Landslide Susceptibility Mapping for Quito with Logistic Regression and Sensitivity Analysis

Fernando Xavier Puente-Sotomayor (fpuente@uce.edu.ec)
Universidad Central del Ecuador https://orcid.org/0000-0002-3430-7831

Ahmed Mustafa
New School

Jacques Teller
Universite de Liege

Research

Keywords: Landslide Susceptibility, Quito, LOGIT, Sensitivity Analysis, Andean Cities, Landslide Risk Reduction

Posted Date: August 27th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-60877/v1

Version of Record: A version of this preprint was published at Geoenvironmental Disasters on August 20th, 2021. See the published version at https://doi.org/10.1186/s40677-021-00184-0.

1. Introduction

Risks related to landslides are one of the main concerns for development in the Andean region. This is due to a combination of physical and social factors. The geodynamic of the Andean region is rather prone to landslides. This condition is aggravated by climate change. Further on urban areas have been growing very rapidly over the last decades. They now gather 70% of the population and the share of urban population keeps growing rapidly. Unplanned urbanization is developing without any consideration for landslide risks and public bodies have limited capacities in the management of urban development (Comunidad Andina, 2017; D'Ercole et al., 2009; UNISDR, 2018). Evidence about landslide-prone conditions in the region has been presented by Kirschbaum & Stanley (2018) and Sepúlveda & Petley (2015). These studies illustrate the concentration of landslide-susceptible area in the irregular orography held of Colombia, Ecuador and Peru, with hundreds of fatalities compared with lower figures of neighboring countries, except for Brazil. Accordingly, disaster risk management should be better integrated with land-use planning for an appropriate diagnostics and effective prevention of risks related to landslides.

Quito, the capital of Ecuador, is the largest metropolitan district of the country with 2'781.641 inhabitants in 2020 (INEC, 2016). This jurisdiction covers 4.235,2 km^2, from which 10% is urban area in which 81% of the 353.595 housing units of the metropolitan district are settled (MDMQ, 2015). As part of the Andean mountain cities, Quito has suffered from multiple natural threats, including landslides, volcano eruptions, floods and earthquakes. Exposure to risks was further exacerbated in the city, given the fast population growth and an uncontrolled urbanization process. Accordingly, Quito started collecting geodata about landslide disaster events systematically during the last two decades. This strengthened its management capacities and its approach towards preventive policies and actions (Rebotier, 2016), besides preparedness and response. Most recently, resiliency has been adopted as an urban policy up to the point to be institutionalized with the creation of a Resiliency Department and the production of the City's Resiliency Strategy as an action (Alcaldia del Distrito Metropolitano de Quito, 2018; MDMQ, 2017)

In relation to the state of the situation of landslide risk reduction policies in Quito, there is approximately one decade of history of landslide-related land-use zoning as part of the local plan. The first studies and policy were started in 2011 (Puente-Sotomayor et al., 2018). Before that building regulations had already included basic risk prevention measures, such as setbacks from ravines,

slope borders and rivers (Concejo del Distrito Metropolitano de Quito, 2003). In punctual cases like lahar slides prone areas, a transfer of responsibility from government to owners had been applied. In practice, this meant that owners who decided to build on flows-prone areas had to submit their project with a notarized responsibility of this condition before the city approval (Concejo del Distrito Metropolitano de Quito, 2011). Since 2013 this is no longer allowed according to national laws, because the responsibility of the generated risk relies on any official that approves the project (Asamblea Nacional del Ecuador, 2014). During the last decade the preventive/reductive approach was materialized by establishing a new category for landslide risk-zone (ZR) in the land use plan. Construction were strictly forbidden in ZR areas. This zoning policy combined intuitively and imprecisely slopes (1:5000 scale), soil stability (1:25000 scale) and field inspection as the only inputs in 2011. The application of this regulation triggered around 40 complaints a year from affected owners, who were affected socially (housing rights) and economically due to previous expectations and investments. In 2013 a reform to this ordinance relaxed it by giving back the right to build on ZRs as long as geotechnical studies submitted and approved by the municipal officials justified the project. This raised problems of technical capacity of users and officials, the objective definition of 'mitigation', and accessibility to technology for low socio-economic strata. A new reform in 2015 cancelled the ZR land-use category and risk areas were converted in an overlay map, which in practice did not change the policy. (Puente-Sotomayor et al., 2018). By that time, the first landslide susceptibility studies were initiated. The outputs of these studies were expected to improve the ZR policy. They were used as input data for this research

A comparison between the ZR layer, which has not changed its limits up to now, the existing landslide susceptibility study (FUNEPSA et al., 2015); and a landslide events database (from 2005-2017, seen in the study area in Figure 1) provided by the Metropolitan Emergency Operations Committee of Quito (COE-M) reveals inconsistencies between the policy, studies and facts. Only 8% of recorded landslide events are contained in ZR polygons, 25% of ZR do not match with high susceptibility areas and 81% of high susceptibility areas are not covered by the ZR polygons. This means that more areas should be considered as landslide prone while, in less proportion, some area does not need a protection policy (Puente-Sotomayor et al., 2018).

The objective of the present study is to go beyond these previous works by proposing a landslide susceptibility map based on logistic regression combined with a sensitivity analysis (SA) in

order to calibrate variables' coefficients used in the model. Developing such an evidence-based landslide susceptibility map considering sensitivity factors to appears as an indispensable input for developing an urban policy backed by a socio-political consensus (Orán Cáceres et al., 2010).

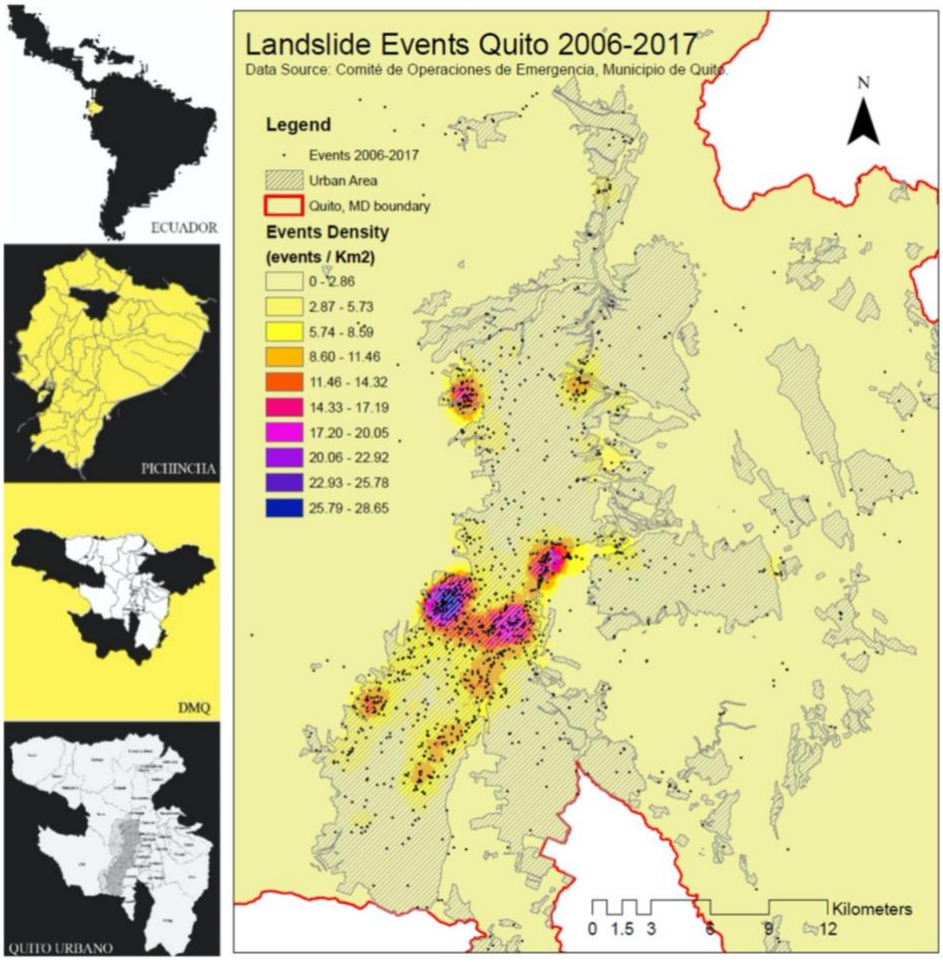

Figure 1. Study Area in the Metropolitan District of Quito, with 2005-2017 landslide events and heatmap.

2. Conceptual Framework

A landslide is here defined as: *"the downslope movement of soil, rock, and organic materials under the effects of gravity"* (Highland & Bobrowsky, 2008, p. 4). Its types vary from slides, falls, topples, flows, lateral spreads and combined, whose causes could be geological, morphological or anthropic and can be triggered by water, seismic and volcanic activities (GEMMA, 2007; USGS, 2004). Landslide disaster risk is the result of a combination of natural hazards (weak soil, intense precipitations and earthquakes), vulnerability (soil cuts and fills, or structural weakness) and exposure (construction in risk prone areas) (Puente-Sotomayor and Teller, 2019). Such an understanding of risk is directly related to landslide susceptibility that, beyond the understanding of disaster risk as a social product, pursues to identify the interaction between natural and built components susceptible to

prompt landslides. 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. They also conclude from a global review on landslide susceptibility modelling that usual determinant factors are slope, geology and aspect, being the two first the ones with higher in prediction power. They also state that results may vary according to methodologies, model validation, landslide types, triggering factors and researchers' background. Other studies include precipitations, population density and land-use as significant variables (Hemasinghe et al., 2018; Sepúlveda and Petley, 2015).

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) statistically based modelling methods. The present work combines the heuristic approach adopted by the municipality as preliminary input. It submits it together with other variables to a statistically based model. Although normalization 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).

Once the data is processed and set available for modelling, one common, though simple theoretical used model is the Multi-Criteria Evaluation (MCE) that can be combined with sensitivity analysis, as in Feizizadeh & Blaschke (2014) and Orán Cáceres et al. (2010). A complementary approach to MCE is the binary logistic regression (Logit), which is amongst the most used statistical methods for landslide susceptibility mapping (Reichenbach et al., 2018). This type of model helps go beyond weighted models, which do not help assess the probability of landslide occurrences (Lombardo and Mai, 2018).

A further step in evaluating LOGIT models is to apply a Sensitivity Analysis (SA) (Reichenbach et al., 2018). The objective of sensitivity analysis is to help adjust the calibration of the studied parameters involved in the LOGIT model to improve its predicting/classification power. Amongst different methodologies, two of them, the simple, univariate, or "one-by-one" method and the

stochastic/random-selection method, also called “Monte Carlo”, whose applications vary according to the needs of the requiring field of practice (Bouyer, 2009) will be explained later in the methodology section and their results presented in the corresponding section.

3. Data Compilation and Methods

The research builds upon data collected during a landslide risk analysis made for the municipality in 2014 and 2015. This delivered a weighted multi-criteria theoretical model with six variables surveyed and processed. These variables are *slope, intense precipitations, soil stability after former large landslides, lithology, land use / vegetation coverage* and *seismic intensity*. Each variable had partial weights 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) by using the map algebra GIS tool to sum up the partial weights as shown in Figure 2 (FUNEPSA et al., 2015).

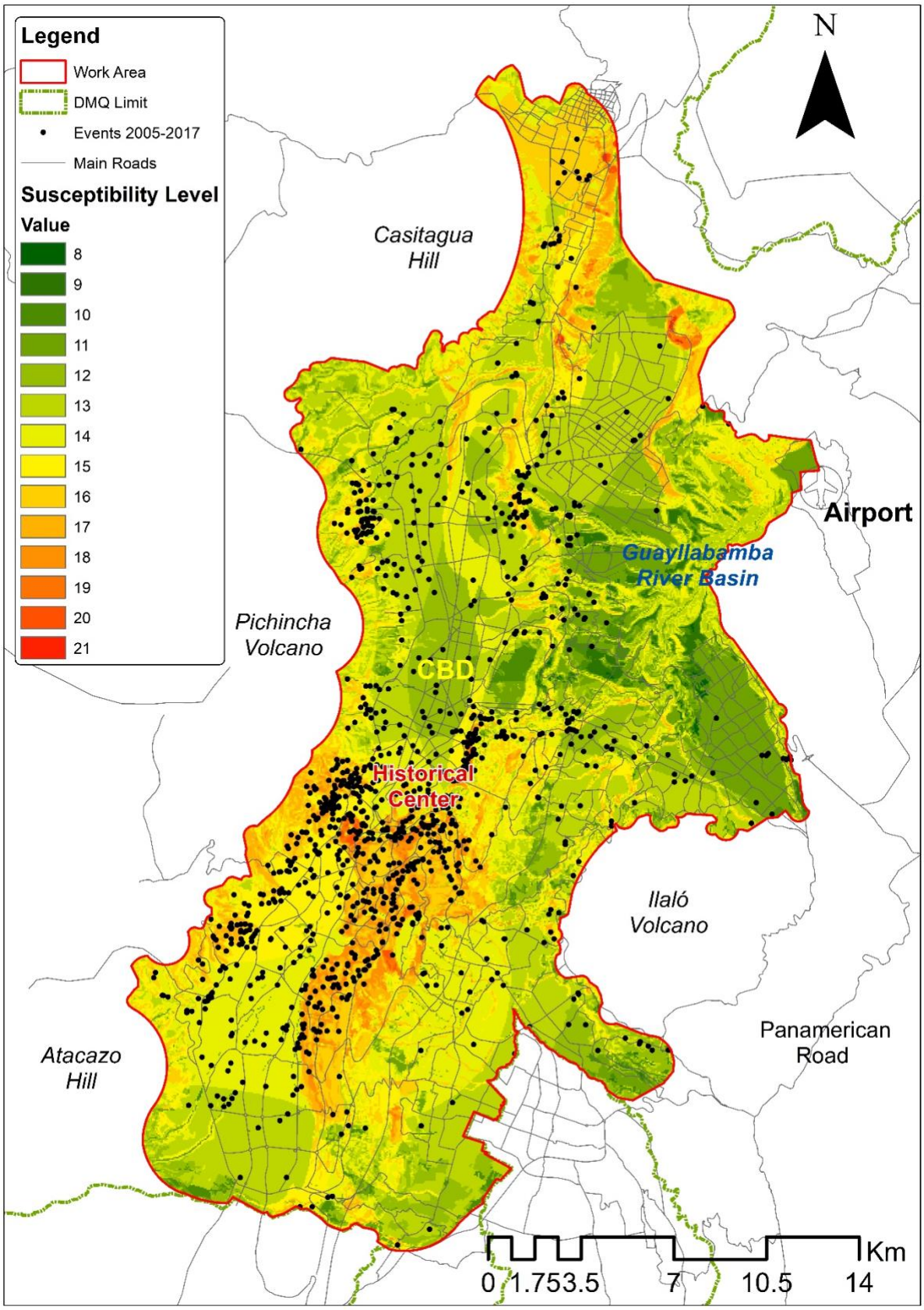

Figure 2. Landslide Susceptibility Map for Quito including 2005.2017 events sites. Data Source: Quito Municipality

Our study proposed to develop a binary logistic regression model on the basis of the variables identified by these experts. A dataset of landslide events that occurred from 2005 to 2017 was therefore collected from the COE-M (Metropolitan Emergency Operations Committee) of Quito. This database includes some 1400 events, including rotational and translational landslides, flows and topples, all considered generically as landslides (USGS, 2004). Four additional variables were included in two steps to test the model. First, *population*, provided by the INEC, and *floor area,* then *road density,* as well as *building footprint area* were included in the model. All four variables were processed and adapted for this research work. Details of all the ten variables included in this study are registered in Table 1.

Table 1. Input Data to Perform Susceptibility Mapping in Quito. Characteristics.

Content	Disaggregation	Specifications, type	Source	Year
Landslide Events	Point	Binary	COE-M Quito	2005-2017
Slope	50 m	Pre-normalized from continuous to weights (weighted classes)	FUNEPSA et al., 2015	2015
***Intense Precipitations* (in 24 hours)**	50 m	Pre-normalized from categorical to weights	FUNEPSA et al., 2015	2015
***Soil Stability* (from former landslides)**	50 m	Pre-normalized from categorical to weights	FUNEPSA et al., 2015	2015
Lithology	50 m	Pre-normalized from categorical to weights	FUNEPSA et al., 2015	2015
Land Use / Vegetation Cover	50 m	Pre-normalized from categorical to weights	FUNEPSA et al., 2015	2015
Seismicity	50 m	Pre-normalized from categorical to weights	FUNEPSA et al., 2015	2015
Population	Block Scale	Continuous, normalized by natural breaks to weights	INEC	2010
Road Density	Street segment	Continuous, normalized by natural breaks to weights	STHV – MDMQ	2016
Floor Area	Building Scale	Continuous, normalized by natural breaks to weights	STHV – MDMQ	2017
Building Footprint Area	Building Scale	Continuous, normalized by natural breaks to weights	STHV – MDMQ	2017

The dependent variable, landslide events (binary), and the ten independent, explanatory variables were processed in raster files, with a pixel size of 50 x 50 m2.

The binary layer of landslide events registers one of two categories for each area unit or pixel: true (one), when a landslide or more occurred in it; or, false (zero), when no landslides occurred in it. For the first six variables of the municipality study (FUNEPSA et al., 2015), the conversion from either classified continuous data or categorical data to weights was set in a scale from 1 to 4. The additional

four variables, all continuous data, were classified by using natural breaks in four classes, to correspond to the 1-to-4 scale. The details of this conversion are shown in Table 2. The complementary data (additional four variables), all in vector format were converted into raster format to fit with the rest of the dataset (initial six variables).

Table 2. Conversion/correspondences between variables classes or categories and weights (weights normalization) of Input Variables for Susceptibility Modeling in Quito. Source: MDMQ, INEC, adapted by authors.

Weights (Partial Susceptibilities)	**1**	**2**	**3**	**4**
***Slope* (degrees):**	• 0°-10°	• 10.1°-25°	• 25.1°-35° • > 50°	• 35.1°-50°
***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
***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	• Airport • Short-cycle crops • Cropped grass • Natural grass	• Quarries • Buildings • Eroded soils

	Saxicola Vegetation			
Soil Stability **(categories)**	• Stabilized	• Latent	• Reactivated • Colluvial	• Active
Intense Precipitations **in 24 hours (a. Máx in mm / n>10 Return time: 100 yrs, b. Average rain in mm) / n<10 yrs)**	• <75.4	• 76.4-91.88	• 91.88-107.34	• 107.34-122.79
Seismic Intensity **(European Macroseismic Scale, ordinal)**	Not applicable	• EMS VII	• EMS VIII	Not applicable
Population **(Inhabitants)**	• 0-1.81	• 1.82-5.56	• 5.57-12.47	• 12.48-31.86
Road Density **(m/Ha)**	• 0-7.50	• 7.51-18.16	• 18.17-31.99	• 32-100.70
Floor Area **(m2)**	• 0-4180.33	• 4180.34-35950.76	• 35950.77-104508.01	• 104508.02-213196.34
Building Footprint Area **(m2)**	• 0-3344.26	• 3344.27-34278.63	• 34278.64-104508.01	• 104508.02-213196.34

A second normalization for the ten explanatory variables was tested and applied. It is based on a percentile normalization. The aim of applying percentile normalization was, first, to have a finer value than the scale from 1 to 4 and second, to correct an error existing in the provided data that was produced by a small portion of outliers that distorted the real range of the dataset. That condition particularly occurred with the data of the *floor area* and *building footprint area*. Table 3 shows how the ten variables were normalized through percentiles. The categorical data weights were assigned a specific percentile, which was each end or limit resulting from the division of the 1-to-100 scale in three equal segments (assuming weights as equal units in a discrete scale). For the remaining continuous data variables, the new values were simply the corresponding percentile.

Table 3. Conversion table of categorical data from weights to percentile normalization, and for continuous data to percentiles.

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

Two sets of units/pixels with an equal number of items were then selected. The first one with pixels that did not register the occurrence of landslides and the second one with pixels that did register one or more events. A generalized linear model regression was then applied to obtain values corresponding to the intercept of the function and the coefficients for all ten variables. With these values the logistic regression (Equation 1) was applied in order to obtain the landslide susceptibility values for all the elements / map units 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 maps. This was done, firstly, for the six variables from the study provided by the municipality; secondly, with the addition of *population* and *floor area* as new variables; and lastly, with the addition of *road density* and *building footprint area*.

The coefficients of the different Logit models were validated with the Receiving Operating Characteristic (ROC).

Equation 1. Logistic Regression Function for landslide susceptibility

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

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 variable x_n
x_n	=	the variable number n

After the generation of a susceptibility map and the validation of the LOGIT model that generated it through the ROC value, a sensitivity analysis was performed to test how sensitive were the model outputs to changes in one, many or all its components. Sensitivity analysis is applied to determine the contribution of input parameters to the accuracy of the model prediction appraised in its outputs (Abbaszadeh Shahri et al., 2019; Poelmans and Van Rompaey, 2010). For this research, the referential metric selected was the ROC value as the output for all the generated simulations as experimented for SA by Poelmans & Van Rompaey (2010). Sensitivity Analyses were performed by two methods. The first one is called simple, univariate or “one-by-one” method, which is easy to assess. It consists 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 variable changes its coefficient, the others remain unaffected (fixed parameters) and left 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 factors ranging from 0.1 to 20 with an interval of 0.1 multiplied each of the coefficients of the ten variables, one at a time, to generate in total 2000 susceptibility maps, from which ROC values (outputs) were generated and plotted. Those ROC values higher than the reference ROC value would indicate that their corresponding models could be better calibrated than the reference one.

A second method to test the sensitivity uses random variations for all parameters, from one to all at a time, also within a range and with an interval. This is also called the Monte-Carlo or stochastic method. For this research, factors multiplying 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, according to the computer processing capacities. Once again, ROC values

(outputs) were generated and those higher than the reference ROC value would indicate that their corresponding models with their values configuration could be a better calibration than the reference itself (Bouyer, 2009).

Another performance test that determines the classification power of the model between occurrence and not occurrence (i.e. 2 samples, one for each category) is the Kolmogorov-Smirnov test, taken in this research as a measure of sensitivity of variables. This test was applied only to the results of the simple sensitivity analysis, only. The metric considered from this tool is a p-value that measures the significance (at an alpha value of 0.05) from the relation kept between the occurrence data sample and the non-occurrence one. This will be explained in detail in the results section.

To summarize, Table 4 shows all the methodology and specific tools applied for this research.

Table 4. Methodology summary, applied tools

	6-Variable Model	**8-Variable Model**	**10-Variable Model**	
	Weights-normalization	Weights-normalization	Weights-normalization	Percentile-normalization
LOGIT and referential landslide susceptibility map and ROC value	yes	yes	yes	yes
Sensitivity Analysis – Simple Method, variations plot and ROC values higher than reference	no	no	yes	yes
K-S test	no	no	yes	yes
Sensitivity Analysis – Monte Carlo Method, visualization of 2 best predictors and ROC values higher than reference	no	no	yes	yes

4. Analysis and Results

The data from the MCE (Multi-Criteria Evaluation) presented by the Municipality in 2015 (FUNEPSA et al.) was processed through map algebra, by adding the partial weights assigned to each of the 6 variables. It delivered scores from 8 (lowest landslide susceptibility) to 21 (highest landslide susceptibility), as shown in Figure 2.

Three logistic regression models were then applied and one variation of the third. The three models considered the weights normalization (also called pseudo-quantitative) with the scale from 1 to 4 of the variables given by the former study. The first one was run with the same data used for the initial map algebra calculation shown in Figure 2**Error! Reference source not found.**, which

contained six variables: *lithology, land use / vegetation coverage, seismic intensities, intense precipitations, soil stability after large events* and *slope*. The second model included two more variables: *population* and *floor area*, while keeping the weights-normalization (classified by natural breaks). The third model included two more variables: *road density* and *building footprint area*, also normalized. It is noticeable that, as new variables were added, the highest prediction values (coefficients) of them make them change its relative order. For the third model (10 variables) a variation in the way variables values were normalized was applied. We then applied a normalization according to percentiles. The two reference sensitivity maps based on 10 variables with different weighting methods can be seen in Figure 3. The results of the four logistic regression models are presented in Table 5 with their corresponding ROC value. In this article, order is referred as the relative position of the explanatory variable if they are sorted according to their coefficient from highest to lowest value.

Table 5. Output values from LOGIT model for Landslide Susceptibility in Quito

Code	Variable	6-Variable Model		8-Variable Model		10-Variable Model					
		Weights-normalization		Weights-normalization		Weights-normalization			Percentile-normalization		
		Coeff-icient	Desc-ending Order	Coeff-icient	Desc-ending Order	Coeff-icient	Desc-ending Order	p-value	Coeff-icient	Desc-ending Order	p-value
0	Intercept	-0.5281		-10.783		-4.1375		5.11e-29	-2.4317		1.75e-10
1	*Lithology*	0.3756	3	0.255	5	0.1905	5	4.35e-11	0.0160	2	3.33e-16
2	*Land use / vegetation coverage*	0.8483	1	0.425	4	0.0125	7	0.0006	0.0122	3	2.30e-12
3	*Seismic Intensity*	-0.1628	6	-0.091	7	-0.2004	9	0.0538	-0.0110	8	5.50e-08
4	*Intense Precipitations*	0.6528	2	0.450	2	0.3943	3	2.05e-09	0.0238	1	7.69e-33
5	*Stability after large events*	0.0247	5	0.116	6	-0.1526	8	0.7374	0.0047	5	0.9740
6	*Slope*	0.3756	4	0.445	3	0.3896	4	1.46e-09	-0.00045	10	0.8802
7	*Population*	-	-	0.684	1	0.5348	2	7.54e-21	0.0034	7	0.1247
8	*Road Density*	-	-	-	-	0.6101	1	1.02e-15	0.0052	4	9.43e-05
9	*Floor Area*	-	-	-0.128	8	0.0566	6	0.5540	-0.0040	6	0.6508
10	*Building Footprint Area*	-	-	-	-	-0.2364	10	0.2513	-0.0038	9	9.81e-05
	ROC value	**0.755**		**0.784**		**0.7928**			**0.7417**		

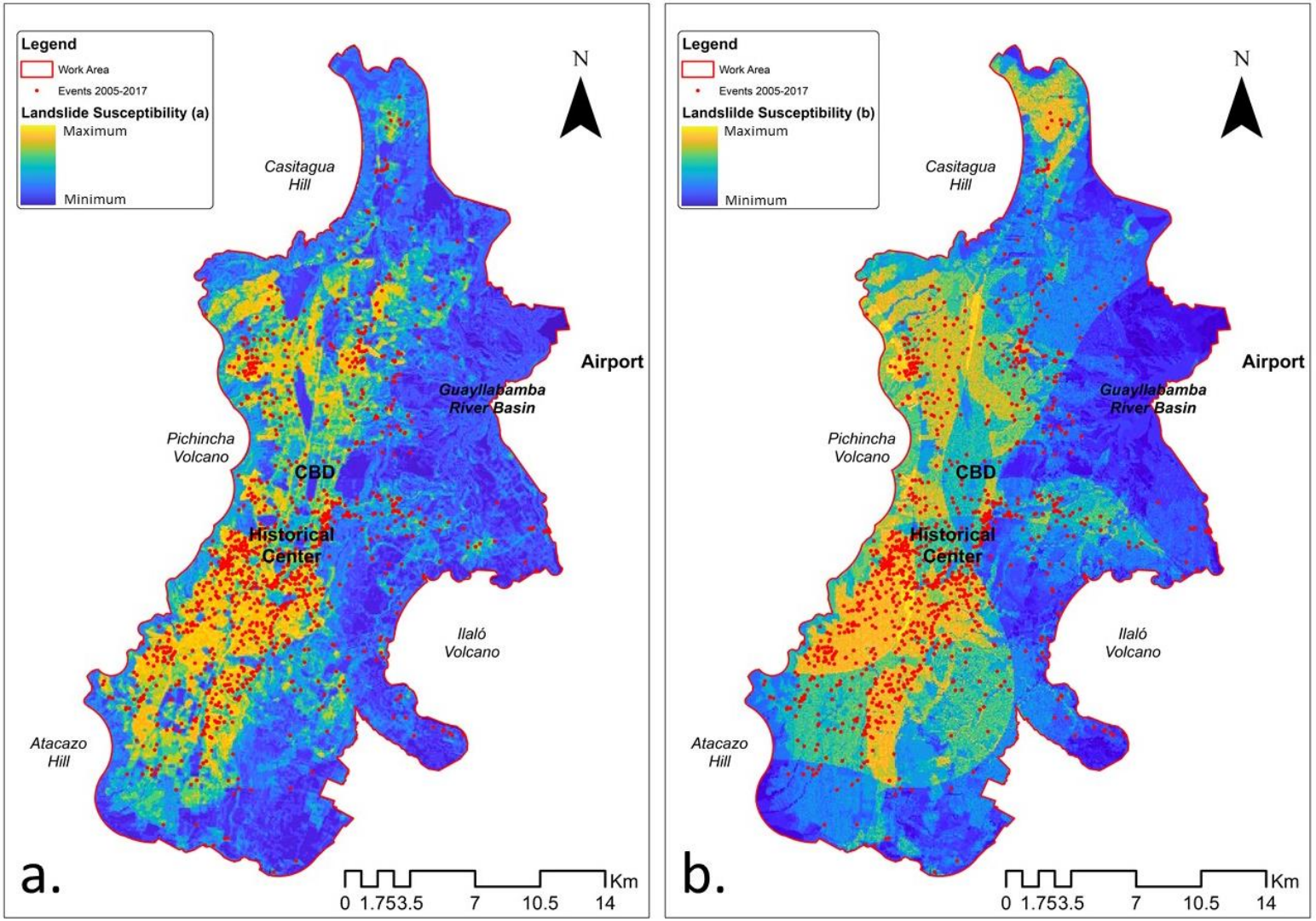

Figure 3. Landslide Susceptibility Reference Maps for Quito, from LOGIT modelling (a) weights normalization (b) percentile normalization.

Once the 10-variable model was executed by both normalization methods (weights and percentiles), a sensitivity analysis was performed for validation. This analysis was applied by the two methods explained in the methodology section of this article. The first, simple "one-by-one", method produced susceptibility maps, whose ROC calculations (determined as metric of the sensitivity), were plotted as observed in Figure 4. From this analysis, and from the range and interval to produce the coefficients' variations, there were 241 ROC values higher from the reference ROC value (0.7928) out of 2000 runs. Nevertheless, as seen in this graph, the ROC improvement related to these tests is relatively marginal to the reference. From the 2000 the simulations, the highest ROC value was 0.7943, which is almost 2% higher than the reference simulation ROC value. The coefficients that have the stronger impacts on results are the *population, slope* and *road density*.

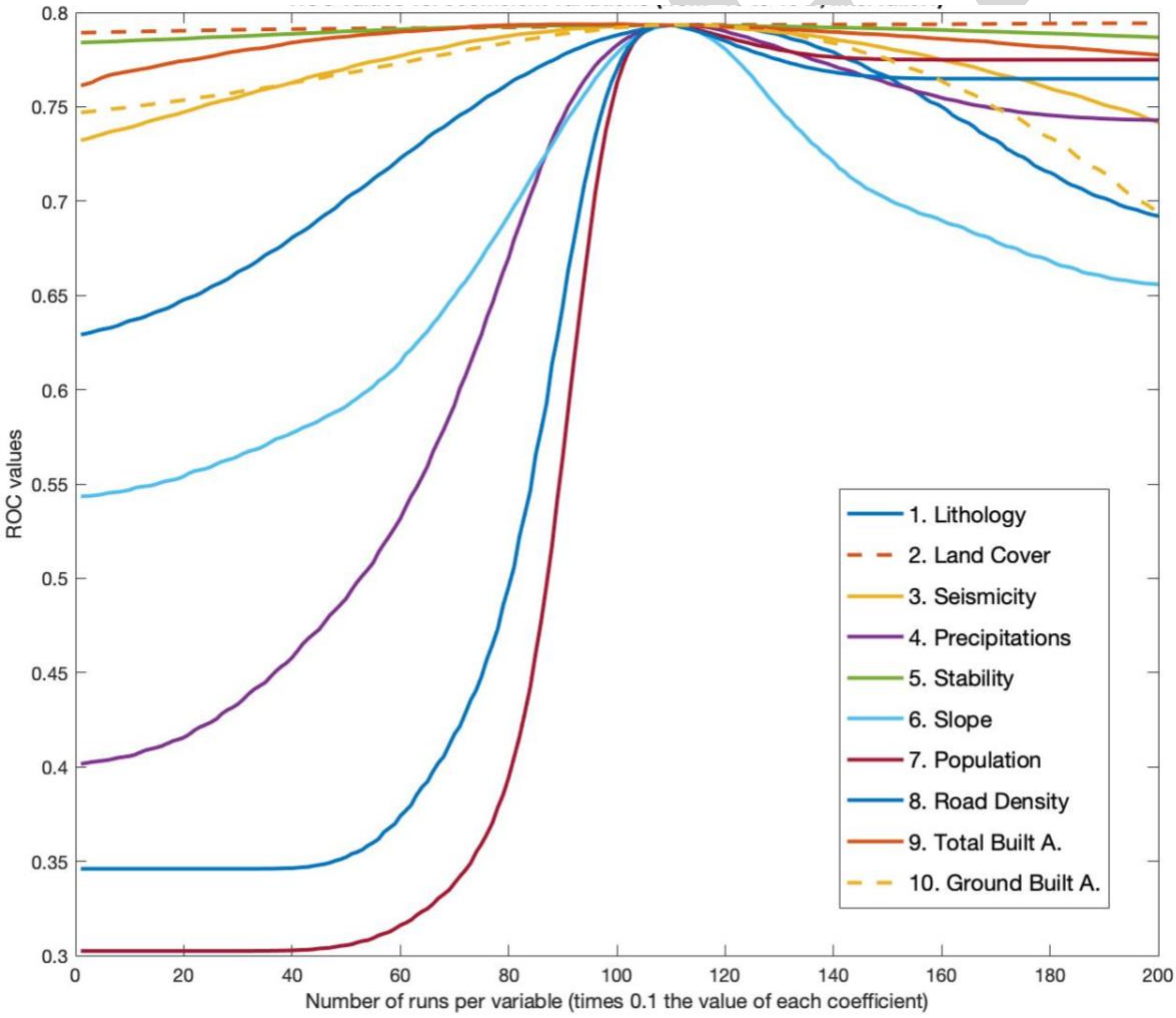

Figure 4. Univariate sensitivity analysis for LOGIT model (weighted normalization): ROC values as metric of coefficient variations.

The same test was applied to the 10-variable LOGIT model. In that case, there were only 34 ROC values higher from the reference ROC value (0.7928) out of 2000 programmed runs. The plotting of all ROC values derived from the one-by-one variations' susceptibility maps/datasets can be seen in Figure 5. *Precipitations, Lithology and Road Density* are here the most relevant factors affecting the sensitivity of the model.

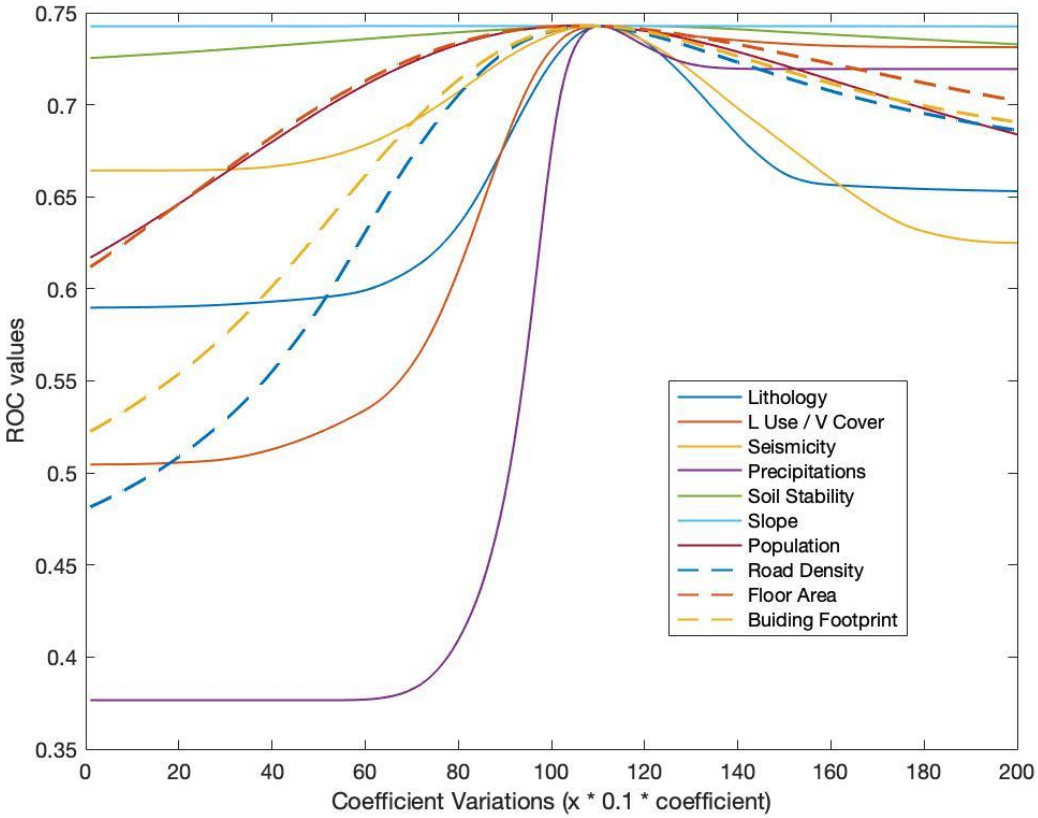

Figure 5. Univariate sensitivity analysis for LOGIT model (percentile normalization): ROC values as metric of coefficient variations.

A Kolmogorov-Smirnov (K-S) test was applied to these results for both weights and percentile-based normalization methods. This is another way to approach sensitivity, by measuring a p-value that measures the probability that, for each simulation and its susceptibility map, the distribution of the cell values corresponding to event occurrence belong to the same distribution of the cell values corresponding to non-event occurrence. For the weights-normalized method, the p-value of the test, with an alpha value of 0.05 showed that in the case of 13 resulting landslide susceptibility maps and datasets out of the 2000 generated were not significant. These 13 corresponded to variations in 3

variables: *intense precipitations, population* and *road density*. For the percentile-normalized method, with the same alpha value and number of maps, the resulting p-values showed that only por the case of 1 map and dataset was non-significant, corresponding to the *intense precipitations* variable. These results are illustrated in Figure 6.

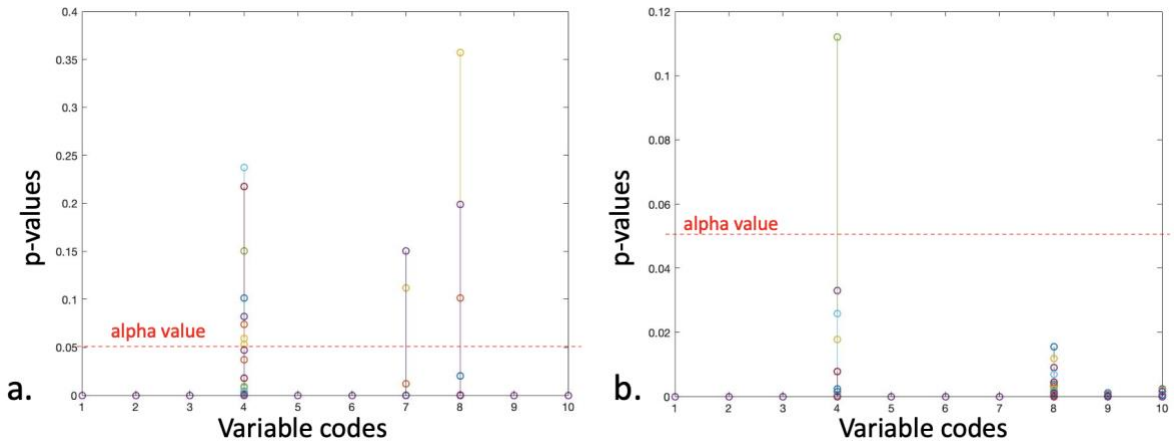

Figure 6. The p-values of K-S test (2-sample) applied to 2000 landslide susceptibility datasets resulting from the simple (1-by-1) susceptibility analysis of LOGIT model. (a) weights normalization (b) percentile normalization.

The second sensitivity analysis method applied was the random/stochastic method, also called the MonteCarlo method. The ROC value was again used as a metric to test the sensitivity through this method for both weights and percentile normalization types. For both normalization types 8000 runs of the models were applied through the MonteCarlo methodology. It can be seen an important difference between the weights and the percentile normalization models. The weights-normalized model generated 350 ROC values (out of the 8000) higher than the referential 0.7928 ROC value, while the percentile-normalized one generated 4440 values (out of the 8000) higher than the referential 0.7417 ROC value. A set of the best 10 out of the 4440 factors combinations to change the referential coefficients are presented in Table 6.

Table 6. Randomly selected combination of factors to multiply coefficients of the LOGIT model (percentile-normalized) to calibrate it for optimal results in defining the landslide susceptibility map for Quito, Ecuador.

ROC	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7	Factor 8	Factor 9	Factor 10
Variable	*Slope*	*Land Use / Veg Cover*	*Seismic Intensity*	*Intense Precipitations*	*Soil Stability*	*Slope*	*Population*	*Road Density*	*Floor Area*	*Building Footprint*
0,7417 **	0,016*	0,0122 *	-0,011 *	0,0238 *	0,0047 *	-0,000454 *	0,0034 *	0,0052 *	-0,004 *	-0,0038 *
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 referential coefficients for each variable provided from the former LOGIT model application (percentile-normalization)
** The referential ROC values provided after the former LOGIT model application (percentile-normalization)

The two variables with the highest values of coefficients (see Table 5) were selected to compare their ROC values with combination of the variations of both coefficients, while the rest 8 remained as the referential results. They were illustrated in graphs to approximately identify the coordinates (combination) that performs the highest ROC value. For the weights-normalized model these variables were *Population* and *Road Density* and the random outputs selected 12 combinations, but none of them were higher than the reference (see Figure 7). For the weights-normalized model these variables were *Population* and *Road Density* and the random outputs selected 18 combinations. In this last case, all combinations performed higher ROC values than the referential (0.7417), as shown in Figure 8.

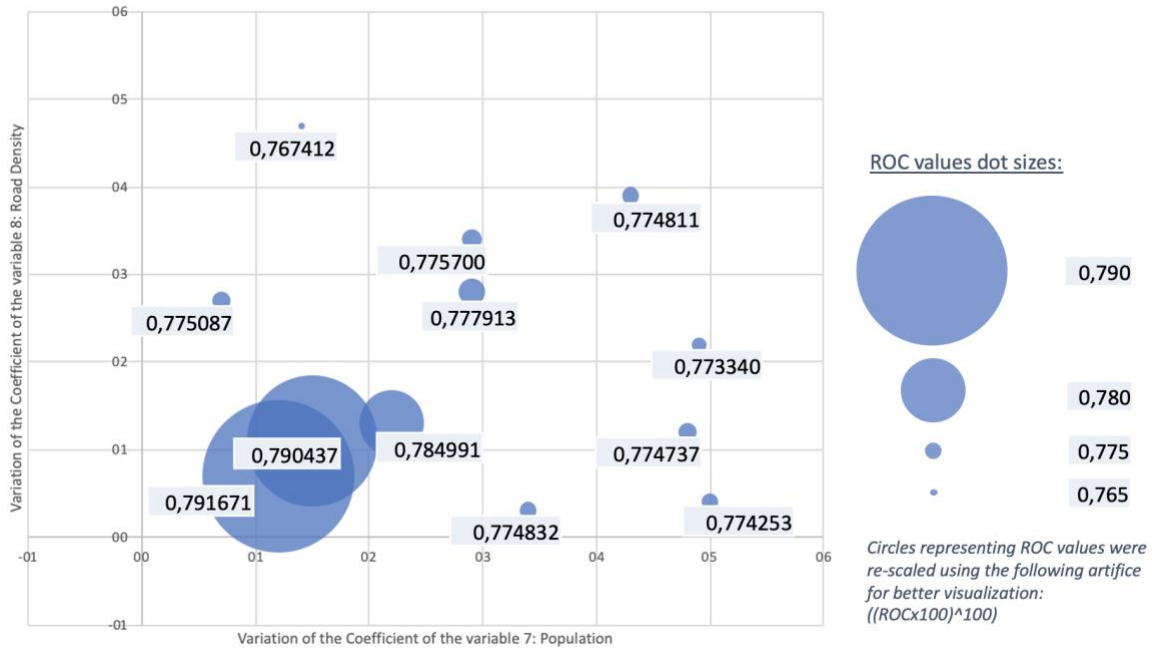

Figure 7. ROC values resulting from random combinations of population and road density coefficients variations from the referential outputs from the LOGIT model (weights normalization).

Figure 8. ROC values resulting from random combinations of Lithology and Precipitations coefficients variations from the referential outputs from the LOGIT model (percentile normalization).

5. Discussion

A relevant issue for further research is to consider as a parameter the way the data is processed. Van Dessel, van Rompaeya, & Szilassi, (2011) stress on the incidence that the quality of the input maps has on the resulting coefficients of logistic regression models applied for landslide susceptibility analysis. This research implements a so-called pseudo-quantitative method, which, initially consider weights from a multicriteria assessment of landslide susceptibility (Leoni et al., 2009; Lombardo and Mai, 2018), which later is, as commonly, migrated to a LOGIT model. This pondering is usually done according to criteria determined by experts and their knowledge of the study area. Other works have adopted other ways of applying this pondering based on the frequency ratio of events that a specific class has (Bui et al., 2020), which would be ideal when complete events report is available, which is not the case for Quito. For this research, part of the processing of the variables was normalization based on both weights and percentiles methods. In this regard, there are differences that should not be overlooked when running the LOGIT between the weighted method and the percentile-normalized one for the variables tuning.

By seizing the availability of data provided by the municipality of Quito, this research considered to include more variables to the previous landslide susceptibility study done by the municipality. These were *population, road density, floor area* and *building footprint area*, which may have helped to

be more specific in characterizing the urban category from the *land use / vegetation coverage* variable from the former study. It could be expected that population, which appears as an important predictor after the LOGIT application, is related to *building footprint area* and *floor area*. Nonetheless, the latter variables did not perform to be relevant predictors. This might be related with the fact that the largest *floor area* volumes are concentrated on the center-north of the main city, an area where self-built / informal construction is low, and building are often medium-rise with appropriate construction techniques for soil management. A further step of this study, beyond a better tuning of the model, could be a complement to assess building scales, from a vulnerability and uncertainty quantification approach, considering the heterogeneity of urban fabrics as the ones performed by Kaynia et al. (2008) and Du et al. (2013), which may indeed steer to enhance the data quality surveying of building conditions and soil management to collect it at a large scale, as the case of Quito.

Another complementary remark to the results in terms of policy implications is that when observing the relevance of the *road density* variable as predictor it does not necessarily mean that roads *per se* promote topples risk. In parallel way, for the case of translational landslides, attention should focus on sloppy soils that alternate soft and hard rock sliding strata (Du et al., 2020), which highlights *lithology* as predictor, but the territory is too heterogeneous to generalize this assumption. In fact, heterogeneity can importantly affect the model (Wang et al., 2020). Considering roads as promoters of development, it is necessary to research on the detail of how vulnerability is produced at the household scale in terms of construction techniques and soil management in the resulting landslide-susceptible areas from this model. Due to the costly and complex nature of these studies, it could be thought as a sampling studies, to build a multiple-case analysis.

When revising the quality of the data corresponding to soil stability after previous large events and the seismic intensities, it is noticeable that the level of detail is poor, corresponds to a smaller scale and collects few information. This might be resulting in imprecise outputs of the models here referred. Micro-zoning seismicity studies are being surveyed at the moment as part of a long-term project. In the future, this could help precise results in the modeling for landslide susceptibility studies, as well as together with better quality and new data, which seems to be promising from the municipal institution of Quito.

6. Conclusions

This article presents the application of a binary logistic regression model with the objective of defining a landslide susceptibility map for the urban area of Quito Metropolitan District in Ecuador (main city and satellite urban areas). A landslide events database covering the period 2005-2017 was used as dependent binary variable. Ten explanatory variables were tested*: lithology, land use / vegetation coverage, seismic intensities, intense precipitations, soil stability based on previous events, slope, population, road density, floor area, and ground-built area*. Two normalization methods were applied. The first method considered weights previously assigned to the variables scores and the second one applied a percentile scale.

After running the LOGIT model, the weighted-scaled method had a Receiver Operating Characteristic (ROC) of 0.7928, while the percentile-scaled one obtained 0.7421. Regarding the resulting coefficients for the explanatory variables, the first method performed more stable values than the second one (more oscillation), after simulation of the models for several times. According to these values, the first method portrayed as best predictors *road density*, *population, intense precipitations* and *slope*, in that order. The second method portrayed *intense precipitations, lithology, land use / vegetation coverage* and *road density* as best predictors, in that order. Both methods presented ground-built area and seismic intensities with negative values and as the lowest coefficients for the LOGIT model. The other variables fluctuate around zero.

A subsequent analysis was applied to the LOGIT model to measure how sensitive were all explanatory variables as parameters for the model. Two methods were used for this matter. These were the one-by-one 'simple' method (variation of the coefficient of one variable at a time, while the others remain unchanged) and the random-variation one, also called MonteCarlo (in which all variables may change randomly). The Area Under the ROC Curve (AUC), also called ROC value has been adopted, considering it is a common measure to evaluate prediction accuracy of models for natural hazards (Abbaszadeh Shahri et al., 2019; Wang et al., 2020). By interpreting the ROC value as a metric from the simple method, it can be observed that the *slope, road density, intense precipitations* and *population* variables' curves importantly vary along with the variations of the factors that multiply the reference coefficients. This is the case for the weights-normalized model (see Figure 4). For the case of the percentile normalized model (see Figure 5), the most sensitive variables' curves are *intense precipitations, land use/vegetation coverage, lithology* and *road density.*

When executing the LOGIT model, the weights-normalized model (ROC=0.7928) proved to be more stable and reliable than the percentile-normalized one (ROC=0.7417), considering the results from the MonteCarlo Sensitivity Analysis, from which out of 8000 runs, 4440 showed higher ROC values than the referential one, reaching values of almost 0.8. This differs substantially from the weights-normalized model, which in the MonteCarlo application only provided 350 higher ROC values from the reference. This means that the calibration of the percentile-normalized model's variables can still be adjusted through their coefficients to improve predictability.

References

Abbaszadeh Shahri, A., Spross, J., Johansson, F., Larsson, S., 2019. Landslide susceptibility hazard map in southwest Sweden using artificial neural network. Catena 183, 104225. https://doi.org/10.1016/j.catena.2019.104225

Alcaldia del Distrito Metropolitano de Quito, 2018. Resolución Administrativa A008, Ordenanzas y Resoluciones. Alcaldia del Distrito Metropolitano de Quito, Quito.

Asamblea Nacional del Ecuador, 2014. Ley Orgánica Reformatoria al Código Orgánico de Organización Territorial, Autonomía y Descentralización. Asamblea Nacional Ecuador, Quito.

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.

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, 104426. https://doi.org/10.1016/j.catena.2019.104426

Comunidad Andina, 2017. Estrategia Andina para la Gestión del Riesgo de Desastres - EAGRD. Comunidad Andina, Perú.

Concejo del Distrito Metropolitano de Quito, 2011. Ordenanza Metropolitana No. 172. Régimen de Suelo de Quito. Concejo del Distrito Metropolitano de Quito, Quito.

Concejo del Distrito Metropolitano de Quito, 2003. Ordenanza Metropolitana No. 095. Régimen de Suelo de Quito. Concejo del Distrito Metropolitano de Quito, Quito.

D'Ercole, R., Hardy, S., Metzger, P., Robert, J., 2009. Vulnerabilidades urbanas en los países andinos. Introducción general. Bull. l'Institut français d'études Andin. 38, 401–410. https://doi.org/10.4000/bifea.2222

Du, J., Yin, K., Nadim, F., Lacasse, S., 2013. Quantitative vulnerability estimation for individual landslides, in: Ponts, P. des (Ed.), Proceedings of the 18th International Conference on Soil Mechanics and Geotechnical Engineering, Paris 2013. Presses des Ponts, Paris, pp. 2181–2184.

Du, Y., Xie, M., Jia, J., 2020. Stepped settlement: A possible mechanism for translational landslides. Catena 187, 104365. https://doi.org/10.1016/j.catena.2019.104365

Feizizadeh, B., Blaschke, T., 2014. An uncertainty and sensitivity analysis approach for GIS-based multicriteria landslide susceptibility mapping. Int. J. Geogr. Inf. Sci. 28, 610–638. https://doi.org/10.1080/13658816.2013.869821

FUNEPSA, Rivera, M., Salazar, A., Carvajal, A., Galárraga, R., Plaza, G., Singaucho, J., Álvarez, B., 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. Serv. Nac. Geol. y Minería 432.

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 Eng. 212, 1046–1053. https://doi.org/10.1016/j.proeng.2018.01.135

Highland, L.M., Bobrowsky, P., 2008. The Landslide Handbook — A Guide to Understanding Landslides. U.S. Geological Survey, Reston.

INEC, 2016. Proyeccion Cantonal Total 2010-2020.

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. Eng. Geol. 101, 33–48. https://doi.org/10.1016/j.enggeo.2008.03.008

Kirschbaum, D., Stanley, T., 2018. Satellite-Based Assessment of Rainfall-Triggered Landslide Hazard for Situational Awareness. Earth's Futur. 1–19. https://doi.org/10.1002/2017EF000715

Leoni, G., Barchiesi, F., Catallo, F., Dramis, F., Fubelli, G., Lucifora, S., Mattei, M., Pezzo, G., Puglisi, C., 2009. GIS methodology to assess landslide susceptibility: Application to a river catchment of central Italy. J. Maps 5, 87–93. https://doi.org/10.4113/jom.2009.1041

Lombardo, L., Mai, P.M., 2018. Presenting logistic regression-based landslide susceptibility results. Eng. Geol. 244, 14–24. https://doi.org/10.1016/j.enggeo.2018.07.019

MDMQ, 2017. Estrategia de Resiliencia Distrito Metropolitano de Quito.

MDMQ, 2015. Diagnóstico Eje Territorial - Plan de Desarrollo y Ordenamiento Territorial. Quito.

Norman, G., 2010. Likert scales, levels of measurement and the "laws" of statistics. Adv. Heal. Sci. Educ. 15, 625–632. https://doi.org/10.1007/s10459-010-9222-y

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 la Univ. Sevilla 971–987.

Pasta, D.J., 2009. Learning When to Be Discrete : Continuous vs. Categorical Predictors. SAS Glob. Forum 2009, Stat. Data Anal. 1–10.

Poelmans, L., Van Rompaey, A., 2010. Complexity and performance of urban expansion models. Comput. Environ. Urban Syst. 34, 17–27. https://doi.org/10.1016/j.compenvurbsys.2009.06.001

Puente-Sotomayor, F., Teller, J., 2019. Land Policies for Landslide Risk Reduction for the Urban Andes. Manuscr. under Rev.

Puente-Sotomayor, F., Villamarin-Jurado, P., Cevallos-Serrano, 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. Asociación Geográfica del Ecuador, Quito, pp. 1–15.

Rebotier, J., 2016. El Riesgo y su Gestión en Ecuador - Una Mirada de Geografía Social y Política. Centro de Publicaciones Pontificia Universidad Católica del Ecuador, Quito.

Reichenbach, P., Rossi, M., Malamud, B.D., Mihir, M., Guzzetti, F., 2018. A review of statistically-based landslide susceptibility models. Earth-Science Rev. 180, 60–91. https://doi.org/10.1016/j.earscirev.2018.03.001

Ronchetti, F., Alessandro, C., Kollarits, S., Leber, D., Papez, J., Plunger, K., Prešeren, T., Schnetzer, I., Stefani, M., 2013. Improve Information Provision for Disaster Management: MONITOR II, EU Project, in: Margottini, C., Canuti, P., Sassa, K. (Eds.), Landslide Science and Practice. Volume 7: Social and Economic Impact and Policies. Springer, Rome, pp. 47–54. https://doi.org/10.1007/978-3-642-31313-4_7

Sepúlveda, S.A., Petley, D.N., 2015. Regional trends and controlling factors of fatal landslides in Latin America and the Caribbean. Nat. Hazards Earth Syst. Sci. 15, 1821–1833. https://doi.org/10.5194/nhess-15-1821-2015

UNISDR, 2018. UNISDR Annual Report 2017. Geneva.

USGS, 2004. Landslide Types and Processes. Highw. Res. Board Spec. Rep.

van Dessel, W., van Rompaeya, A., Szilassi, P., 2011. Sensitivity analysis of logistic regression parameterization for land use and land cover probability estimation. Int. J. Geogr. Inf. Sci. 25, 489–508. https://doi.org/10.1080/13658810903194256

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, 104425. https://doi.org/10.1016/j.catena.2019.104425

Williams, R., 2019. Ordinal Independent Variables. Notre Dame.

Abbreviations

AUC	Area Under the Curve (ROC value)
EMS	European Macroseismic Scale
EPMMOP	Empresa Pública Metropolitana de Movilidad y Obras Públicas (Metropolitan Public Enterprise of Mobility and Public Works)
GIS	Geographic Information Systems
INEC	Instituto Nacional de Estadísticas y Censos de Ecuador (National Institute of Statistics and Census of Ecuador)
MCE	Multi-Criteria Evaluation
MDMQ	Municipio del Distrito Metropolitano de Quito (Government of the Metropolitan District of Quito)
ROC	Receiving Operator Characteristic
SA	Sensitivity Analysis
SGP	Secretaría General de Planificación (General Planning Secretariat of the MDMQ)
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)

Figures

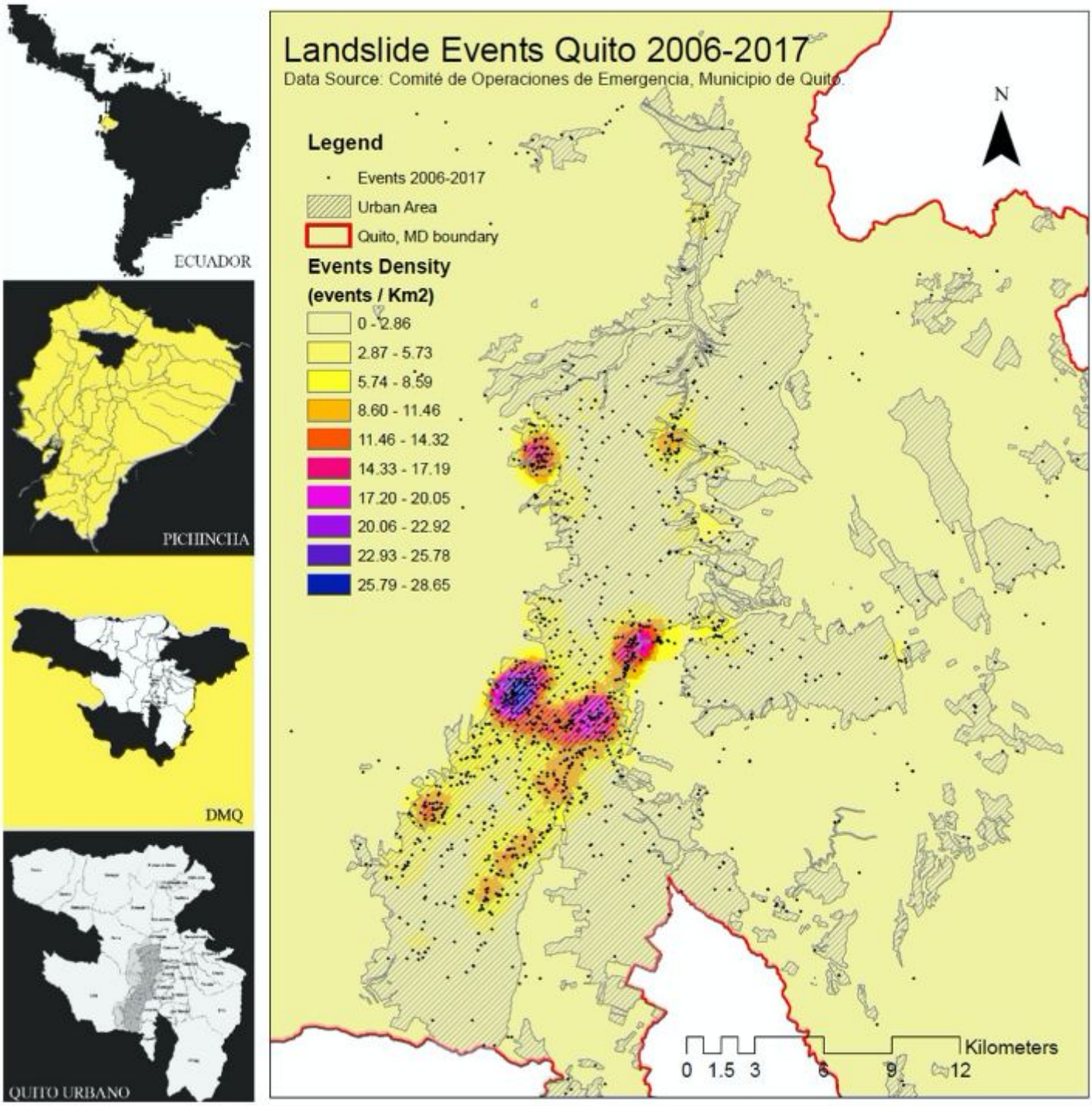

Figure 1

Study Area in the Metropolitan District of Quito, with 2005-2017 landslide events and heatmap.

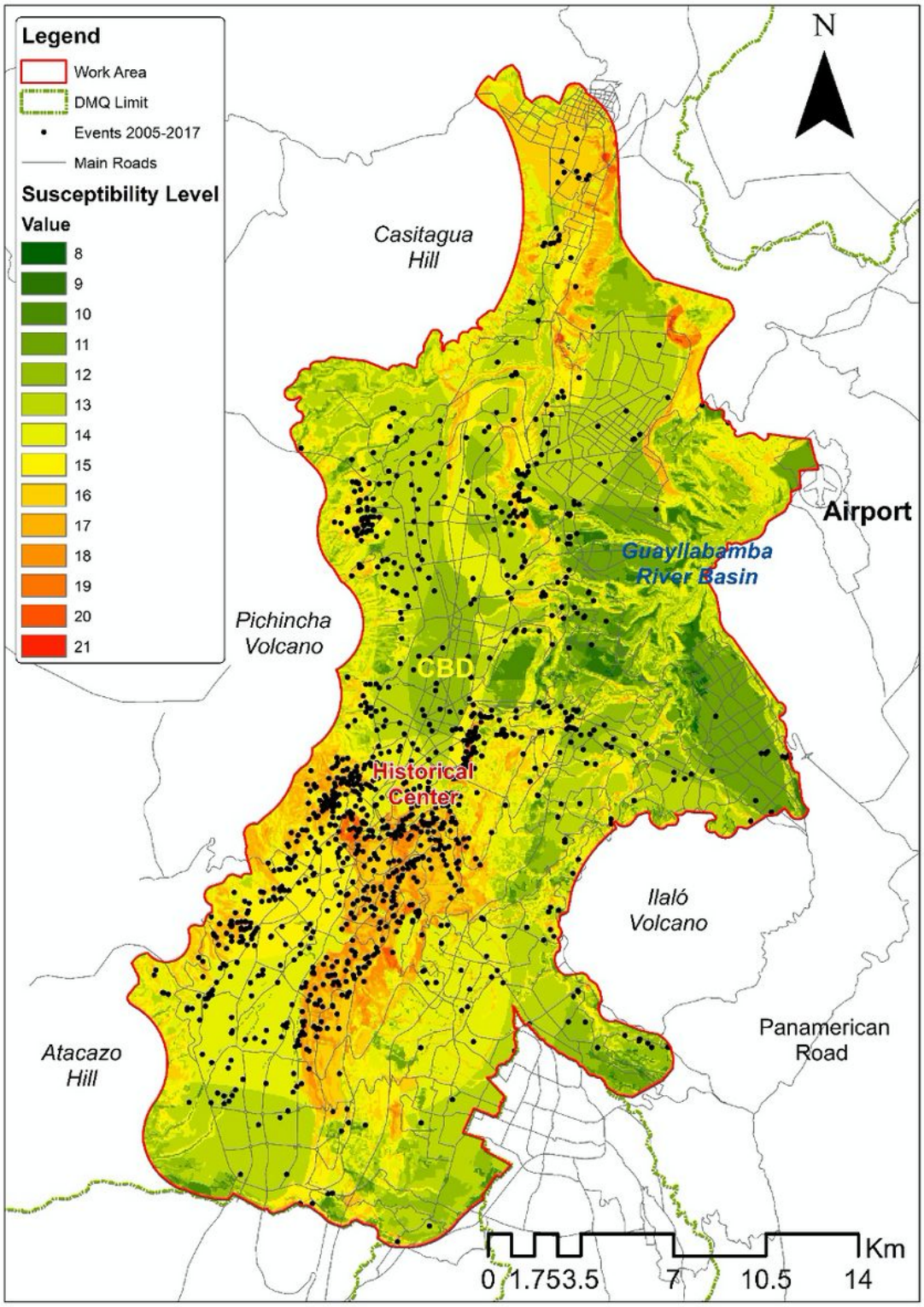

Figure 2

Landslide Susceptibility Map for Quito including 2005.2017 events sites. Data Source: Quito Municipality

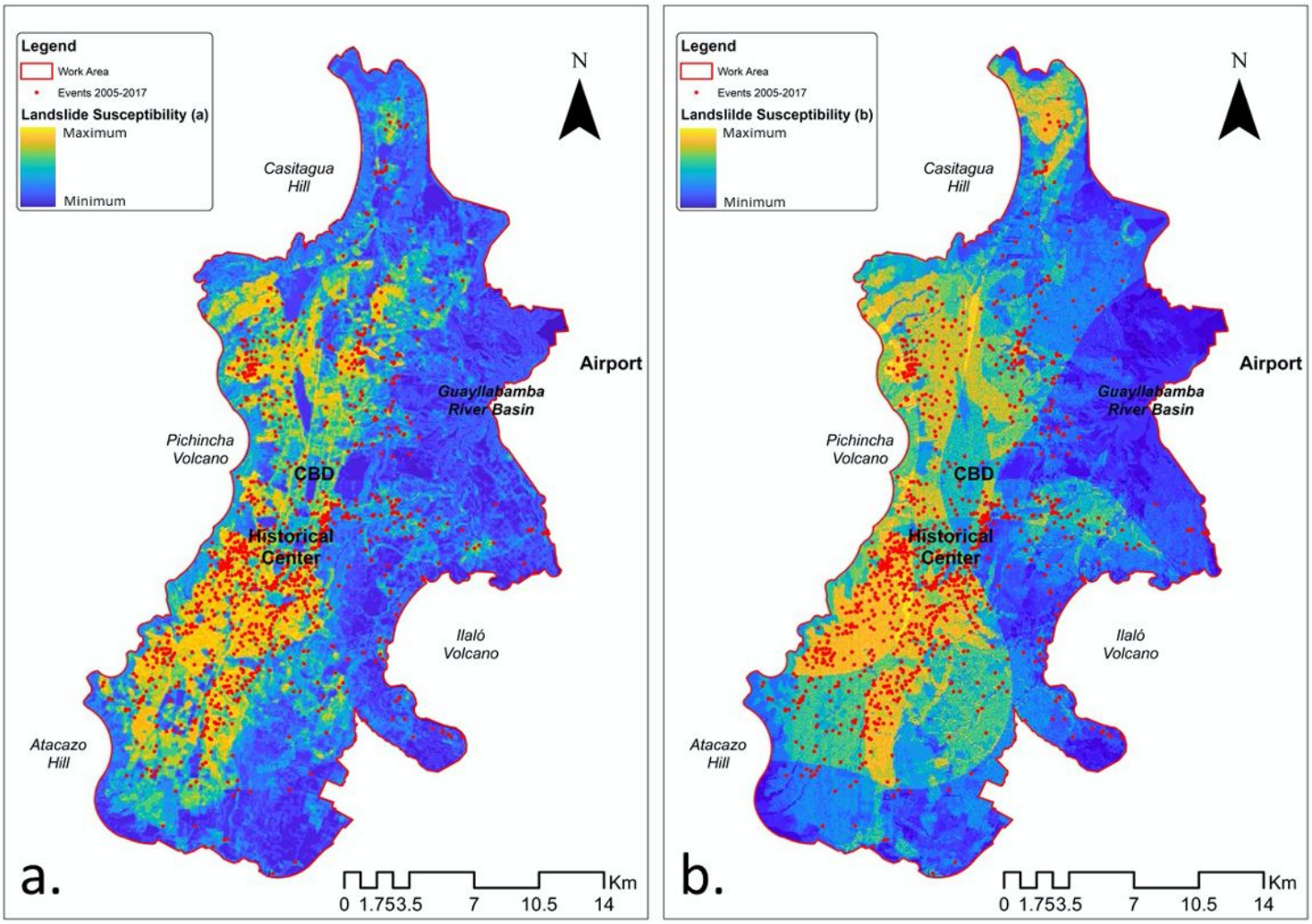

Figure 3

Landslide Susceptibility Reference Maps for Quito, from LOGIT modelling (a) weights normalization (b) percentile normalization.

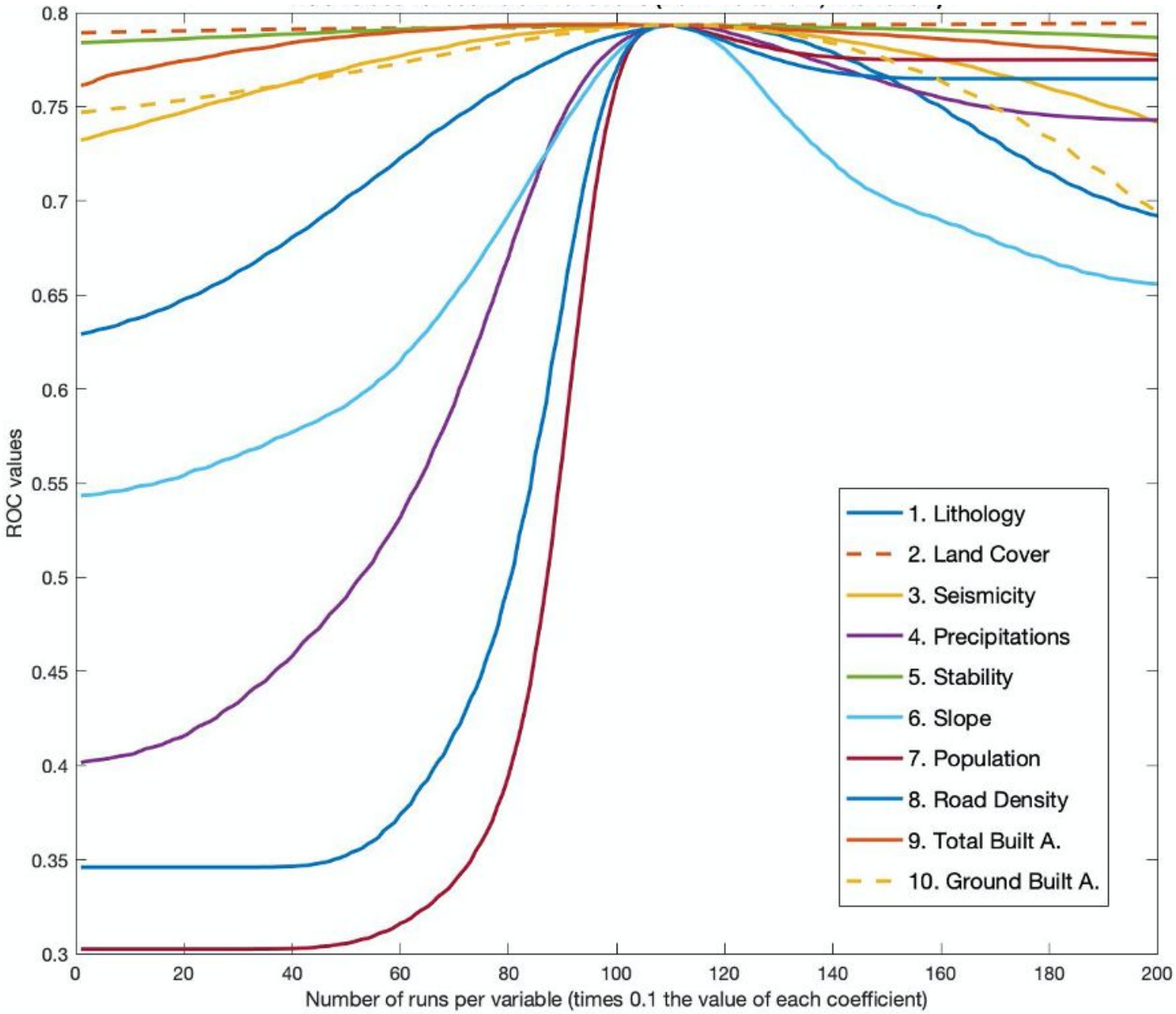

Figure 4

Univariate sensitivity analysis for LOGIT model (weighted normalization): ROC values as metric of coefficient variations.

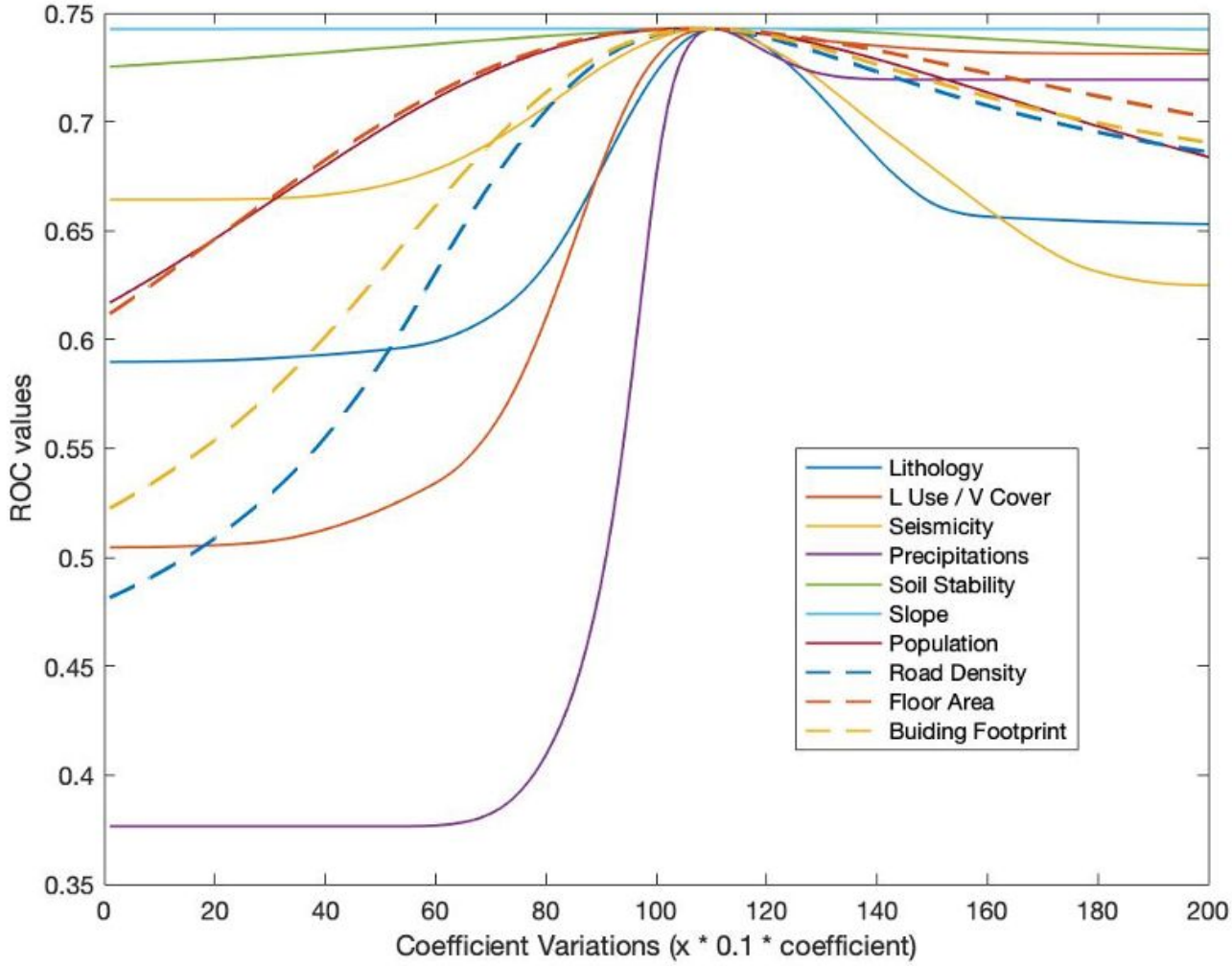

Figure 5

Univariate sensitivity analysis for LOGIT model (percentile normalization): ROC values as metric of coefficient variations.

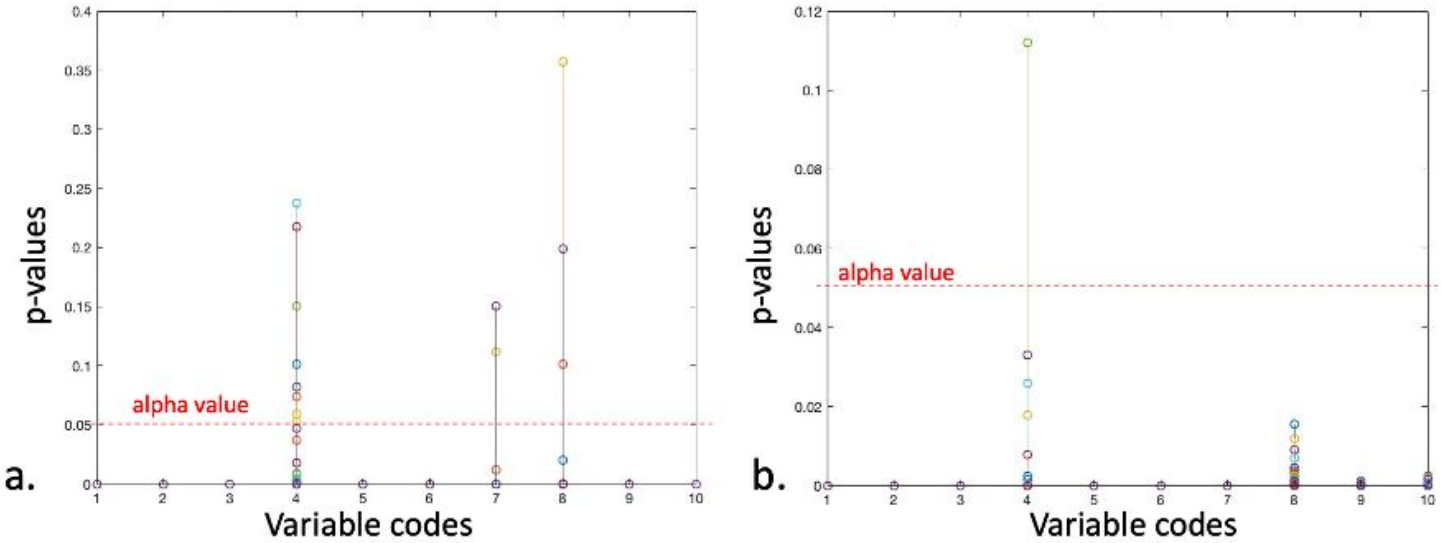

Figure 6

The p-values of K-S test (2-sample) applied to 2000 landslide susceptibility datasets resulting from the simple (1-by-1) susceptibility analysis of LOGIT model. (a) weights normalization (b) percentile normalization.

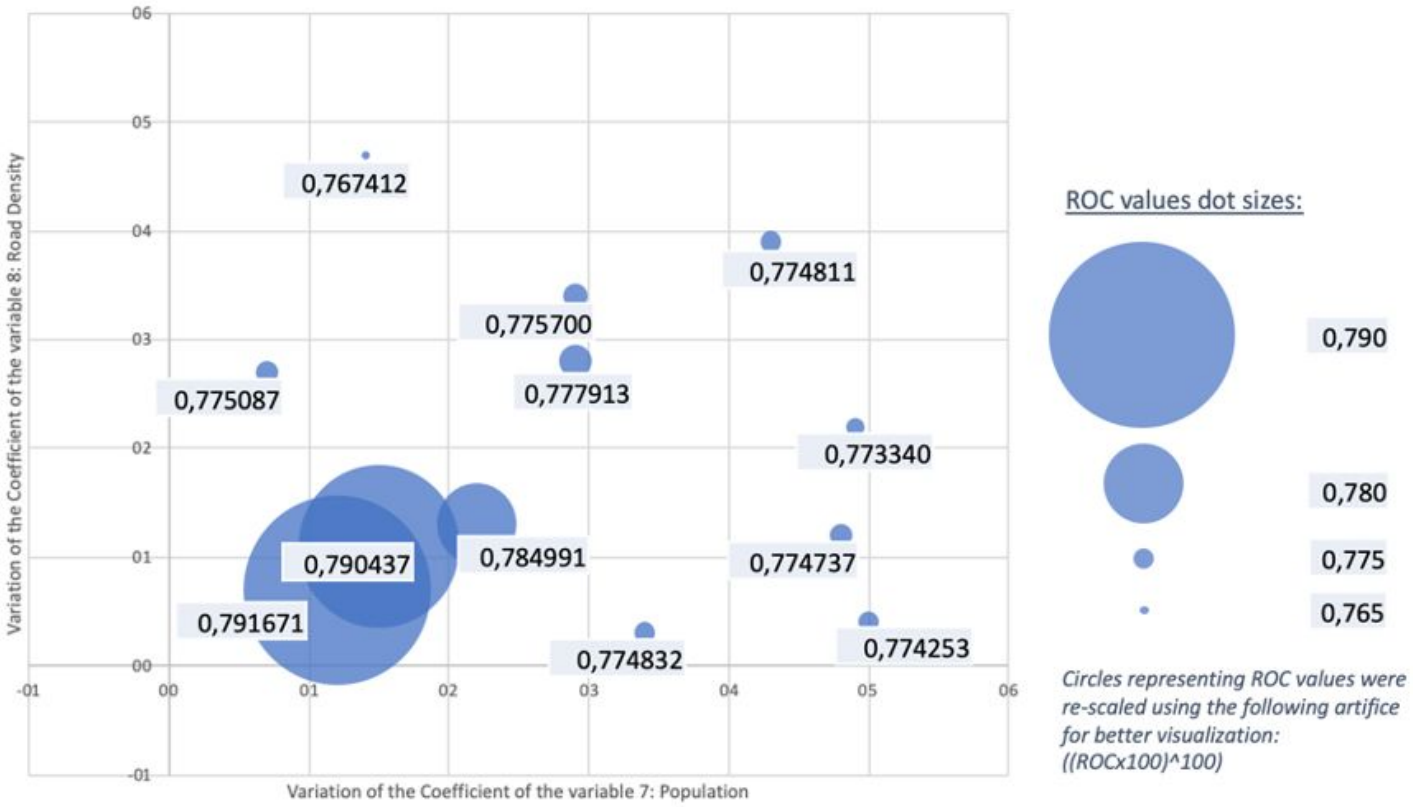

Figure 7

ROC values resulting from random combinations of population and road density coefficients variations from the referential outputs from the LOGIT model (weights normalization).

Figure 8

ROC values resulting from random combinations of Lithology and Precipitations coefficients variations from the referential outputs from the LOGIT model (percentile normalization).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- 04tab1inputdataa.xls
- 05tab2weightsnormalizationa.xls
- 06tab3percentilenormalizationa.xls
- 09tab6alternativecoefffactorsa.xls
- 08tab5outputsLOGITa.xls
- 07tab4methodsa.xls