicated that, for the weights' method, 13 simulations in the weights-encoded cases were not significant in classification power at alpha = 0.05, with the variations of road density, precipitations, and population coefficients being sensitive beyond the D-critical value. For the percentile-discretization cases, only two simulations were not significant, with precipitations the sensitive factor, using the same alpha value.

This research aimed to contribute to the study of LSM, not only with the inclusion of Quito as an Andean city in an LRisk context, but also by observing LR modelling behavior when incorporating novel "urban" factors, which is rarely found in the LSM literature. In this regard, results highlight the importance of the street/road network and population factors on the overall classification power of the model. In contrast, the building-related factors (i.e., floor area and building footprint) do not appear to have the same influence and their inverse effect remains to be explored. Other factors, such as slope and rainfall, appeared to be relatively relevant in the LOGIT-LSM in this urban context, although the quality of the local geology, for the most part, helps to reduce the incidence of LRisk, according to local experts.

It is expected that further research for the case of Quito and similar Andean cities will benefit from the generation of better quality and more detailed official data, particularly regarding factors addressing the urban form and physical and social vulnerability. Ultimately, improved performance of LSM may support LRR policy design, considering the diverse implications that accurate delimitation of risk zones has for LRisk generation, in addition to making land suitable for LRR, the construction of safe housing, and urban development.

References

Abbaszadeh Shahri, A., Spross, J., Johansson, F., & Larsson, S. (2019). Landslide susceptibility hazard map in southwest Sweden using artificial neural network. *Catena*, *183*(June 2018), 104225.

Alcaldia del Distrito Metropolitano de Quito. Resolución Administrativa A008, Ordenanzas y Resoluciones 1–5 (2018). Quito: Alcaldia del Distrito Metropolitano de Quito. Retrieved from http://www7.quito.gob.ec/mdmq_ordenanzas/Resoluciones de Alcaldía/Año 2018/RA-2018-008-DIRECCION METROPOLITANA DE RESILIENCIA-CREACION.PDF

Asamblea Nacional del Ecuador. Ley Orgánica Reformatoria al Código Orgánico de Organización Territorial, Autonomía y Descentralización (2014). Quito: Asamblea Nacional Ecuador. Retrieved from http://www.misionpichincha.gob.ec/transparencia/organizacion-interna-base-legal/normas de regulacion/pdfs/Registro Oficial 166 Ley Reformatoria al Cootad.pdf

Bathrellos, G. D., Kalivas, D. P., & Skilodimou, H. D. (2009). GIS-based landslide susceptibility mapping models applied to natural and urban planning in Trikala, central Greece. *Estudios Geologicos*, *65*(1), 49–65.

Blanchard-Boehm, R. D. (2004). Natural Hazards in Latin America: Tectonic Forces and Storm Fury. *The Social Studies*, *95*(3), 93–105.

Bouyer, J. (2009). *Modélisation et simulation des microclimats urbains: Étude de l'impact de l'aménagement urbain sur les consommations énergétiques des bâtiments*. Université de Nantes. Retrieved from https://tel.archives-ouvertes.fr/tel-00426508

Bui, D. T., Tsangaratos, P., Nguyen, V. T., Liem, N. Van, & Trinh, P. T. (2020). Comparing the prediction performance of a Deep Learning Neural Network model with conventional machine learning models in landslide susceptibility assessment. *Catena*, *188*(July 2019), 104426.

Catani, F., Lagomarsino, D., Segoni, S., & Tofani, V. (2013a). Exploring model sensitivity issues across different scales in landslide susceptibility. *Natural Hazards and Earth System Sciences Discussions*, *1*(2), 583–623.

Catani, F., Lagomarsino, D., Segoni, S., & Tofani, V. (2013b). Landslide susceptibility estimation by random forests technique: Sensitivity and scaling issues. *Natural Hazards and Earth System Sciences*, *13*(11), 2815–2831.

Catani, Filippo, Lagomarsino, D., Segoni, S., & Tofani, V. (2013). Sensitivity analysis and scale issues in landslide susceptibility mapping, *15*, 9000.
Chang, K. T., Merghadi, A., Yunus, A. P., Pham, B. T., & Dou, J. (2019). Evaluating scale effects of topographic variables in landslide susceptibility models using GIS-based machine learning techniques. *Scientific Reports*, *9*(1), 1–21.
Comunidad Andina. Estrategia Andina para la Gestión del Riesgo de Desastres - EAGRD, Pub. L. No. 819 (2017). Perú: Comunidad Andina. Retrieved from http://www.comunidadandina.org/StaticFiles/2017522151956ESTRATEGIA ANDINA.pdf
Concejo del Distrito Metropolitano de Quito. Ordenanza Metropolitana No. 095. Régimen de Suelo de Quito (2003). Quito: Concejo del Distrito Metropolitano de Quito. Retrieved from http://www7.quito.gob.ec/mdmq_ordenanzas/Ordenanzas/ORDENANZAS AÑOS ANTERIORES/ORDM-095 - NUEVO REGIMEN DEL SUELO.pdf
Concejo del Distrito Metropolitano de Quito. Ordenanza Metropolitana No. 172. Régimen de Suelo de Quito (2011). Quito: Concejo del Distrito Metropolitano de Quito. Retrieved from http://www7.quito.gob.ec/mdmq_ordenanzas/Concejo Abierto/Ordenanzas/ORDENANZAS MUNICIPALES/MUNICIPAL (172)/MUNICIPAL_0172_517.pdf
D'Ercole, R., Hardy, S., Metzger, P., & Robert, J. (2009). Vulnerabilidades urbanas en los países andinos. Introducción general. *Bulletin de l'Institut Français d'études Andines*, *38*(3), 401–410.
Dragićević, S., Lai, T., & Balram, S. (2014). GIS-based multicriteria evaluation with multiscale analysis to characterize urban landslide susceptibility in data-scarce environments. *Habitat International*, *45*(P2), 114–125.
Du, G., Zhang, Y., Iqbal, J., Yang, Z., & Yao, X. (2017). Landslide susceptibility mapping using an integrated model of information value method and logistic regression in the Bailongjiang watershed, Gansu Province, China. *Journal of Mountain Science*, *14*(2), 249–268.
Du, J., Yin, K., Nadim, F., & Lacasse, S. (2013). Quantitative vulnerability estimation for individual landslides. In P. des Ponts (Ed.), *Proceedings of the 18th International Conference on Soil Mechanics and Geotechnical Engineering, Paris 2013* (pp. 2181–2184). Paris: Presses des Ponts. Retrieved from http://www.cfms-sols.org/sites/default/files/Actes/2181-2184.pdf
Feizizadeh, B., & Blaschke, 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.
FUNEPSA, Rivera, M., Salazar, A., Carvajal, A., Galárraga, R., Plaza, G., … Salazar, D. (2015). *Actualización de la Zonificación por Amenaza de Deslizamiento en el Distrito Metropolitano de Quito*. Quito.
GEMMA. (2007). Movimientos en masa en la región Andina: Una guía para la evaluación de amenazas. *Servicio Nacional de Geología y Minería*, (4), 432. Retrieved from http://bvpad.indeci.gob.pe/doc/pdf/esp/doc2212/doc2212.htm
Grzenda, W. (2020). The Role of Discretization of Continuous Variables in Socioeconomic Classification Models on the Example of Logistic Regression Models and Artificial Neural Networks. In K. Jajuga, J. Batóg, & M. Walesiak (Eds.), *Classification and Data Analysis. Theory and Applications* (First, pp. 35–52). Cham: Springer.
Gudiyangada Nachappa, T., Kienberger, S., Meena, S. R., Hölbling, D., & Blaschke, T. (2020). Comparison and validation of per-pixel and object-based approaches for landslide susceptibility mapping. *Geomatics, Natural Hazards and Risk*, *11*(1), 572–600.
Hemasinghe, H., Rangali, R. S. S., Deshapriya, N. L., & Samarakoon, L. (2018). Landslide susceptibility mapping using logistic regression model (a case study in Badulla District, Sri Lanka). *Procedia Engineering*, *212*, 1046–1053.
Highland, L. M., & Bobrowsky, P. (2008). *The Landslide Handbook — A Guide to Understanding Landslides*. (M. Kidd, Ed.). Reston: U.S. Geological Survey. Retrieved from https://www.researchgate.net/publication/276206970_The_Landslide_Handbook_-_A_Guide_to_Understanding_Landslides
INEC. (2016). Proyeccion Cantonal Total 2010-2020. Quito: INEC. Retrieved from http://www.inec.gob.ec/estadisticas/index.php?option=com_content&view=article&id=329&Itemid=328&lang=es
Inostroza, L. (2017). Informal urban development in Latin American urban peripheries. Spatial assessment in Bogota, Lima and Santiago de Chile. *Landscape and Urban Planning*, *165*(1), 267–279.
Kaynia, A. M., Papathoma-Köhle, M., Neuhäuser, B., Ratzinger, K., Wenzel, H., & Medina-Cetina, Z. (2008). Probabilistic assessment of vulnerability to landslide: Application to the village of Lichtenstein, Baden-Württemberg, Germany. *Engineering Geology*, *101*(1–2), 33–48.
Kirschbaum, D., & Stanley, T. (2018). Satellite-Based Assessment of Rainfall-Triggered Landslide Hazard for

Situational Awareness. *Earth's Future*, 1–19.
Klimeš, J., & Rios Escobar, V. (2010). A landslide susceptibility assessment in urban areas based on existing data: An example from the Iguaná Valley, Medellín City, Colombia. *Natural Hazards and Earth System Science*, *10*(10), 2067–2079.
Lara, M., Sepúlveda, S. A., Celis, C., Rebolledo, S., & Ceballos, P. (2018). Landslide susceptibility maps of Santiago city, Andean foothills, Chile. *Andean Geology*, *45*(3), 433–442.
Lee, S., Baek, W. K., Jung, H. S., & Lee, S. (2020). Susceptibility mapping on urban landslides using deep learning approaches in mt. Umyeon. *Applied Sciences (Switzerland)*, *10*(22), 1–18.
Leoni, G., Barchiesi, F., Catallo, F., Dramis, F., Fubelli, G., Lucifora, S., … Puglisi, C. (2009). GIS methodology to assess landslide susceptibility: Application to a river catchment of central Italy. *Journal of Maps*, *5*(1), 87–93.
Lombardo, L., & Mai, P. M. (2018). Presenting logistic regression-based landslide susceptibility results. *Engineering Geology*, *244*(January), 14–24.
MDMQ. (2015). *Diagnóstico Eje Territorial - Plan de Desarrollo y Ordenamiento Territorial*. Quito. Retrieved from http://www7.quito.gob.ec/mdmq_ordenanzas/Sesiones del Concejo/2015/Sesion Extraordinaria 2015-02-13/PMDOT 2015-2025/Volumen I/6. Diagnostico Territorial.pdf
MDMQ. (2017). Estrategia de Resiliencia Distrito Metropolitano de Quito. Quito: Imprenta Municipal de Quito. Retrieved from http://www.pichincha.gob.ec/pichincha/cantones/item/23-distrito-metropolitano-de-quito.html
Meena, S. R., Ghorbanzadeh, O., & Blaschke, T. (2019). A comparative study of statistics-based landslide susceptibility models: A case study of the region affected by the Gorkha earthquake in Nepal. *ISPRS International Journal of Geo-Information*, *8*(2).
Melchiorre, C., Castellanos Abella, E. A., van Westen, C. J., & Matteucci, M. (2011). Evaluation of prediction capability, robustness, and sensitivity in non-linear landslide susceptibility models, Guantánamo, Cuba. *Computers and Geosciences*, *37*(4), 410–425.
Norman, G. (2010). Likert scales, levels of measurement and the "laws" of statistics. *Advances in Health Sciences Education*, *15*(5), 625–632.
Orán Cáceres, P., Gómez Delgado, M., & Bosque Sendra, J. (2010). Una propuesta complementaria de análisis de sensibilidad de un modelo basado en técnicas sig y evaluación multicriterio. *Publicaciones de La Universidad de Sevilla*, 971–987.
Pascale, S., Parisi, S., Mancini, A., Schiattarella, M., Conforti, M., Sole, A., … Sdao, F. (2013). Landslide Susceptibility Mapping Using Artificial Neural Network in the Urban Area of Senise and San Costantino Albanese (Basilicata, Southern Italy). In *International Conference on Computational Science and Its Applications* (pp. 473–488).
Pasta, D. J. (2009). Learning When to Be Discrete : Continuous vs. Categorical Predictors. *SAS Global Forum 2009, Statistics and Data Analysis*, 1–10.
Pawluszek, K., Borkowski, A., & Tarolli, P. (2018). Sensitivity analysis of automatic landslide mapping: numerical experiments towards the best solution. *Landslides*, *15*(9), 1851–1865.
Poelmans, L., & Van Rompaey, A. (2010). Complexity and performance of urban expansion models. *Computers, Environment and Urban Systems*, *34*(1), 17–27.
Psomiadis, E., Papazachariou, A., Soulis, K. X., Alexiou, D. S., & Charalampopoulos, I. (2020). Landslide mapping and susceptibility assessment using geospatial analysis and earth observation data. *Land*, *9*(5).
Puente-Sotomayor, F., Egas, A., & Teller, J. (2021). Land policies for landslide risk reduction in Andean cities. *Habitat International*, *107*(December 2019), 102298.
Puente-Sotomayor, F., Villamarin, P., & Cevallos, A. (2018). Riesgos de deslizamiento en Quito, ¿funciona la política? In Asociación Geográfica del Ecuador (Ed.), *Territorios en transición: Transformaciones de la Geografía del Ecuador en el siglo XXI - Memorias del 1er Congreso Nacional de Geografía del Ecuador* (pp. 1–15). Quito: Asociación Geográfica del Ecuador. Retrieved from https://congresonacionalagec.files.wordpress.com/2018/02/memoria_cng_2018.pdf
Ramos-Bernal, R. N., Vázquez-Jiménez, R., Sánchez Tizapa, S., & Arroyo Matus, R. (2019). Characterization of Susceptible Landslide Zones by an Accumulated Index. In R. Ram & M. Lazzari (Eds.), *Landslides - Investigation and Monitoring* (First, pp. 1–26). Intech Open. Retrieved from https://www.intechopen.com/books/advanced-biometric-technologies/liveness-detection-in-biometrics
Rebotier, J. (2016). *El Riesgo y su Gestión en Ecuador - Una Mirada de Geografía Social y Política*. (C. de P. P. U. C. del Ecuador, Ed.). Quito: Centro de Publicaciones Pontificia Universidad Católica del Ecuador. Retrieved from https://biblio.flacsoandes.edu.ec/shared/biblio_view.php?bibid=143165&tab=opac
Reichenbach, P., Rossi, M., Malamud, B. D., Mihir, M., & Guzzetti, F. (2018). A review of statistically-based landslide susceptibility models. *Earth-Science Reviews*, *180*(March), 60–91.
Ronchetti, F., Alessandro, C., Kollarits, S., Leber, D., Papez, J., Plunger, K., … Stefani, M. (2013). Improve

Information Provision for Disaster Management: MONITOR II, EU Project. In C. Margottini, P. Canuti, & K. Sassa (Eds.), *Landslide Science and Practice. Volume 7: Social and Economic Impact and Policies* (First, pp. 47–54). Rome: Springer.

Saltelli, A., Aleksankina, K., Becker, W., Fennell, P., Ferretti, F., Holst, N., … Wu, Q. (2019). Why so many published sensitivity analyses are false: A systematic review of sensitivity analysis practices. *Environmental Modelling and Software*, *114*(March 2018), 29–39.

Sepúlveda, S. A., & Petley, D. N. (2015). Regional trends and controlling factors of fatal landslides in Latin America and the Caribbean. *Natural Hazards and Earth System Sciences*, *15*(8), 1821–1833.

Shahabi, H., & Hashim, M. (2015). Landslide susceptibility mapping using GIS-based statistical models and Remote sensing data in tropical environment. *Scientific Reports*, *5*, 1–15.

Sîrbu, F., Drăguț, L., Oguchi, T., Hayakawa, Y., & Micu, M. (2019). Scaling land-surface variables for landslide detection. *Progress in Earth and Planetary Science*, *6*(1), 1–13.

UNISDR. (2018). *UNISDR Annual Report 2017*. Geneva. Retrieved from https://www.unisdr.org/files/58158_unisdr2017annualreport.pdf

USGS. (2004). Landslide Types and Processes. *Highway Research Board Special Report*. U.S. Department of the Interior, U.S. Geological Survey. Retrieved from https://pubs.usgs.gov/fs/2004/3072/pdf/fs2004-3072.pdf

van Dessel, W., van Rompaeya, A., & Szilassi, P. (2011). Sensitivity analysis of logistic regression parameterization for land use and land cover probability estimation. *International Journal of Geographical Information Science*, *25*(3), 489–508.

van Lindert, P. (2016). Rethinking urban development in Latin America: A review of changing paradigms and policies. *Habitat International*, *54*, 253–264.

Wang, Y., Feng, L., Li, S., Ren, F., & Du, Q. (2020). A hybrid model considering spatial heterogeneity for landslide susceptibility mapping in Zhejiang Province, China. *Catena*, *188*(July 2019), 104425.

Williams, R. (2019). *Ordinal Independent Variables*. Notre Dame. Retrieved from https://www3.nd.edu/~rwilliam/xsoc73994/OrdinalIndependent.pdf

Wubalem, A. (2020). Modeling of Landslide susceptibility in a part of Abay Basin, northwestern Ethiopia. *Open Geosciences*, *12*(1), 1440–1467.

Zhang, L., Ray, H., Priestley, J., & Tan, S. (2020). A descriptive study of variable discretization and cost-sensitive logistic regression on imbalanced credit data. *Journal of Applied Statistics*, *47*(3), 568–581.

ANNEX: Abbreviations

AHP	Analytical Hierarchical Process
ANN	Artificial Neural Networks
AUC	Area Under the Curve (ROC value)
CF	Certainty Factor
COE-M	Metropolitan Emergency Operations Committee of Quito
DL	Deep Learning
DMQ	Metropolitan District of Quito
DRM	Disaster Risk Management
DRR	Disaster Risk Reduction
DTM	Digital Terrain Model
EMS	European Macroseismic Scale
EPMMOP	Empresa Pública Metropolitana de Movilidad y Obras Públicas (Metropolitan Public Enterprise of Mobility and Public Works)
FFNN	Feed-Forward Neural Network
FR	Data-Driven Frequency Ratio
GIS	Geographic Information Systems
INEC	Instituto Nacional de Estadísticas y Censos de Ecuador (National Institute of Statistics and Census of Ecuador)
IV	Information Value
JT	Jackknife Test
K-S	Kolmogorov-Smirnov test
LCVs	Landslide Conditioning Factors
LOGIT	Binary Logistic Regression Model
LR	Logistic Regression
LRisk	Landslide Risk
LRR	Landslide Risk Reduction
LS	Landslide Susceptibility

LSI	Landslide Susceptibility Index
LSM	Landslide Susceptibility Mapping or Landslide Susceptibility Map
LUP	Land-Use Planning
MCE	Multi-Criteria Evaluation
MDMQ	Municipio del Distrito Metropolitano de Quito (Government of the Metropolitan District of Quito)
ML	Maximum Likelihood
MuF	Multiple Factor Model
MURs	Mapping Unit Resolutions
NDVI	Normalized Difference Vegetation Index
OAT	One-At-a-Time
OWA	Ordered Weighted Average
PBA	Pixel Based Approaches
PIX	Potential Erosion Index
PRE	Potential Rain Effect
RF	Random Forest
ROC	Receiving Operator Characteristic
SA	Sensitivity Analysis
SGP	Secretaría General de Planificación (General Planning Secretariat of the MDMQ)
SI	Susceptibility Index
SPI	Stream Power Index
SSG	Secretaría de Seguridad y Gobernabilidad (Security and Governability Secretariat of the MDMQ)
STHV	Secretaría de Territorio, Hábitat y Vivienda (Territory, Habitat and Housing Secretariat of the MDMQ)
SVM	Support Vector Machine
TPI	Topographic Position Index
TRI	Topographic Roughness Index
TWI	Topographic Wetness Index
WeF	Weight Factor Model
WLC	Weighted Linear Combination
ZR	Landslide Risk Zone or “Zona de Riesgo” (a LUP category in Quito)