Article

Different Approaches to Use Morphometric Attributes in Landslide Susceptibility Mapping Based on Meso-Scale Spatial Units: A Case Study in Rio de Janeiro (Brazil)

Vanessa Canavesi [1,2,*], **Samuele Segoni** [1], **Ascanio Rosi** [1], **Xiao Ting** [3], **Tulius Nery** [2], **Filippo Catani** [1] **and Nicola Casagli** [1]

[1] Department of Earth Science, Florence University, Via Giorgio La Pira, 4, 50121 Florence, Italy; samuele.segoni@unifi.it (S.S.); ascanio.rosi@unifi.it (A.R.); filippo.catani@unifi.it (F.C.); nicola.casagli@unifi.it (N.C.)

[2] National Center for Monitoring and Early Warning of Natural Disasters (CEMADEN), Estrada Doutor Altino Bondensan, 500-Distrito de Eugênio de Melo, São José dos Campos SP 12247-016, Brazil; tulius.nery@cemaden.gov.br

[3] Faculty of Engineering, China University of Geosciences, Wuhan 430074, China; shouting1993@cug.edu.cn

[*] Correspondence: vanessa.canavesi@cemaden.gov.br; Tel.: +55-1232050130

check for updates

Abstract: Landslide susceptibility maps are widely used in landslide hazard management. Although many models have been proposed, mapping unit definition is a matter that still needs to be fully examined. In the literature, the most reported mapping units are pixels and slope units, while in this work, developed in the Rio de Janeiro region (Brazil), the use of drainage basins as a mapping unit is examined; even if their use leads to the definition of maps with a coarser spatial resolution than pixels-based maps, they convey information that can be easily and rapidly handled by civil defense organizations. However, for the morphometrical characterization of entire basins, a standardized procedure does not exist, and the susceptibility results may be sensitive to the approach used. To investigate this issue, a random forest model was used to assess landslide susceptibility, using 12 independent variables: four categorical (land use, soil type, lithology and slope orientation) and eight numerical variables (slope gradient, elevation, slope curvature, profile curvature, planar curvature, flow accumulation, topographic wetness index, stream power index). For each basin, the numerical variables were aggregated according to different approaches, which, in turn, were used to set up four different model configurations: i) maximum values, ii) mean values, iii) standard deviation values, iv) joint use of all the above. The resulting maps showed noticeable differences and a quantitative validation procedure showed that the best configurations were the ones based on mean values of independent variables, and the one based on the combination of all the values of the numerical variables. The main outcomes of this work consist of a landslide susceptibility map of the study area, to be used in operational procedures of risk management and in some insights on the best approaches to aggregate raster cell data into wider spatial units.

Keywords: landslide; susceptibility; basin; spatial unit; sensitivity; random forest

1. Introduction

Landslides are a geomorphological process responsible for deaths and damages worldwide [1] and landslide susceptibility maps (LSM), which represent the spatial probability of landslide occurrence in a given location [2], are one of most widely used tools for the related hazard management. Landslide

susceptibility is usually assessed by a study of the spatial distribution of static predisposing factors and their relationship with the location of the landslides occurred in the past [3–5], as it is postulated that future slope failures will be more likely to occur under the conditions that led to past and present instability [6,7].

LSM have been produced applying a wide range of mathematical techniques, from the most traditional statistic approaches like frequency ratio [8], discriminant analysis [9,10] and logistic regression [11–13], to more recent and more advanced techniques, like artificial neural network [5,14], machine learning [15,16] and multi criteria decision analysis [17]. Many studies highlighted that different models can produce different results but, to date, no model has been proved to be better than all the others [5,8,10,12,16,17]. Indeed, another series of studies showed that the same model in the same test site can provide very different outcomes, depending on the approaches used to prepare the input data [18–20].

Independently from the mathematical model used in the landslide susceptibility assessment, geographic information systems (GIS) remain a fundamental tool for LSM studies, as they can be used in data preparation, data analysis and in drawing the outputs [21–23]. One of the advantages of the use of GIS in LSM is the possibility of overlaying different thematic maps and analyzing the spatial distribution of their attributes across appropriate spatial units. Concerning the spatial units of analysis and of representation of the resulting outputs, a recent review by Reichenbach et al. [24] showed that pixels are by far the most used mapping unit in the international literature, and only a minor number of works exists that use unique condition units (units defined in terms of the possible combinations of a limited set of input parameters in a given space), slope units or other techniques of segmentation of the territory [13,25–28].

As a consequence, we believe that the application of unit-based landslide susceptibility models instead of pixel-based models has not been deeply investigated. For instance, the recent literature is rich of examples of comparisons among different models [8], sensitivity analyses to model settings [19], sampling strategies [29], parameterization of variables [30] or pixel cell size [31], but despite a few attempts [32,33], scarce attention has been paid on the different approaches to generalize input data from the original raster format to the polygonal nature that characterizes other spatial units.

The main objectives of the present work are twofold. The first practical objective is to produce an LSM for the mountain region north of Rio de Janeiro, Brazil, to assist civil defense agencies in hazard management procedures. Another important research objective of this work is to cope with a source of uncertainty that affects LSM based on spatial units larger than the pixel size of the input parameters: how to generalize the fine-level pixel information at the level of the chosen spatial unit. In this study, a state-of-the-art machine learning algorithm (random forest [19,34]) has been used for landslide susceptibility modeling, and the hydrologic basins of the State of Rio Janeiro have been used as spatial units for the analyses. The approaches of characterizing the basins in term of the mean, the maximum and the standard deviation of the values of the numerical predisposing factor has been tested, highlighting noticeable differences in the resulting LSM and providing insights to decide which is the best option and if these strategies can be combined.

2. Materials and Methods

2.1. Test Site Description

The study area comprises the municipalities of Petrópolis, Teresópolis and Nova Friburgo (Figure 1), which, together, occupy an area of 2500 km^2. The climate of the region is classified as Cwb (altitude subtropical climate), according to the Köppen classification, resulting in dry winters and rainy and fresh summers.
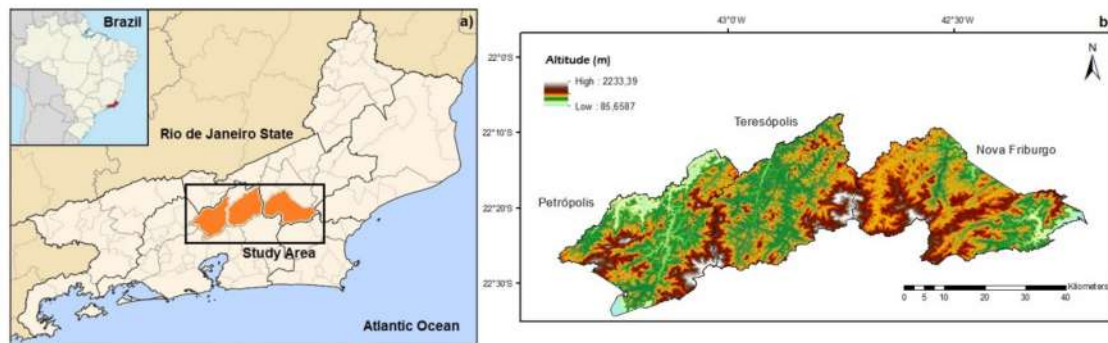
Figure 1. Study area location (**a**) and its elevation map (**b**).

The study area is located in the Serra do Mar complex, an important mountain chain that extends for approximately 1500 km along the southeastern coast of Brazil. The geology of the area is associated with a wide range of folds (Proterozoic age), composed predominantly of high-grade metamorphic rocks (gneisses) with well-defined foliation in the SW-NE direction and fractures in several orientations. Associations of syn-tectonic igneous rocks (granitoids) are also common [35]. Petrópolis and Teresópolis are located in a particular Serra da Mar portion, called the Serra dos Órgãos, which consists of a mountainous wall, raised by tectonics, which delimits the bottom of the Guanabara Bay basin. The unit is characterized by an abrupt orographic barrier with WSW-ENE direction. The highest portion of the Serra dos Órgãos, between Petrópolis and Teresópolis, is a plateau raised to over 2000 m of altitude, which contrasts with the extremely rugged relief of the surrounding mountain escarpments, characterized by very steep slopes and sharp tops. Between Teresópolis and Nova Friburgo, the Serra dos Órgãos maintains its imposing monolithic aspect, with altitudes between 1100 and 1300 m to the west, and 1400 to 2000 m to the east.

At lower altitudes, the relief is smoother, with intramontane hills characterized by convex-concave slopes and rounded tops, with colluvium and alluvium sedimentation at the footslope. In these areas, the regolith is composed by thick saprolitic and colluvial deposits that together can occasionally reach up to 50 m of thickness. The valley bottoms are narrow and develop along persistent tectonic fractures, in which only the larger rivers generate fluvial deposits. Near the valleys, there are escarpments with rocky outcrops and steep slopes, higher than 35 degrees. Talus and colluvium deposits with abundant rock blocks are usually present at the base [36].

The original vegetation was tropical forest, named "Mata Atlântica", but the human occupation and the need for areas available for planting and animal husbandry substantially reduced the natural vegetation. Therefore, areas with high slope, which, at first, were occupied by primary forests, became unstable after the removal of natural vegetation [37]. Also, in the 1960s, the urban area began to expand, occupying places with steep slopes, and consequently, houses were built in hazardous areas [38].

Mass movements of the study area are characterized mainly by rapid shallow translational landslides and debris flows [37]. Typically, the depth of the landslides is between 0.5 and 2.0 m, within the saprolites, in the slope colluvium or at the soil-rock interface [37]. In the study area, shallow translational slides and debris flows are triggered by rainfall. In some cases, they are combined together (e.g., complex landslides formed by translational slides evolving into debris flows) and are influenced by very similar controlling factors; indeed, in the landslide susceptibility literature, these kinds of phenomena are frequently analyzed together [10,39].

In January 2011, 11th and 12th, this region experienced one of Brazil's biggest disasters, as it rained about 320 mm in 48 hours, leading to hundreds of landslides [40]. This event caused 918 deaths (officially) and the 3 studied municipalities were the most affected. Currently, these municipalities are monitored by the National Center for Monitoring and Early Warning of Natural Disasters—Brazil (CEMADEN) 24 hours per day, and receive warnings for landslides and flooding. It is therefore

necessary to constantly update the mappings of the susceptible areas, in order to improve the alerts of natural disasters, and to direct attention to areas with the highest potential for landslides.

2.2. Random Forest

To generate the landslide susceptibility maps, in this work, the random forest model has been used. The random forest (RF) model is a nonparametric, multivariate, machine learning technique, which was proposed by Breiman [34,41] and first used in landslide susceptibility studies by Brenning [42]. Since then, it has rapidly gained widespread consolidation through many researches and case studies, as it is considered a relatively powerful approach in classification, regression and unsupervised learning [16,43–45]. RF contains a series of binary tree predictors, which are generated by using a random selection of the explanatory variables (or independent variables) to split each binary node (yes/no), and to perform a classification of the target dependent variable (in LSM studies, the presence or absence of landslides). Some of the observations are used for internal testing to evaluate the predictive capability of each predictor tree. This information is used to iterate the procedure hundreds of times by growing other random trees (hence the name "random forest"), and to iteratively adjust the prediction effectiveness. Once the best predictor tree is identified, it is applied to the whole study area, to define the LSM. RF has a great predictive performance, and runs fast by summarizing a large number of classification trees, especially when dealing with large amounts of data.

The random forest algorithm used in this work includes an internal test procedure, which can be used to assess the relative error that would be committed by the model, in case an independent variable is not used (out of bag error (OOBE)). This feature allows for ranking the variables according to their importance, and to implement a procedure of forward selection of parameters, in which the redundant, pejorative or uninfluential parameters can be automatically identified and discarded. Comparing to methods with single classifier combinations, the random forest has several advantages: both categorical and numerical variables can be used simultaneously; the algorithm is deemed suitable for analyzing nonlinear variables without considering the matter of multi-collinearity; it is quite robust with respect to overfitting problems; no particular statistical distribution of the data is required [19,34].

2.3. Model Configuration

The study area was divided into drainage basins, using the digital elevation model (DEM) from the Shuttle Radar Topography Mission (SRTM), modified by Valeriano [46], with 30 m spatial resolution, and the ArcHydro tool, present in the ArcGIS software, totaling 1253 basins, with an average area of 1.9 km^2. The choice of this spatial unit is mainly linked to the operational procedures in which the resulting map is planned to be used. As stated before, pixel is by far the most used spatial unit for landslide susceptibility analysis and mapping [24], and it has the advantage of a very fine spatial resolution. However, the pixel output is very complicated to understand and to exploit for untrained personnel, as it carries a very fine spatial detail that is difficult to properly address in the operative response. Basins are coarser spatial units, but they have the advantage of referring to a portion of the territory that is easy to understand and in which a unitary counteraction can be implemented in case of alarm.

The random forest model used as dependent variable the landslides and as independent variables a series of physical parameters (land use, soil type, lithology and slope aspect as categorical variables and slope gradient, elevation, slope curvature, profile curvature, planar curvature, flow accumulation, topographic wetness index, stream power index as numerical variables), as reported in Table 1. For the landslide data, the information used for the model was the presence or absence of landslides (binary classification between basins affected by landslides or not). Regarding the independent variables, different approaches were pursued to upscale the information to the basin scale. For categorical parameters, the most frequent class of each basin was considered as the most representative (majority approach); for numerical parameters, different variables were created considering the maximum value, the mean value and the standard deviation of the values encountered in each basin. For instance, the physical parameter "slope gradient" was used to derive three possible independent variables,

namely "mean slope" (the mean value characterizing each basin), "maximum slope" (the highest value encountered in each basin) and "standard deviation of slope" (the standard deviation of the values encountered in each basin).

Table 1. The parameters and derived independent variables used in the various configurations of the susceptibility model. Suffixes MAJ, MAX, MEAN and STD stand for majority (most represented class), maximum value, mean value and standard deviation, respectively.

	Parameters	Independent Variables	Classes/Range of Numerical Values
Categoric	Land Use	LULC_MAJ	Urban Area, Pasture, Eucalyptus, Secondary Forest, Primary Forest and Water
	Soil	Soil_MAJ	Argisol (Ultisol), Haplic Cambisol, Red-Yellow Latosol (Oxisol), Red Latosol (Oxisol), Red-Yellow Argisol (Ultisol), Litholic Neosol, Rock outcrops
	Lithology	Lito_MAJ	Granite, Migmatite, Gneiss, Gabbro, Orthogneiss, Granada-Biotite gnaisse, Biotite-gneiss, Metadiorite, Quartzite, Colluvial Sediment /Talus, Alluvial Sediment, Colluvium
	Aspect	Aspect_MAJ	Flat, North, Northeast, Northwest, East, Southeast, South, Southwest, West
Numeric	Slope	Slope_MAX; Slope_MEAN; Slope_STD	0 to 77°
	Elevation	Elevation_MAX; Elevation_MEAN; Elevation_STD	85 to 2233 m a.s.l.
	Curvature	Curv_MAX; Curv_MEAN; Curv_STD	−8 to 11.58
	Profile Curvature	Curv_Prof_MAX; Curv_Prof_MEAN; Curv_Prof_STD	−5.94 to 5.88
	Planar Curvature	Curv_Plan_MAX; Curv_Plan_MEAN; Curv_Plan_STD	−3.06 to 5.82
	Flow Accumulation	Flow_Acc_MAX; Flow_Acc_MEAN; Flow_Acc_STD	0 to 766 pixels
	Topographic Wetness Index (TWI)	TWI_MAX; TWI_MEAN; TWI_STD	5.4 to 29.27
	Stream Power Index (SPI)	SPI_MAX; SPI_MEAN; SPI_STD	0 to 184513

A crucial step in LSM analysis is the approach used to sample the variables to train and validate the model. As in any other statistical procedure, the size of the dataset influences the results, therefore the higher the number of samples to perform the statistical calibration/validation of the model, the more reliable the results obtained [19]. The same authors [19] demonstrated that a random sampling improves the predictive capability of the map and that the susceptibility model should also be trained/validated with respect to information about non-landslide locations, to avoid a generalized hazard overestimation. Regarding the proportion between the calibration and dataset samples, it is common practice to split them according to a 20/80 or 30/70 ratio [5,8,25,26,37,38]. As a consequence, in our application, all landslide basins were included in the analysis dataset. Afterwards, to get an unbiased susceptibility assessment, SPSS software was used to randomly select for the analysis also a number of non-landslide basins equal to that of landslide-affected basins. From the whole set of used basins (50% with landslides and 50% without landslides), 432 (70% of the dataset) basins were

randomly selected (with SPSS software) for model training, and 196 (30% of the dataset) basins were selected for validation.

To investigate the research questions posed in the introduction, four susceptibility assessments were performed separately, using four different configurations of input parameters. The difference in these configurations consists in the method to consider numerical variables in each unit basin: (i) MEAN VALUE, (ii) MAXIMUM VALUE, (iii) STANDARD DEVIATION, (iv) ALL the three above were used conjunctly. The motivations behind this series of tests is that all these methods have some points of strength: mean values would provide the best possible "description" of the average morphometric characteristics of each basin; maximum values would highlight the most extreme morphometric situations, which are often more linked to the initiation of landslides than averaged representations; standard deviation of morphometric attributes could be used as indicators of the variability of the landforms. The workflow of the research is summarized in Figure 2.
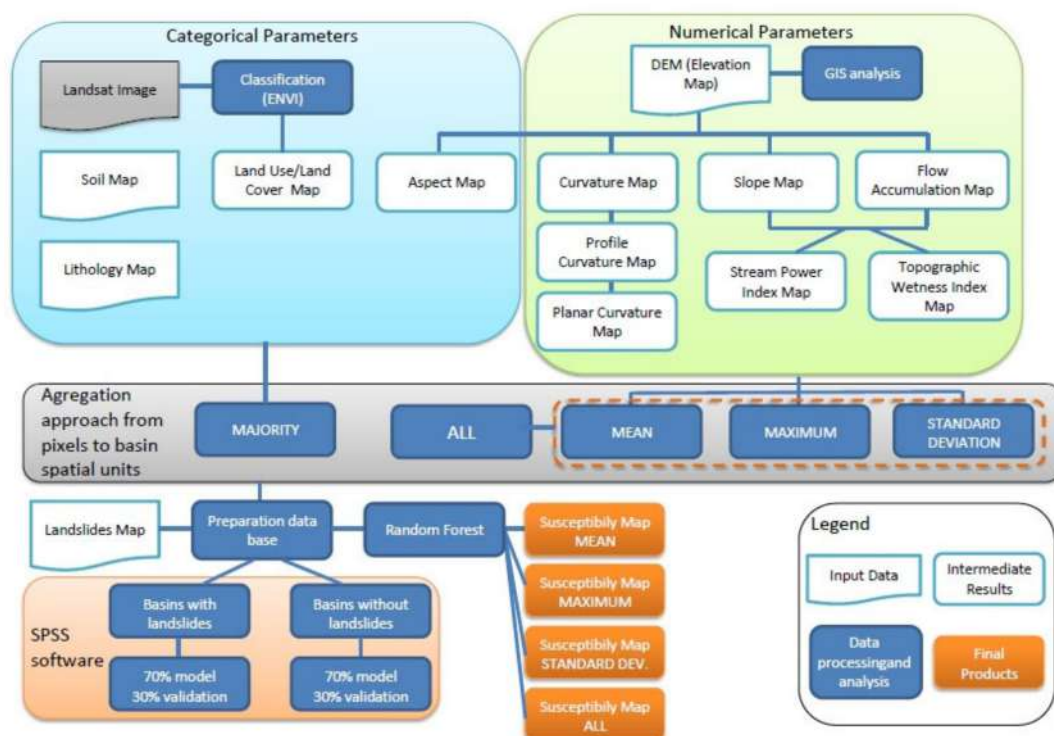

Figure 2. Workflow showing the main steps of the research.

2.4. Landslide Inventory

The availability of a complete and accurate landslide inventory is a fundamental requisite for landslide susceptibility assessments [47]. Traditionally, landslide inventories can be compiled using different sources of information including remote sensing, field surveys, internet, official report events [48–51]. There is no single database of natural disaster events in Brazil, so different data sources have been used, and the collected data have been organized and homogenized.

In this study, landslide data from 2008 to 2018 were obtained from scientific papers, event reports provided by CEMADEN, media, or mapped using GIS and satellite images. Overall, 3483 landslides were inventoried in a GIS-based geodatabase used for model calibration and validation: 2930 in Nova Friburgo, 348 in Petrópolis and 205 in Teresópolis. Data of the January 2011 landslide event for Nova Friburgo were provided by the municipal civil defense, which were mapped using satellite images available by the Google Earth in 01/19/2011 and visual interpretation (Figure 3). The scars were mapped along their entire length with polygon geometry. Landslide occurrences during the same event in Petrópolis and Teresópolis have been mapped by visual inspection of the Google Earth images

acquired in 04/11/2011 and 04/18/2011, and were mapped as points positioned where the slide started. Other events that occurred before and after January 2011 were obtained from CEMADEN, which uses data provided by the local civil defense and media, and also searches for landslides occurrences on the internet. In the aftermath of an event, the civil defense provides CEMADEN with a report containing the localization of the landslides. This information was used to include each landslide in the database as a point with a varying degree of precision (exact coordinates or approximate location). For Nova Friburgo, landslide occurrences found in the work of Oliveira [52] were also used, which were provided by the Nova Friburgo civil defense from November 2008 to December 2011.
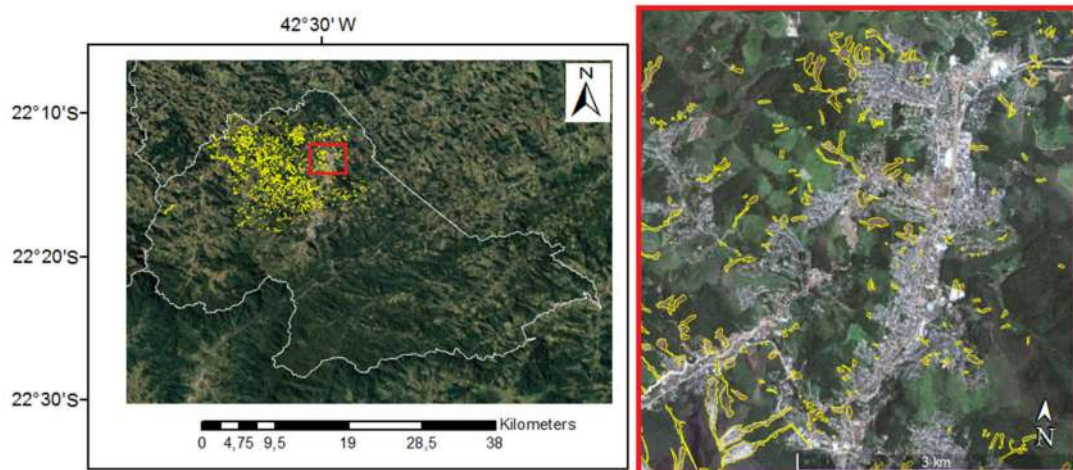
**Figure 3.** Nova friburgo landslide scars mapped on the google earth image in 19 January 2011.

These different sources provide very heterogenic information about landslides: sometimes they are mapped as polygons, sometimes as points. In addition, polygons may represent the only depletion zone or the whole landslide (from depletion to accumulation zone), while points sometimes are referred to the triggering area, and sometimes to the impact with infrastructures. Lastly, the spatial accuracy of the mapped elements is also not homogeneous: many landslides are mapped with high spatial accuracy, but in some occurrences, the spatial approximation is in the order of some meters. However, this level of uncertainty and inhomogeneity is not a problem if drainage basins are used as a spatial unit for analysis and for mapping, since, for each landslide, the spatial information needed is the basin in which it occurred, and there is no need to exactly pinpoint its location as in pixel-based LSM analyses. All the landslides were integrated in a geo-database in a GIS system, and drainage basins were classified as landslide-basins and no-landside-basins. To assess whether a basin is affected by landslides or not, the spatial detail of the elements of the geodatabase is sufficiently accurate and the mapping approach (whether the scars or the landslide bodies were mapped by polygons or points) is uninfluential, as, usually, each landslide evolves within the same basin.

2.5. Explanatory Variables

The physical parameters and the derived explanatory variables (aggregated at the basin scale) used in the susceptibility model are listed in Table 1. The aggregation technique varies, depending on the type of the variable. For the categorical parameters, the majority class present in the entire basin was considered. For the numerical parameters, the mean, the maximum value and the standard deviation were calculated on each basin, thus generating three different variables from each parameter. A brief overview of all the parameters is reported hereafter.

The land use and land cover (LULC) map is a categorical parameter, and was created using Landsat TM5 images, which have 30 m spatial resolution. Four images were used to cover the entire

study area: orbit 216, scenes 075 and 076 (09/04/2011) and orbit 217, scenes 075 and 076 (08/28/2011), obtained by the National Institute for Space Research (INPE), are shown in Figure 4.
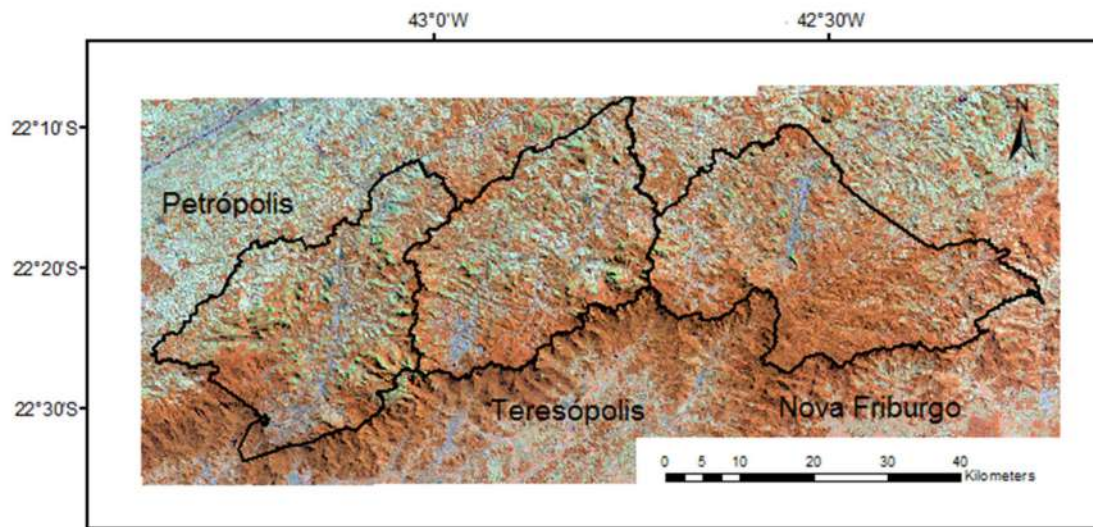
Figure 4. Landsat images used for the classification process (orbit 216, scenes 075 and 076 (09/04/2011) and orbit 217, scenes 075 and 076 (08/28/2011)). color composition rgb 543.

To define the LULC map, the year 2011 was selected, because most of the landslides occurred in this year. The method chosen to make the classification of land use was the neural network tool provided by the ENVI software. This technique replicates the neural structure of the human brain, and can be trained to recognize classification patterns in a complex scenario, such as a satellite image. Pixel samples were selected for the urban area, pasture, eucalyptus, secondary forest, primary forest and water classes (Figure 5). The omission and commission errors in the final map was corrected using raster editing and high-resolution spatial images available in Google Earth™. Land cover is important for studies of mass movements in the sense that the type or absence of vegetation cover directly influences the soil infiltration capacity [35,53,54]. It is evident that the degradation of vegetation cover coincides with widespread landslides of the slopes, but there are also cases of large landslides trigged by heavy rains in regions covered by forests. As the natural vegetation has been removed and replaced by other land covers, the weakening and destabilization of the slopes occurs [55]. Three different types of forest were mapped, because they have a very different structure, with different implications in slope stability. The primary forest is denser, and the root system is more compact and can better sustain the soil. The secondary forest is sparser, and is the forest in regeneration stage. The roots system is not well-formed and does not support the soil in the landslide events very well. The eucalyptus forest in Brazil is a non-native forest, and is planted for commercial purposes (cellulose, wood and other products). The large spacing between the trees can be a disadvantage for hillslope resistance to landslides.

The soil type parameterization was derived from the Soil Map provided by the Rio de Janeiro Environmental Secretariat, at the scale 1:100,000 (Figure A1). Different soil types have different physicochemical characteristics and, consequently, varying degrees of susceptibility in relation to landslides. The soil types more connected to shallow landslides are little cohesive, shallow, colluvial and porous [56]. Unstructured soils tend to be more common in steeper slopes or in talus deposits.

Lithological classes were derived from a lithology map provided by the Brazilian Geological Survey (CPRM), at the scale 1:100,000 (Figure A2). The most common lithology types present in the study area are shown in Table 1. Lithology is responsible for the composition, texture and degree of resistance of rocks and soils, as well as other factors, such as permeability and shear strength,

which directly influence the susceptibility to landslides [30,57]. Different rocks have different degrees of stability.
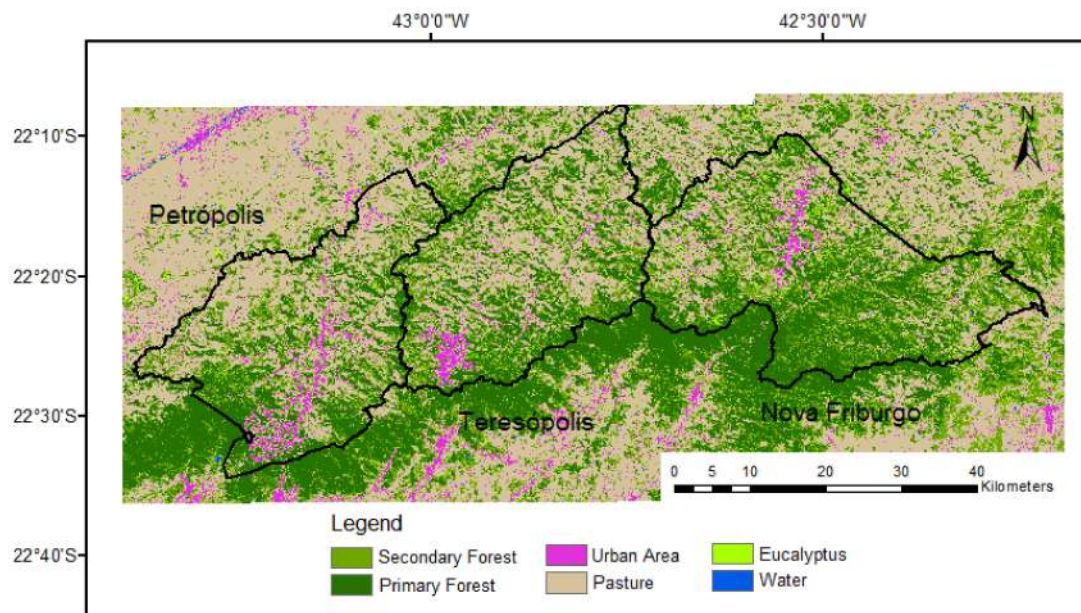
Figure 5. Land use and land cover map in the study area.

All the other parameters were derived from the digital elevation model (DEM) in a geographic information system software. The DEM was obtained by CPRM database and was generated from SRTM or Aster data, with spatial resolution of 30 m, and represents the elevation map (Figure 1). The aspect (Figure A3) indicates the slope orientation, and it was selected because in the study area the slopes facing south and southeast are those receiving more rain during the arrival of a cold front from the south, which is the condition that causes more landslides in the country.

The slope angle (Figure A4) is probably the most used parameter in landslide susceptibility studies [24]. In the study area, it varies from 0° to 77°, and it is a widely acknowledged as being the predisposing factor for debris flows and shallow landslides.

The curvature (Figure A5) is formed by the intersection of a random plane with the terrain surface [58], and controls the superficial and subsurface water regime of the slopes and, consequently, the predisposition to landslides.

The profile curvature (Figure A6) is the curvature in the vertical plane parallel to the slope direction. It is the measure of the rate of change of slope and may be planar, concave or convex. Negative values indicate concave portions of the terrain, while positive values indicate convex portions. Terrain with concave curvature are more prone to landslides [59]. A study conducted by Carvalho et al. [60] in a basin of Nova Friburgo municipality showed that the landslides occurred mainly in concave slopes

Planar curvature (Figure A7) is described as the curvature of a contour line formed by intersection of a horizontal plane with the surface [12], and is used to characterize areas of convergence and divergence of surface and subsurface water flow. It is related to the rate of variation of the slope along the contour lines and measures the propensity of the surface and sub-surface flow water to converge or diverge. Positive values correspond to divergent terrain, negative to convergent terrain and null to planar terrain. Convergent or V-shaped profiles are more conducive to trigger mass movements [61].

The flow accumulation (Figure A8) is derived from a flow direction map, and performs a cumulative count of the number of pixels that naturally drain into outlets and measures the area of a watershed that contributes runoff to each pixel. Flow accumulation highlights the sites that tend to concentrate water after precipitation, which are likely to have a high incidence of landslide triggers [19,62].

Flow accumulation and slope gradient were combined to derive two widely established hydrological indexes, namely the topographic wetness index (TWI)—Equation (1), developed from Beven and Kirkby [63], and the stream power index (SPI)—Equation (2), created by Bagnold [64], which are related to surface and sub-surface water and soil moisture, and thus involved in landslide triggering (Figures A9 and A10, respectively).

$$\text{TWI} = \log_e(\frac{A}{b}\tan\beta), \tag{1}$$

$$\text{SPI} = A\tan\beta/b, \tag{2}$$

where A = upstream catchment area (m^2) or flow accumulation, b is the width of a cell through which water flows (m) and β is the slope angle (radian). High values of TWI are associated with flat sites and valley bottoms, with a large contributing area. These areas are usually related to the deposit zones, with material coming from past landslides. The places where landslides started normally have low TWI values, since they are concentrated in areas with steep slope, also associated with the presence of a thin layer of soil [65]. The SPI is derived from slope and hydrological contributing area. The SPI is higher near the streamlines, and the maximum value occurs where the contribution area is larger, and the morphology is flatter. Then, it is expected that higher SPI values indicate places where runoff is high, which makes it difficult to establish well developed and thick soil layers [65].

3. Results

The landslide susceptibility indices calculated in each basin were imported in ArcGIS software and used to draw the reclassified susceptibility maps (Figure 6). The reclassification is based on equal intervals (0–25 low; 25–50 moderate; 50–75 high; 75–100 very high), because this classification scheme is independent from the statistical distribution of the mapped values and, thus, the same class-break values can be used for every map. As a consequence, the differences encountered depend only on the model results, and are not influenced by the classification approach, allowing for a more direct and straightforward comparison among the results of the four models. All the model configurations identify some large susceptibility hotspots around Nova Friburgo and Petropolis, while in many basins scattered away from the most intensely landslide affected areas, the spatial prediction of susceptibility is modelled with different outcomes. Figure 7 can be used to get an idea of the discrepancies between the resulting maps, as it represents the difference between the maximum and the minimum susceptibility values calculated in each basin by the four models. As can be seen, relevant differences can be encountered in some basins. This outcome highlights the sensitivity of the LSM results to the approach used to transform numerical parameters into basin-level variables. The subsequent step of the research consists of assessing which one of the resulting maps is the most reliable.

According to a consolidated literature [24,66], the area under the receiver operating characteristic (ROC) curve (AUC) was used to validate and evaluate the accuracy of each model's results. The ROC curve is a cutoff-independent evaluation model, and the AUC value indicates the performance of the corresponding model configuration, as shown in Figure 8.

The validation statistics show that the most accurate spatial prediction is obtained using the mean value of all numerical variables to characterize each basin (AUC = 74.3%). However, the fourth configuration, which uses all the parameters (maximum values, minimum values and standard deviation), shows a comparable accuracy (AUC = 73.7%). Both outcomes can be considered indexes of a "good" prediction accuracy, according to the scale proposed by Arabameri et al. [67].

As a further test of the modeling results, Table 2 provides a summary of the subdivision of the basins into the four susceptibility classes. These statistics could be used to interpret the meaning of the AUC values derived by the validation of the LSMs. In all four applications, higher susceptibility classes are a lower number, and most of the landslides concentrate there. This is a reasonable outcome for LSM, but some differences can be identified among the four approaches. The LSM derived by the

MEAN approach provides the highest AUC, because it is the one with the most balanced classification: it has one of the highest numbers of high and very high susceptible basins, similar to MAXIMUM, but the percentage of correctly classified basin is higher. The use of the MAXIMUM approach leads to an overestimation of hazard (which is reflected by a lower AUC value), as the derived LSM is the one with the highest number of very high susceptible basins and a relevant number of highly susceptible ones, but the number of basins that have actually been affected by landslide are the lowest percentage of all the four approaches. On the contrary, STANDARD DEVIATION underestimate hazard, as the number of basins mapped as highly susceptible is lower than all other approaches.
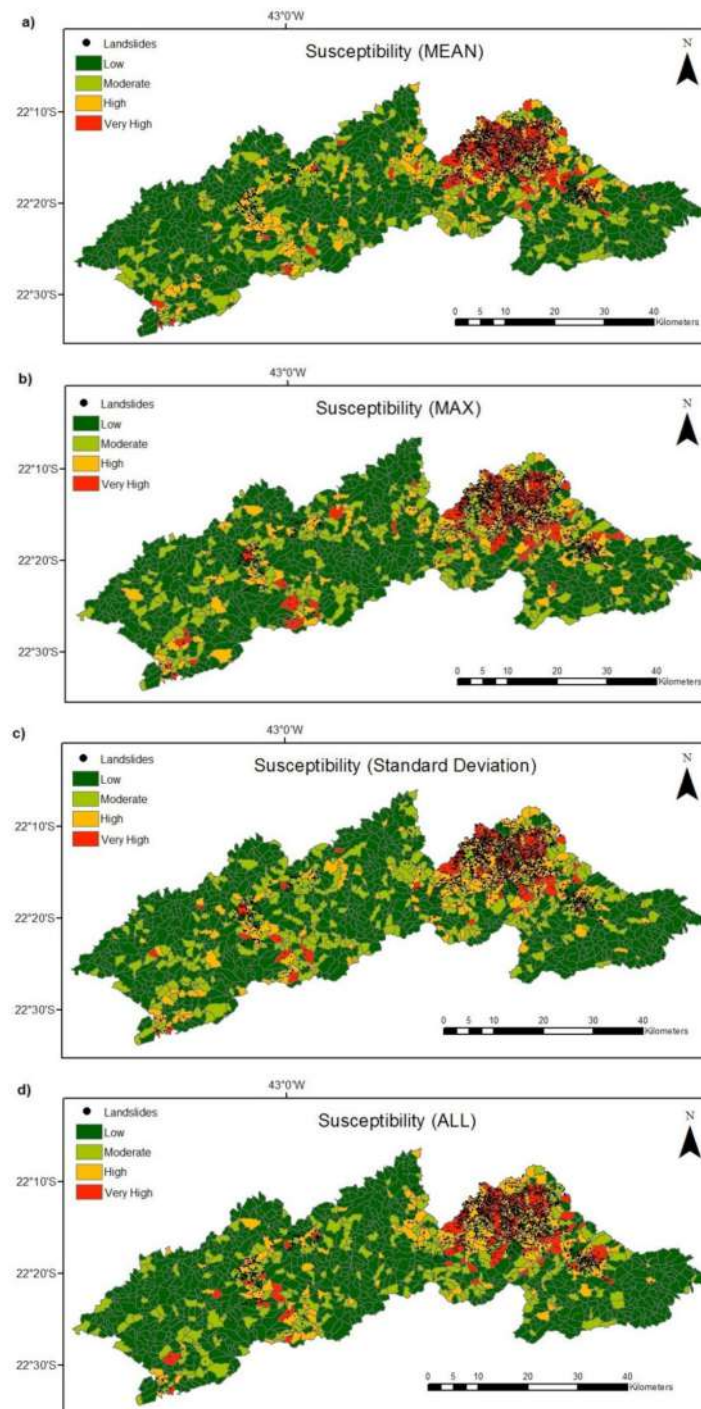
Figure 6. Susceptibility maps generated with the four different model configurations: (**a**) mean, (**b**) maximum, (**c**) standard deviation and (**d**) all.
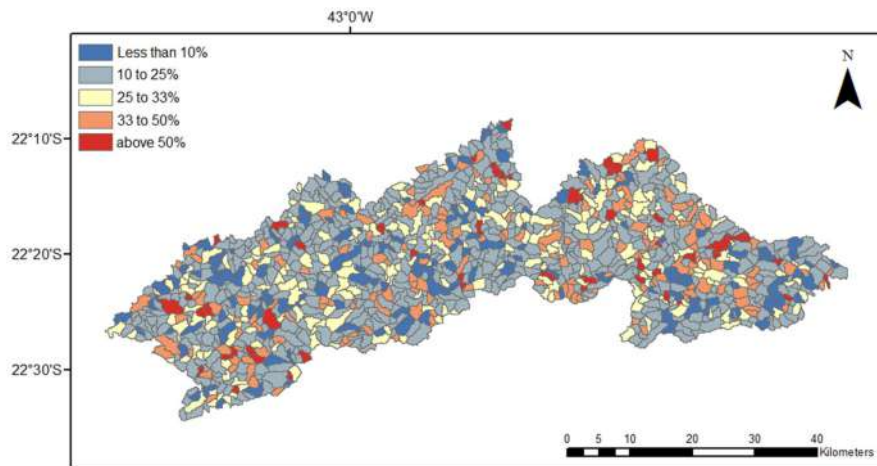
Figure 7. Difference among the maximum and minimum susceptibility values mapped by the four model configurations.
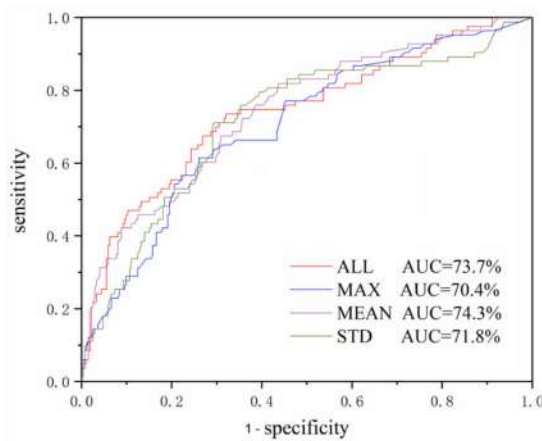

Figure 8. Receiver operating characteristic (roc) curves of the four model configurations.

Table 2. Summary statistics of the subdivision of the basins into susceptibility classes, according to the different approaches pursued in the study.

Approaches	Susceptibility Class	Number of Basins	Percentage of Basins	Number of Landslide Basins	Percentage of Basins Affected by Landslides
Susceptibility ALL	Low	794	63.36%	49	6.17%
	Moderate	246	19.63%	83	33.73%
	High	141	11.25%	113	80.14%
	Very High	72	5.74%	67	93.05%
Susceptibility MEAN	Low	747	59.61%	49	6.55%
	Moderate	287	22.90%	90	31.35%
	High	140	11.17%	102	72.85%
	Very High	79	6.30%	71	89.87%
Susceptibility MAXIMUM	Low	775	61.85%	55	7.09%
	Moderate	250	19.95%	83	33.20%
	High	148	11.81%	103	69.59%
	Very High	80	6.38%	71	88.75%
Susceptibility STANDARD DEVIATION	Low	757	60.41%	50	6.60%
	Moderate	278	22.18%	94	33.81%
	High	156	12.45%	109	69.87%
	Very High	62	4.94%	59	95.16%

Another issue worth discussing is the importance of the explanatory variables used in the modeling. The relative importance of each variable can be quantitatively assessed with the "out of bag error" [19,68,69], an index measuring the relative error that would be committed by the model in case an independent variable is not used. Figure 9 shows the importance of the parameters used in the best performing configuration (based on mean values (Figure 9a)) and in the fourth configurations (all parameters (Figure 9b)). The latter can be used to get an overview on how the relative importance of a parameter can increase or decrease if it is aggregated at the basin level considering the mean value, the maximum value or the standard deviation. Indeed, in the first ranks, the parameters of various typology are present: some based on mean values (e.g., mean curvature), some based on maximum values (e.g., maximum slope), some based on standard deviation (e.g., standard deviation of curvature) and some categorical parameters (e.g., soil type).
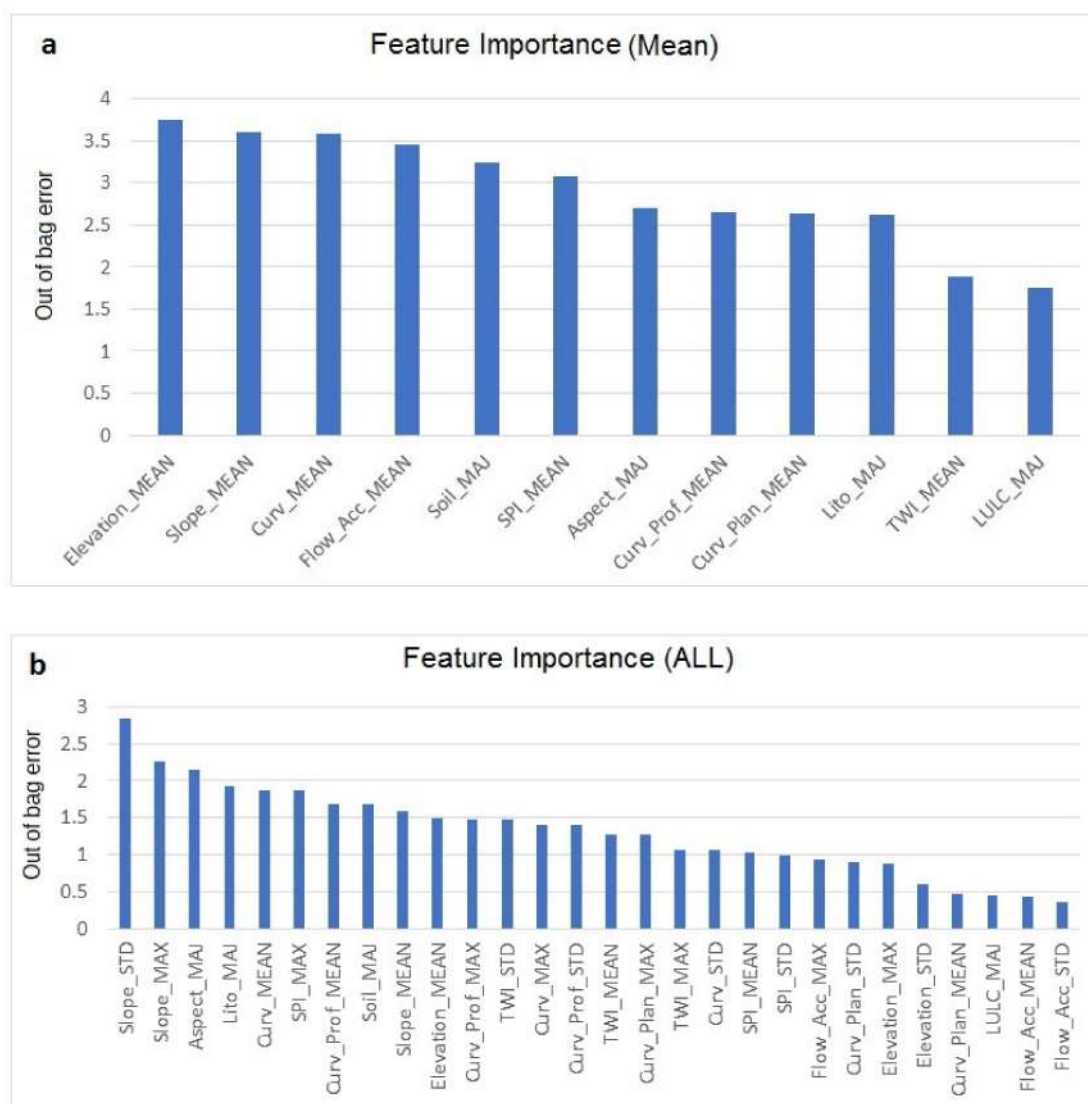
Figure 9. Ranking of the parameters used in the best performing configurations, according to their importance estimated by the out of bag error (**a**): using mean configuration and (**b**): using all parameters together). see Table 1 for details on the variables used.

4. Discussion

A closer investigation of the results obtained can help getting some insights on the use of numerical variables with spatial units larger than the cell size of the raster DEM used to derive them. Indeed,

the results showed that the method used to aggregate the values of the parameters over the spatial units of analysis had a noticeable impact on the results, both in terms of overall predictive capability (AUC values) and in terms of spatial distribution of the susceptibility values (maps in Figure 6).

Using the mean value to characterize the whole basin is the strategy that produced the most accurate predictions in terms of AUC (74.3%, as shown in Figure 8). This means that when the environmental information needs to be upscaled from the pixels to relatively larger spatial units like basins, using the averaged values of the hydrologic and morphologic parameters derived from raster DEMs seems to better reflect the average spatial probability of landslide occurrence. In contrast, the model configuration using maximum values to characterize the whole unit basins is the one that returns the lowest AUC values (70.4%).

Indeed, it should be stressed that the approach of using hydrographic basins as a spatial unit is based on the characterization of each basin with a relative and averaged spatial probability of being interested by landslides. Consequently, there is not a full correspondence with the physical mechanism of shallow landslide triggering, for which extreme values are more significant as predisposing factors (for instance, other factors being equal, a shallow landslide is expected more likely when extreme values of morphological parameters are present). In addition, the use of the maximum value alone may not be entirely representative of the geomorphological system of the hillslopes at catchment scale, and could provide biased predictions, in which the hazard is overestimated, hence, the largest number of basins mapped in the highest susceptibility classes (Table 2) and the lowest AUC value obtained after the validation.

Using only the standard deviation derived variables, the overall results are intermediate (AUC = 71.8%). This outcome was quite surprising, as the standard deviation of geomorphic parameters is rarely used in LSM studies requiring spatial aggregations; however, some authors used standard deviation as a proxy for the variability of the conditions encountered in a large spatial unit [19,70,71], and similar indexes may be reasonable indicators of predisposing conditions.

Beyond the generalizations inferred by the comparison of AUC values, it should be stressed that the above considerations depend largely on the morphometric parameters considered. For some indexes and attributes with a hydrologic meaning (SPI, flow accumulation and plan curvature) it can be seen from Figure 9b that the parameter derived from the maximum value has a higher importance in the modelling than the homologue parameters derived from the mean or the standard deviation. For elevation, curvature and profile curvature, mean values seems the best method of aggregation (Figure 9a), probably because these parameters are more suited to account for the general characteristics of each basin in terms of morphology or climatic setting; for instance, elevation is clearly related to the mean rainfall amounts falling on the test area and, thus, its average value appears to be a good indicator in the susceptibility analysis. According to the ranking of the importance of the single parameters, slope and TWI are the only factors for which the standard deviation-derived parameter has the highest statistical importance. This outcome can be interpreted as observing that the standard deviation of TWI and slope could be symptomatic of abrupt changes in the hillslope morphology and sub-surface flow. It is well known that the mechanism of triggering in some circumstances can be favored neither by exceptionally high nor by mean values, but by abrupt changes in the hillslope system, such as a marked contrast in geotechnical properties, or sudden changes in surface or subsurface morphologies that reduce hydraulic transmissivity, thus fostering return flows or overland flows that may evolve in soil slips or mobilization of debris. In the case of many other factors, however, the standard deviation is not the most important derivate parameter, but it can still be conveniently used (thus, the second highest performance of the "all-parameters" configuration in terms of AUC) as an index of how much the morphology is varied in each basin, thus giving an approximate idea of the terrain roughness and energy of relief. Indeed, the standard deviation of several morphometric attributes has been already used in landslide susceptibility assessment with this connotation [19,72].

It is not surprising, then, that a good spatial prediction of landslide susceptibility is obtained with the configuration, encompassing all the factors derived from the maximum, mean and standard

deviation (AUC 73.7%); this strategy exploits each approach used to aggregate pixel values at the basin level, and can also be considered a further confirmation of the flexibility and predictive power of random forest, which can handle very complex spatial information, even when provided by a large number of input variables affected by mutual relationships. Among the 28 used parameters, some resulted to have little importance; however, their contribution was still evaluated positively by the internal tests of the RF algorithm (Figure 9b). To further corroborate this assumption, some model configurations encompassing a reduced set of parameters have been tested and validated: the validation statistics showed that a more parsimonious parameterization does not improve the model performance: the AUC value is reduced to 70.3% if the least important explanatory variable (Flow_Acc_STD) is excluded from the analysis, and to 68.8% if the second least important (Flow_Acc_MEAN) is also excluded. Similarly, a configuration in which the redundancy and the parameterization is reduced to the minimum was tested: for every parameter only the variable that showed the highest OOBE value was selected. This configuration therefore uses each parameter only once, and combines the variables aggregated using the mean value (elevation, curvature and profile curvature), some using the maximum (planar curvature, SPI and flow accumulation) and some using the standard deviation (slope and TWI) and gets lower AUC scores (0.700) than the four main configurations tested. However, the AUC value reported by the "ALL" configuration is lower than the "MEAN" configuration, suggesting that the random forest algorithm is not completely capable of resolving the problems of collinearity between such a high number of variables.

Another important input for discussion is the quality of the results obtained. The validation of "ALL" and "MEAN" configurations returned the AUC scores 74.3% and 73.7%, respectively. These values fall in the 70–80% range, which is considered "good", according to the classification proposed by some authors [67]; however, it is lower than validation values commonly reported in the international literature. This indicates the limits of the proposed approach, in which the predisposing factors at the triggering point of each landslide cannot be ascertained and included in the analysis, because only the general characteristics of each basin is taken into account. This approximation has been discussed before, here, it is worth stressing that some predisposing factors dramatically lose their significance if considered according to this approach: for instance, the flow accumulation has a very low predictive importance for every aggregation method (as reported in Figure 9b, Flow_Acc_MEAN and Flow_Acc_STD are the least important variables, and even Flow_Acc_MAX has a very low relative importance).

In addition, another reason for the relatively low AUC scores is that the use of basins as spatial computation unit reduces the training and validation sample. In the study area the landslide inventory contains a high number of landslides (namely, 3483). In a pixel-based approach, that would allow to identify at least 3483 points to sample predisposing conditions (of course, only if all the landslide locations were known with good accuracy). In the proposed approach, only 1253 basins were defined, and only 312 of them included landslides; consequently, the training and validation samples are smaller. It has not been excluded that in the future, with a larger and updated dataset, the results could be improved. However, it should be noted that a pixel-based approach could not be pursued in this test site, due to limitations in the landslide dataset: even small uncertainties in the locations and geometry of the landslides prevent all of them from being related exactly to a pixel, and therefore a pixel-based analysis would be impossible.

Another main drawback of the proposed approach is the coarse spatial resolution represented by the hydrographic basins. However, this choice was made for the practical reasons of integration in the framework of existing and planned hazard management procedures, since it is a good practice to consider also the needs of decision makers in the planning of a research work [73]. Raster cells would allow for a finer spatial resolution, but cannot be used as a spatial unit for warnings and operative response; thus, a spatial reaggregation in larger spatial units is needed. Basins would be easier to understand for populations, and untrained personnel and could be conveniently exploited in decision making and in operational risk management procedures.

Indeed, the strict collaboration between CEMADEN and local civil defense authorities was useful to direct the research towards practical and operative solutions. As a result, the final LSM will be used as an operational tool for hazard management. Regarding risk prevention activities, the identification of basins with higher landslide susceptibility will be useful in terms of setting the priorities of intervention for risk mitigation strategies. Use of the LSM is also planned during the prediction and hazard response activities to be undertaken in the situation room: joint use with weather monitoring and forecasting products will allow for a better constraint of the spatial extent of the incoming hazardous events. However, regarding this operational use, it should be stressed that, like any other LSM, the resulting map is a static instrument that provides a spatial assessment of the relative spatial probability of landslide occurrence, based on the analysis of predisposing factors. Therefore, triggering factors (e.g., rainfall intensity) are not directly accounted for, and no temporal information is provided. To overcome this drawback common to all LSMs, in the future, progress of the research and integration of the susceptibility map with a warning system based on rainfall thresholds [40] is planned, using the recently proposed combination techniques [73–75].

Moreover, the option of basin-scale aggregation is the only option that makes it possible to handle a multi-source landslide inventory with inhomogeneous characteristics (e.g., different landslide geometries and mapping approaches). If all landslides are associated with the hydrographic basin spatial unit, it makes no differences if landslides are mapped as points or polygons, and if the element mapped is the scar, the crown or the whole accumulation body, as long as, of course, each landslide begins, evolves and ends within the same catchment. In addition, small uncertainties in the spatial location of the mapped elements also become irrelevant, or at least would have a much lower influence than in a pixel-based approach.

5. Conclusions

In this study, the random forest technique was applied to map landslide susceptibility in a mountain region north of Rio de Janeiro, Brazil, where hundreds of landslides occurred in 2011, and impacted the environment and the population, causing hundred deaths and economics losses.

The study area was divided into drainage basins to be used as meso-scale spatial unit for the analysis and the resulting map. This approach was chosen for two main reasons. First, the choice of this spatial unit is mainly linked to the operational hazard management procedures in which the resulting map is planned to be used: basins are coarse spatial units, but they have the advantage of referring to a portion of the territory that is easy to understand, and in which a unitary counteraction can be implemented by Civil Defense organizations in case of alarm. Second, the hydrographic basin approach makes it possible to overcome the problems of spatial accuracy and inhomogeneity encountered in the multi-source landslide inventory used; such limitations would prevent a pixel-based analysis, because the uncertainties in locations and geometry of the landslides would make it impossible to exactly relate each landslide to a pixel.

Twenty-eight variables, derived from four categorical and eight numerical parameters (described in Table 1), have been used for the model construction. To aggregate their values from the pixel to the basin level, different approaches were pursued. For categorical parameters, the most representative (majority) class was assigned to each basin, and for the numeric parameters, three different aggregation schemes were implemented, based on the maximum value, mean value and the standard deviation. Four susceptibility maps were generated varying the approach used to characterize the numerical morphometric parameters: maximum values, mean values, standard deviation and all values together. The best results were obtained using mean values (AUC 74.3%) and all values together (AUC 73.7%), while the standard deviation and maximum values produced lower validation scores (AUC was 71.8% and 70.4%, respectively).

The LSM obtained with this work will be used by CEMADEN and local civil defense authorities as an operational tool for hazard management. The identification of basins with higher landslide susceptibility will be useful to set priorities of intervention for risk mitigation strategies. Moreover,

joint use in the situation room with weather monitoring and forecasting products will allow for a better constraint of the spatial extent of the incoming hazardous events.

In addition to producing a meso-scale susceptibility map for the study area, the present work made it possible to draw some conclusions concerning a common source of uncertainty that affects LSM based on spatial units larger than the pixel size of the input parameters: that is, how to generalize the fine-level pixel information at the level of a larger spatial unit like basins:

- In the analyzed case of study, with peculiar physiographic characteristics and a limited number of training data, the best generalization approach was the use of the mean value of each numerical parameter to derive the explanatory variable to be used in the susceptibility analysis.

- Nevertheless, investigating the relative importance of each variable, it was discovered that each morphometric factor could be more conveniently exploited using a different generalization scheme, related to the different physical meaning of each of them.

- For instance, in the proposed application, most of the hydrologic factors are better represented by the maximum value, elevation and curvature by the mean value, slope and TWI by standard deviation.

- Given that, if the susceptibility assessment is based on a model that can handle multicollinearity, large number of parameters and complex configurations, another option worth testing could be using all the parameters derived with the three different approaches.

Author Contributions: V.C. and X.T. organized and analyzed the data and wrote the paper, S.S. was the responsible of the project, developed the conceptual scheme of the research and wrote the paper, A.R. wrote and reviewed the paper, T.N. provided landslide data, F.C. and N.C. supervised the work. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A
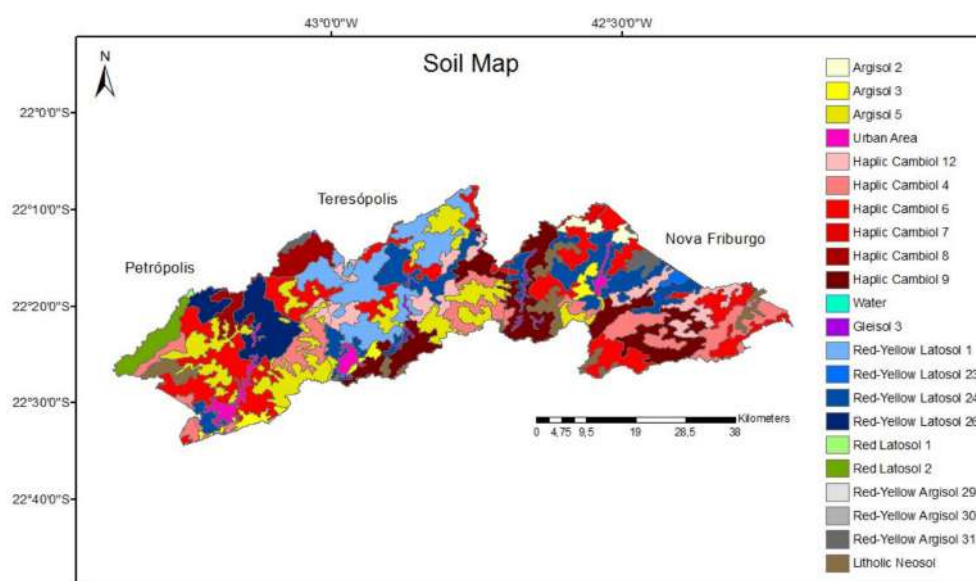
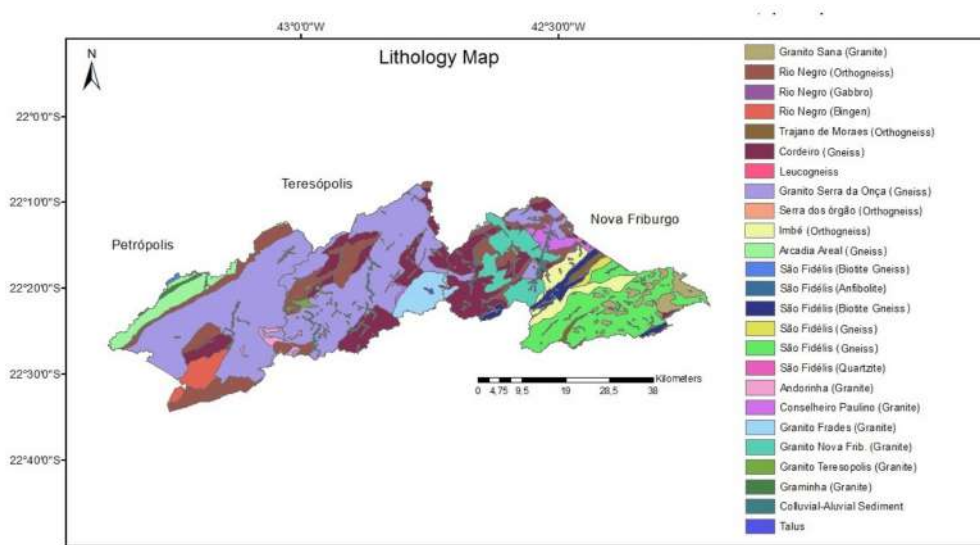**Figure A1.** Soil Map in the study area. Scale 1:100.000. Source the Rio de Janeiro Environmental Secretariat.

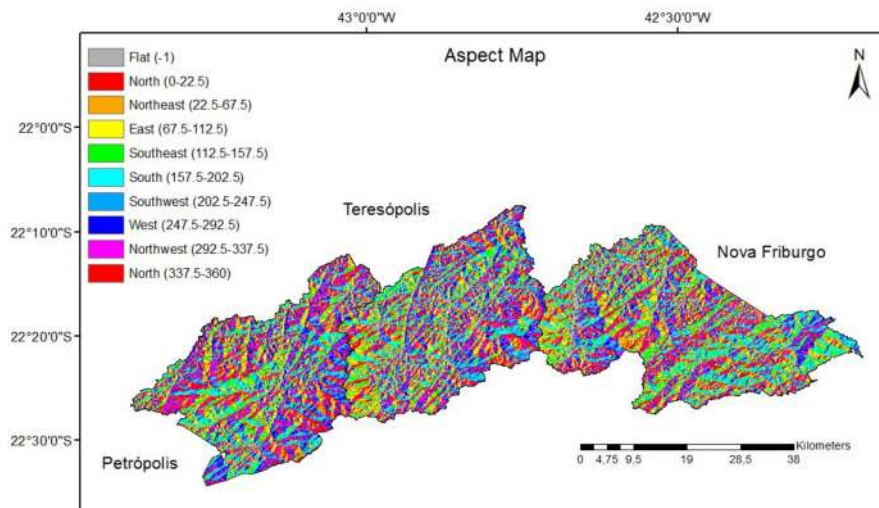**Figure A2.** Lithology Map. Scale 1:100.000. Source: CPRM—Brazilian Geological Survey.
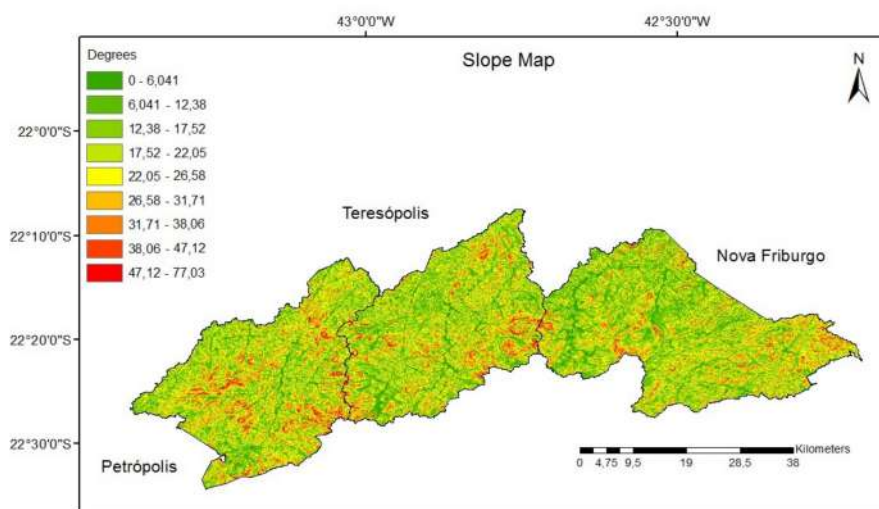

Figure A3. Aspect Map.
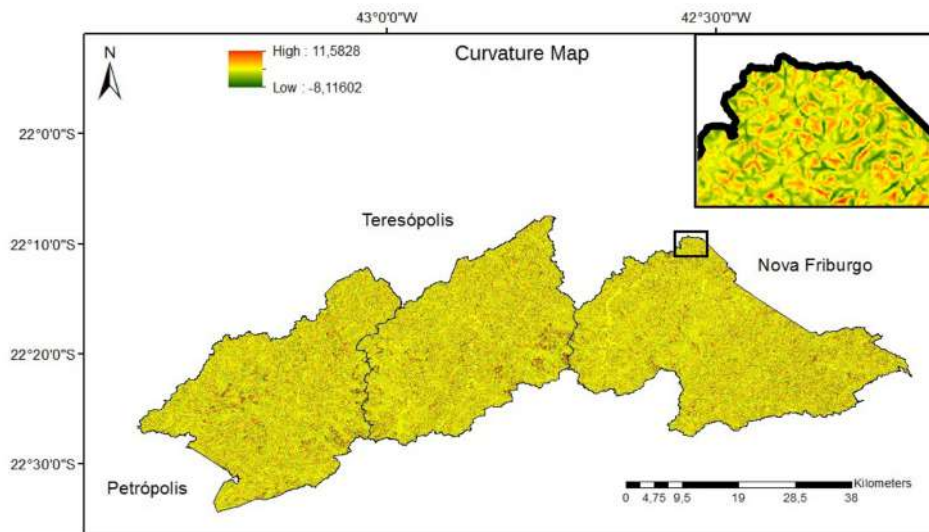

Figure A4. Slope Map.
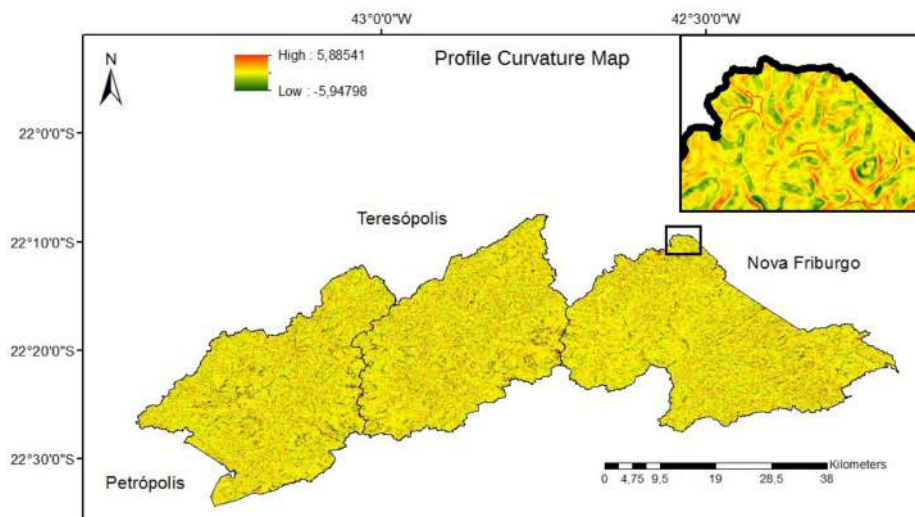
Figure A5. Curvature Map.
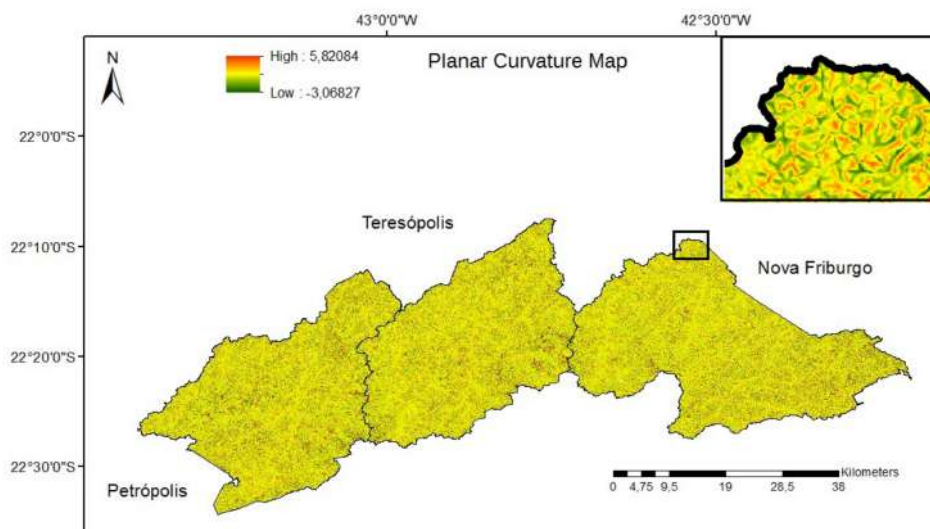

Figure A6. Profile Curvature Map.

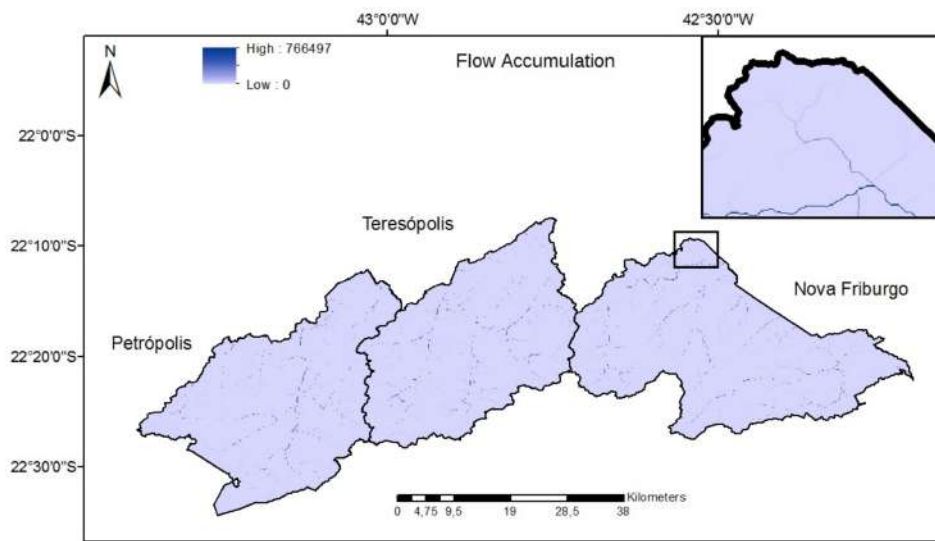
Figure A7. Planar Curvature Map.
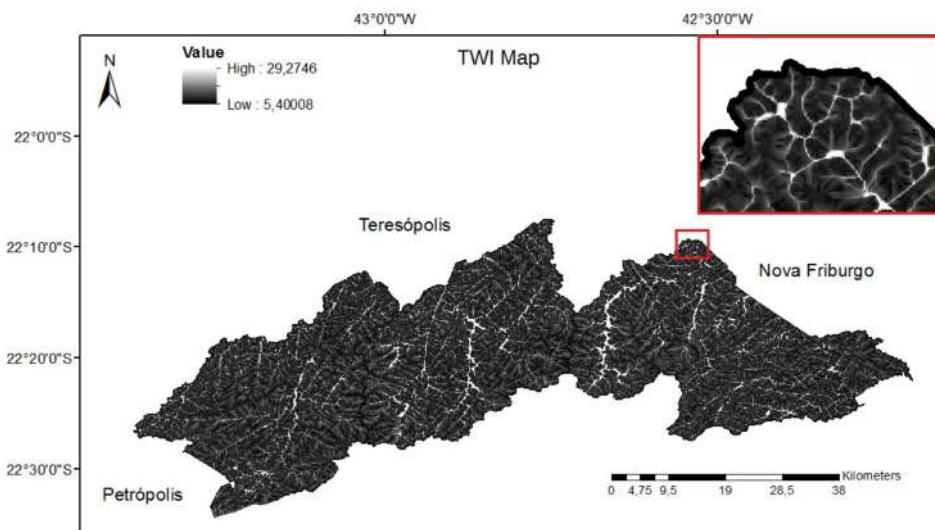
Figure A8. Flow Accumulation Map.
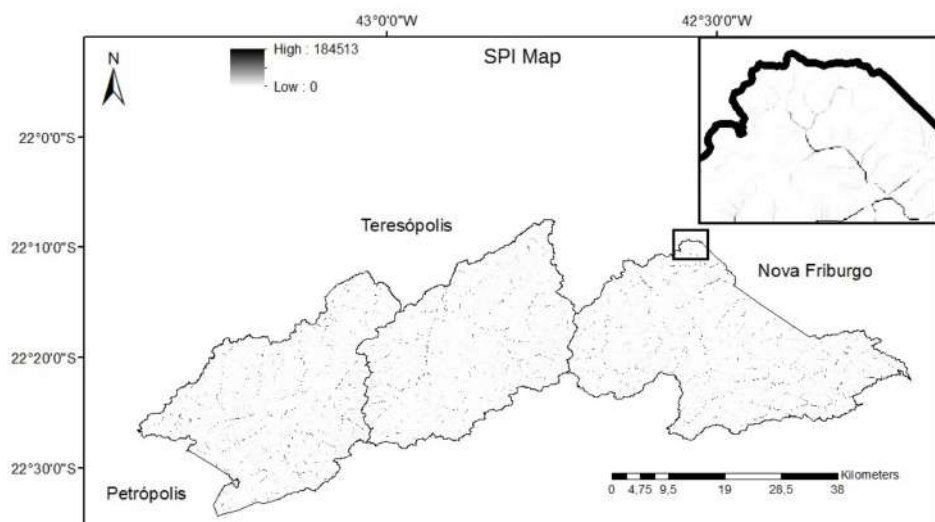

Figure A9. TWI Map.

Figure A10. SPI Map.

References

1. Froude, M.J.; Petley, D. Global fatal landslide occurrence from 2004 to 2016. *Nat. Hazards Earth Syst. Sci.* **2018**, *18*, 2161–2181. [CrossRef]
2. Guzzetti, F.; Reichenbach, P.; Cardinali, M.; Galli, M.; Ardizzone, F. Probabilistic landslide hazard assessment at the basin scale. *Geomorphology* **2005**, *72*, 272–299. [CrossRef]
3. Van Westen, C.J.; Terlien, M.T.J. An approach towards deterministic landslide hazard analysis in GIS. A case study from Manizales (Colombia). *Earth Surf. Process. Landforms* **1996**, *21*, 853–868. [CrossRef]
4. Van Westen, C.J.; Castellanos, E.; Kuriakose, S.L. Spatial data for landslide susceptibility, hazard, and vulnerability assessment: An overview. *Eng. Geol.* **2008**, *102*, 112–131. [CrossRef]
5. Tien Bui, D.; Tuan, T.; Klempe, H.; Pradhan, B.; Revhaug, I. Spatial prediction models for shallow landslide hazards: A comparative assessment of the efficacy of support vector machines, artificial neural networks, kernel logistic regression, and logistic model tree. *Landslides* **2016**, *13*, 361–378. [CrossRef]
6. Varnes, D.J.; IAEG. Commission on landslide and other mass movement on slopes. In *Landslide Hazard Zonation: A Review of Principles and Practice*; UNESCO: Paris, France, 1984; p. 63.
7. Hutchinson, J.N. Landslide Hazard Assessment. In *Landslides, Proceeding of the 6th International Symposium on Landslides, Christchurch, New Zeland, 10–14 February 1992*; Bell, D.H., Ed.; Balkema: Amsterdam, The Netherlands, 1995; Volume 1, pp. 1805–1841.
8. Yilmaz, I. Landslide susceptibility mapping using frequency ratio, logistic regression, artificial neural networks and their comparison: A case study from Kat landslides (Tokat—Turkey). *Comput. Geosci.* **2009**, *35*, 1125–1138. [CrossRef]
9. Carrara, A. Multivariate models for landslide hazard evaluation. *J. Int. Assoc. Math. Geol.* **1983**, *15*, 403–426. [CrossRef]
10. Trigila, A.; Frattini, P.; Casagli, N.; Catani, F.; Crosta, G.; Es-positon, C.; Iadanza, C.; Lagomarsino, D.; Scarascia, M.G.; Segoni, S.; et al. *Landslide Science and Practice*; Margottini, C., Canuti, P., Sassa, K., Eds.; Springer: Berlin, Germany, 2013; pp. 287–295.
11. Lee, S. Application of logistic regression model and its validation for landslide susceptibility mapping using GIS and remote sensing data. *Int. J. Remote Sens.* **2005**, *26*, 1477–1491. [CrossRef]
12. Duman, T.Y.; Can, T.; Gokceoglu, C.; Sonmez, H. Application of logistic regression for landslide susceptibility zoning of Cekmece Area, Istanbul, Turkey. *Environ. Geol.* **2006**, *51*, 241–256. [CrossRef]
13. Manzo, G.; Tofani, V.; Segoni, S.; Battistini, A.; Catani, F. GIS techniques for regional-scale landslide susceptibility assessment: The Sicily (Italy) case study. *Int. J. Geogr. Inf. Sci.* **2013**, *27*, 1433–1452. [CrossRef]
14. Ermini, L.; Catani, F.; Casagli, N. Artificial Neural Networks applied to landslide susceptibility assessment. *Geomorphology* **2005**, *66*, 327–343. [CrossRef]
15. Marjanović, M.; Kovačević, M.; Bajat, B.; Voženílek, V. Landslide susceptibility assessment using SVM machine learning algorithm. *Eng. Geol.* **2011**, *123*, 225–234. [CrossRef]
16. Xiao, T.; Segoni, S.; Chen, L.; Yin, K.; Casagli, N. A step beyond landslide susceptibility maps: A simple method to investigate and explain the different outcomes obtained by different approaches. *Landslides* **2020**, *17*, 627–640. [CrossRef]
17. Akgun, A. A comparison of landslide susceptibility maps produced by logistic regression, multi-criteria decision, and likelihood ratio methods: A case study at İzmir, Turkey. *Landslides* **2012**, *9*, 93–106. [CrossRef]
18. Greco, R.; Sorriso-Valvo, M. Influence of management of variables, sampling zones and land units on LR analysis for landslide spatial prevision. *Nat. Hazards Earth Syst. Sci.* **2013**, *9*, 2209–2221. [CrossRef]
19. Catani, F.; Lagomarsino, D.; Segoni, S.; Tofani, V. Landslide susceptibility estimation by random forests technique: Sensitivity and scaling issues. *Nat. Hazards Earth Syst. Sci.* **2013**, *13*, 2815–2831. [CrossRef]
20. Liu, L.; Li, S.; Li, X.; Jiang, Y.; Wei, W.; Wang, Z.; Bai, Y. An integrated approach forlandslide susceptibility mapping by considering spatial correlation and fractal distributionof clustered landslide data. *Landslides* **2019**, *16*, 715–728. [CrossRef]
21. Carrara, A.; Cardinali, M.; Detti, R.; Pasqui, V.; Reichenbach, P. GIS Techniques and Statistical-Models in Evaluating Landslide Hazard. *Earth Surf. Process. Landf.* **1991**, *16*, 427–445. [CrossRef]
22. Carrara, A.; Guzzetti, F. Use of GIS technology in the prediction and monitoring of landslide hazard. *Nat. Hazards* **1999**, *20*, 117–135. [CrossRef]
23. Magliulo, P.; Di Lisio, A.; Russo, F.; Zelano, A. Geomorphology and landslide susceptibility assessment using GIS and bivariate statistics: A case study in southern Italy. *Nat. Hazards* **2008**, *47*, 411–435. [CrossRef]

24. Reichenbach, P.; Rossi, M.; Malamud, B.D.; Mihir, M.; Guzzetti, F. A review of statistically-based landslide susceptibility models. *Earth Sci. Rev.* **2018**, *180*, 60–91. [CrossRef]

25. Luzi, L.; Pergalani, F.; Terlien, M.T.J. Slope vulnerability to earthquakes at subregional scale, using probabilistic techniques and geographic information systems. *Eng. Geol.* **2000**, *58*, 313–336. [CrossRef]

26. Carrara, A.; Crosta, G.; Frattini, P. Comparing models of debris-flow susceptibility in the alpine environment. *Geomorphology* **2008**, *94*, 353–378. [CrossRef]

27. Chung, C.J.F.; Fabbri, A.G.; Van Westen, C.J. Multivariate regression analysis for landslide hazard zonation. In *Geographical Information Systems in Assessing Natural Hazards*; Advances in Natural and Technological Hazards Research; Carrara, A., Guzzetti, F., Eds.; Springer: Dordrecht, Germany; Berlin, Germany, 1995; Volume 5, pp. 107–133.

28. Guzzetti, F.; Carrara, A.; Cardinali, M.; Reichenbach, P. Landslide hazard evaluation: A review of current techniques and their application in a multi-scale study, Central Italy. *Geomorphology* **1999**, *31*, 181–216. [CrossRef]

29. Nefeslioglu, H.A.; Gokceoglu, C.; Sonmez, H. An assessment on the use of logistic regression and artificial neural networks with different sampling strategies for the preparation of landslide susceptibility maps. *Eng. Geol.* **2008**, *97*, 171–191. [CrossRef]

30. Segoni, S.; Pappafico, G.; Luti, T.; Catani, F. Landslide susceptibility assessment in complex geological settings: Sensitivity to geological information and insights on its parameterization. *Landslides* **2020**, 1–11. [CrossRef]

31. Tian, Y.; Xiao, C.; Liu, Y.; Wu, L. Effects of raster resolution on landslide susceptibility mapping: A case study of Shenzhen. *Sci. China Technol. Sc.* **2008**, *51*, 188–198. [CrossRef]

32. Alvioli, M.; Marchesini, I.; Reichenbach, P.; Rossi, M.; Ardizzone, F.; Fiorucci, F.; Guzzetti, F. Automatic delineation of geomorphological slope units with r. slopeunits v1. 0 and their optimization for landslide susceptibility modeling. *Geosci. Model Dev.* **2016**, *9*, 3975. [CrossRef]

33. Schlögel, R.; Marchesini, I.; Alvioli, M.; Reichenbach, P.; Rossi, M.; Malet, J.P. Optimizing landslide susceptibility zonation: Effects of DEM spatial resolution and slope unit delineation on logistic regression models. *Geomorphology* **2018**, *301*, 10–20. [CrossRef]

34. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]

35. Vargas, M. Revisão histórico-conceitual dos escorregamentos na serra do mar. *Bol. Solos e Rochas* **1999**, *22*, 53–83.

36. Dantas, M.E. *Geomorfologia do Estado do Rio de Janeiro*; Embrapa Solos: Niterói, Rio de Janeiro, Brazil, 2000; pp. 22–31. Available online: http://www.cprm.gov.br/publique/Gestao-Territorial/Geologia%2C-Meio-Ambiente-e-Saude/Projeto-Rio-de-Janeiro-3498.html/Geomorfologia. (accessed on 14 May 2020).

37. Avelar, A.S.; Netto, A.L.C.; Lacerda, W.A.; Becker, L.B.; Mendonça, M.B. Mechanisms of the recent catastrophic landslides in the mountainous range of Rio de Janeiro, Brazil. In *Landslide Science and Practice*; Margottini, C., Canuti, P., Sassa, K., Eds.; Springer: Berlin, Germany, 2013; pp. 265–270.

38. Vieira, B.C.; Fernandes, N.F. Landslides in Rio de Janeiro: The role played by variations in soil hydraulic conductivity. *Hydrol. Process.* **2004**, *18*, 791–805. [CrossRef]

39. Park, D.W.; Lee, S.R. Landslide and debris flow susceptibility zonation using TRIGRS for the 2011 Seoul landslide event. *Nat. Hazards Earth Syst. Sci.* **2013**, *13*, 2833–2849. [CrossRef]

40. Rosi, A.; Canavesi, V.; Segoni, S.; Dias Nery, T.; Catani, F.; Casagli, N. Landslides in the Mountain Region of Rio de Janeiro: A Proposal for the Semi-Automated Definition of Multiple Rainfall Thresholds. *Geosciences* **2019**, *9*, 203. [CrossRef]

41. Breiman, L. Bagging predictors. *Mach. Learn.* **1996**, *24*, 123–140. [CrossRef]

42. Brenning, A. Spatial prediction models for landslide hazards: Review, comparison and evaluation. *Nat. Hazard Earth Syst.* **2005**, *5*, 853–862. [CrossRef]

43. Chen, W.; Xie, X.; Wang, J.; Pradhan, B.; Hong, H.; Bui, D.T.; Duan, Z.; Ma, J. A comparative study of logistic model tree, random forest, and classification and regression tree models for spatial prediction of landslide susceptibility. *Catena* **2017**, *151*, 147–160. [CrossRef]

44. Lagomarsino, D.; Tofani, V.; Segoni, S.; Catani, F.; Casagli, N. A tool for classification and regression using random forest methodology: Applications to landslide susceptibility mapping and soil thickness modeling. *Environ. Model. Assess.* **2017**, *22*, 201–214. [CrossRef]

45. Pourghasemi, H.R.; Rahmati, O. Prediction of the landslide susceptibility: Which algorithm, which precision? *Catena* **2018**, *162*, 177–192. [CrossRef]

46. Valeriano, M.M. *Topodata: Guia Para Utilização de Dados Geomorfológicos Locais*; INPE: São José dos Campos, Brazil, 2008; Available online: http://mtc-m18.sid.inpe.br/col/sid.inpe.br/mtc-m18@80/2008/07.11.19.24/doc/publicacao.pdf (accessed on 15 October 2019).

47. Hervás, J.; Bobrowsky, P. Mapping: Inventories, susceptibility, hazard and risk. In *Landslides–Disaster Risk Reduction1*; Sassa, K., Canuti, P., Eds.; Springer: Berlin, Germany, 2009; pp. 321–349.

48. Nichol, J.; Wong, M.S. Satellite remote sensing for detailed landslide inventories using change detection and image fusion. *Int. J. Remote Sens.* **2005**, *26*, 1913–1926. [CrossRef]

49. Guzzetti, F.; Mondini, A.C.; Cardinali, M.; Fiorucci, F.; Santangelo, M.; Chang, K.T. Landslide inventory maps: New tools for an old problem. *Earth-Sci. Rev.* **2012**, *112*, 42–66. [CrossRef]

50. Battistini, A.; Segoni, S.; Manzo, G.; Catani, F.; Casagli, N. Web data mining for automatic inventory of geohazards at national scale. *Appl. Geogr.* **2013**, *43*, 147–158. [CrossRef]

51. Battistini, A.; Rosi, A.; Segoni, S.; Lagomarsino, D.; Catani, F.; Casagli, N. Validation of landslide hazard models using a semantic engine on online news. *Appl. Geogr.* **2017**, *82*, 59–65. [CrossRef]

52. Oliveira, N.S. Relação Entre Chuva e Deslizamento em Nova Friburgo/RJ. Master's Thesis, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil, July 2014.

53. Fressard, M.; Thiery, Y.; Maquaire, O. Which data for quantitative landslide susceptibility mapping at operational scale? Case study of the Pays d'Auge plateau hillslopes (Normandy, France). *Nat. Hazards Earth Syst. Sci.* **2014**, *14*, 569–588. [CrossRef]

54. Luti, T.; Segoni, S.; Catani, F.; Munafò, M.; Casagli, N. Integration of Remotely Sensed Soil Sealing Data in Landslide Susceptibility Mapping. *Remote Sens.* **2020**, *12*, 1486. [CrossRef]

55. Gray, G.L.; Leiser, A.J. Role of vegetation in stability and protection of slopes. In *Biotecnical Slope Protection and Erosion Control*; Van Nortrand Reinhol Company: New York, NY, USA, 1982; pp. 2–25.

56. Dietrich, W.E.; Wilson, C.J.; Reneau, S.L. Hollows, collu-viurn, and landslides in soil-mantled landscapes. In *Hillslope Processes*; Abrahams, A.D., Ed.; Allen and Unwin: Winchester, KY, USA, 1986; pp. 361–388.

57. Kanungo, D.P.; Arora, M.K.; Sarkar, S.; Gupta, R.P. Landslide susceptibility zonation (LSZ) mapping-A review. *J. South Asia Disaster Stud.* **2009**, *2*, 81–105.

58. Wilson, J.P.; Gallant, J.C. *Terrain Analysis: Principles and Applications*; John Wiley, Sons, Inc.: Toronto, ON, Canada, 2000; pp. 51–54.

59. Ohlmacher, G.C. Plan curvature and landslide probability in regions dominated by earth flows and earth slides. *Eng. Geol.* **2007**, *91*, 117–134. [CrossRef]

60. Carvalho, C.N.; Carvalho, R.C.; Salgado, C.M. Condicionantes geomorfológicos e da cobertura da terra na ocorrência de movimentos de massa na região serrana do Rio de Janeiro. *Cad. Geogr.* **2019**, *29*, 27–44.

61. Montgomery, D.R.; Dietrich, W.E. Where do channels begin. *Nature* **1988**, *336*, 232–234. [CrossRef]

62. Dahal, R.K.; Hasegawa, S.; Nonomura, A.; Yamanaka, M.; Masuda, T.; Nishino, K. GIS-based weights-of-evidence modelling of rainfall-induced landslides in small catchments for landslide susceptibility mapping. *Environ. Geol.* **2008**, *54*, 311–324. [CrossRef]

63. Beven, K.J.; Kirkby, M.J. A physically-based variable contributing area model of basin hydrology. *Hydrol. Sci. Bull.* **1979**, *24*, 43–69. [CrossRef]

64. Bagnold, R.A. *An Approach to the Sediment Transport Problem from General Physics*; Geolical Survey Profissional Paper, 422-I; United States Government Printing Office: Washington, DC, USA, 1966; p. 37.

65. Vorpahl, P.; Elsenbeer, H.; Märker, M.; Schröder, B. How can statistical models help to determine driving factors of landslides? *Ecol. Model.* **2012**, *239*, 27–39. [CrossRef]

66. Frattini, P.; Crosta, G.; Carrara, A. Techniques for evaluating the performance of landslide susceptibility models. *Eng. Geol.* **2010**, *111*, 62–72. [CrossRef]

67. Arabameri, A.; Rezaei, K.; Cerdà, A.; Conoscenti, C.; Kalantari, Z. A comparison of statistical methods and multi-criteria decision making to map flood hazard susceptibility in Northern Iran. *Sci. Total Environ.* **2019**, *660*, 443–458. [CrossRef] [PubMed]

68. Ng, V.W.; Breiman, L. Bivariate variable selection for classification problem. In *Technical Report*; Department of Statistics, University of California-Berkeley: Berkeley, CA, USA, 2005; pp. 1–22.

69. Segoni, S.; Lagomarsino, D.; Fanti, R.; Moretti, S.; Casagli, N. Integration of rainfall thresholds and susceptibility maps in the Emilia Romagna (Italy) regional-scale landslide warning system. *Landslides* **2015**, *12*, 773–785. [CrossRef]

70. Kayastha, P.; Dhital, M.R.; De Smedt, F. Landslide susceptibility mapping using the weight of evidence method in the Tinau watershed, Nepal. *Nat. Hazards* **2012**, *63*, 479–498. [CrossRef]

71. Günther, A.; Reichenbach, P.; Malet, J.-P.; Van Den Eeckhaut, M.; Hervás, J.; Dashwood, C.; Guzzetti, F. Tier-based approaches for landslide susceptibility assessment in Europe. *Landslides* **2013**, *10*, 529–546. [CrossRef]

72. Lee, C.T.; Huang, C.C.; Lee, J.F.; Pan, K.L.; Lin, M.L.; Dong, J.J. Statistical approach to earthquake-induced landslide susceptibility. *Eng. Geol.* **2008**, *100*, 43–58. [CrossRef]

73. Segoni, S.; Rosi, A.; Fanti, R.; Gallucci, A.; Monni, A.; Casagli, N. A Regional-Scale Landslide Warning System Based on 20 Years of Operational Experience. *Water* **2018**, *10*, 1297. [CrossRef]

74. Hong, Y.; Adler, R.F.; Huffman, G. An experimental global prediction system for rainfall-triggered landslides using satellite remote sensing and geospatial datasets. *IEEE Trans. Geosci. Remote* **2007**, *45*, 1671–1680. [CrossRef]

75. Segoni, S.; Tofani, V.; Rosi, A.; Catani, F.; Casagli, N. Combination of rainfall thresholds and susceptibility maps for dynamic landslide hazard assessment at regional scale. *Front. Earth Sci.* **2018**, *6*, 85. [CrossRef]