Landslide susceptibility mapping on the islands of Vulcano and Lipari (Aeolian Archipelago, Italy), using a multi-classification approach on conditioning factors and a modified GIS matrix method for areas lacking in a landslide inventory

Palenzuela JA*, Scifoni S**, Marsella M**, De Astis G***, Irigaray C*

*Department of Civil Engineering of the University of Granada, Granada, Spain
**Department of Civil, Building and Environmental Engineering, Sapienza University of Rome, Rome, Italy
***Sections of Seismology and Tectonophysics of the National Institute of Geophysics and Volcanology, Rome, Italy

Abstract

In areas prone to landslides, the identification of potentially unstable zones has a decisive impact on the risk assessment and development of mitigation plans. Active volcanic islands are particularly prone to instability phenomena as they are always in the early stage of dynamic unrest. A historical example of slope instability is the landslide which occurred in 1988 along the northwestern flank of La Fossa Cone on the island of Vulcano (Aeolian Archipelago). Based on this past activity, a susceptibility assessment using the bivariate technique of the GIS matrix method (GMM) was carried out on the islands of Lipari and Vulcano. Nevertheless, this case is congruent with those where a part of the surface was not assigned to stable or unstable areas, since a comprehensive inventory was only available for the island of Lipari. Some of the implemented steps of the susceptibility matrix method were modified to enable the model developed in the Lipari area to be applied to both islands. Considering the important role that the classification of conditioning factors plays in susceptibility analysis, the degree of association with landslide spatial distribution for the multiple classifications of each factor was assessed. Furthermore, an innovative clustering approach based on text and data mining techniques (self-organizing map neural network) was applied and compared with a heuristic classification of the categorical variable of lithology units. In addition to the extensive contingency analysis, up to 14 factor combinations were submitted to the GMM, validated and compared so as to select the one that best explains the susceptibility zoning. The effects of these incorporated processes in the previous phase of classification were discussed and reliminary susceptibility map was generated for both islands. After the validation of the susceptibility assessment, it is shown that the highest classes (High and Very High) matched 76.9% (relative accuracy) of the test inventory, while the lower susceptibility classes (Very Low and Low) resulted in a degree of fit of 14.39% (relative error).

Keywords
SOM, Modified GMM, DOF, Data mining, Text mining, Landslide susceptibility, GIS (geographic information system)

1. Introduction

Population growth, extending to mountainous areas, implies an increase in the number of properties, infrastructures and lives at risk. Landslides are often triggered by earthquakes and intense precipitations, but they are also induced by anthropogenic activity (Varnes 1984; Panizza 1996; Aleotti and Chowdhury 1999; Guzzetti et al. 1999; Lacasse and Nadim 2009). Thus, progress in landslide risk management mapping and its effective use for land planning is necessary to mitigate destructive effects. In the specific case of volcanic edifices, high and steep slopes in pyroclastic deposits are prone to slope instability phenomena. In the short term, a volcano edifice is a transient landform where unstable pyroclastic materials are easily involved in slope and surface water processes. In the long term, the magnitude of these processes decreases, but their impact on landscapes, especially if this includes human activity, can be major. Moreover, the combined effect of meteoric water and condensation water from the fumaroles degassing during periods of unrest (Marsella et al. 2015a) can infiltrate into these kinds of rocks. Thus, a major weakening of the mechanical strength of the rocks, caused by the negative effects of high water pore pressure, enhances the triggering of landslides. Considering this geological setting, slope failures, especially when concurrent to rainfalls, can produce debris avalanches and lahars that can rapidly travel long distances from the source (Thouret 1999).

Flank instability in volcanic environments is a phenomenon that has affected the study area, for example a landslide that affected the NE-flank of La Fossa volcano on the island of Vulcano. On April 20, 1988, after a period of unrest from June 1987 to 1990 (Barberi et al. 1991), which also caused a small tsunami with waves less than 1 m high (Tinti et al. 1999). Taking into account the potential emergence of this phenomenon, the scientific community has initiated a ground deformation monitoring program on the Lipari and Vulcano volcanic complex by using a dense GPSnetwork (Bonforte and Guglielmino 2008) and the Differential Synthetic Aperture Radar Interferometry technique (DInSAR) (Scifoni et al. 2015). In addition, research has been carried out to model the landslide mechanisms on the island of Vulcano (Marsella et al. 2013; Tinti et al. 1999). Building on the previous landslide risk assessment work, this current paper addresses the landslide susceptibility zoning for the problem area, as well as for those areas with limited data.

Despite the advantages brought by geographic information system (GIS) tools for geomorphologic mapping (Chacón et al. 2006; Wan 2009), the documentation available (aerial photographs, topographic map) and the scale at which it is produced can make the generation of a comprehensive inventory difficult. In addition, there are several aspects influencing the reliability (Malamud et al. 2004; Pavel et al. 2008, 2011) of the landslide inventory and its degree of association with each conditioning factor. Among these aspects, the land use, the cartographer's level of experience and their interpretation criteria and subjectivity (Wan 2009) are common examples. Consequently, geoscientists and engineers have to deal with non-assigned features—i.e. nonexistence of positive (zones with landslides) and/or negative data (zones without landslides) (Melchiorre et al. 2008)—which prevents the direct application of conventional techniques for susceptibility and hazard mapping. To overcome the constraints of nonassigned areas, the principle of similar landslides occurring in areas with similar conditions can be considered (Soeters and van Westen 1996). As discussed in literature (Ayalew and Yamagishi 2005), the analysis of the cause-effect relationships in susceptibility assessment is highly dependent on the selection of conditioning factors (Rahmati et al. 2016) and their classification.

In this work, the method GIS matrix method (GMM) (DeGraff and Romesburg 1980), which involves a statistical bivariate analysis applicable to areas with a set of simple parameters (starting with limited information), has been adopted for the assessment of landslide susceptibility. More specifically, landslide susceptibility zoning for the islands of Lipari and Vulcano was obtained by modifying the common GMM to make it applicable to areas without a landslide inventory. To improve the performance of the adopted model, two approaches for the classification of lithological units were used; a heuristic method and a self-organizing map (SOM)-based clustering. In addition, multiple conditioning factor combinations were

tested and validated. The proposed method may be adopted for a preliminary landslide susceptibility map in other areas without a pre-existing landslide inventory.

2. Study area

The Aeolian Archipelago (Southern Tyrrhenian Sea, Italy) is located between the northeastern Sicilian coast and the Tyrrhenian back-arc basin, almost on the boundary between the African and Eurasian plates (Ventura 2013) (Fig. 1). The seven Aeolian Islands are composite volcanic structures formed by the superposition of multiple centres and recurrent volcanic and tectonic events, with many sectors and central collapses. These islands lie on the Calabro–Peloritano basement, a block of the European continent that detached itself from Sardinia–Corsica and migrated southwestwards during the Miocene–Quaternary opening of the Tyrrhenian Sea. The Aeolian arc is cut by a complex pattern of fault systems (Ventura 2013). Structural, seismological and volcanological data (Ventura et al. 1999; De Astis et al. 2003; Ventura 2013) suggest that under present-day strain conditions, there is a strike-slip deformation in the Aeolian central sector—the Tindari– Letojanni tectonic system (Locardi and Nappi 1979; Ventura 1994; Mazzuoli et al. 1995; Ventura 1995; De Astis et al. 2003; Ventura 2013), where Vulcano is located, at the southeastern end of NNW– SSE-elongated Salina–Lipari belt.

Accordingly, the structural pattern of Vulcano is dominated by a major NW–SE- to NNW–SSE-strike fault system (Frazzetta et al. 1982; Ventura 1994; Mazzuoli et al. 1995; Ventura 1995; Ventura et al. 1999), which is the shallow expression of the Tindari– Letojanni (TL) system (Continisio et al. 1997; Lanzafame and Bousquet 1997). This study is focused on the islands of Vulcano and Lipari, which have been characterized by different types of landslides (such as rock fall and debris flow). On Vulcano, the areas with the most active slope processes are at the La Fossa cone and caldera. The northern sector of the La Fossa cone summit is the site of intense gas emissions from several fumaroles that have reached Tmax~700 °C (January 1993), decreasing to the present Tmax of about 400 °C. The cone is quite regular in shape, with the exception of the northeastern sector of the Forgia Vecchia (FV) parasitic crater where quite a large depression corresponding to the crater rim is clearly visible (Marsella et al. 2011, 2015b). Different types of instability processes were documented or recognized on the La Fossa cone from morphological evidence (Tommasi et al. 2007). The 1988 landslide (~ 200,000 m3 in volume) affected some layers of pyroclastic material (Tinti et al. 1999).

The island of Lipari is located in the central sector of the Aeolian Archipelago and is the largest (total area of 38 km2) of the seven islands. Several, partially overlapping, volcanic edifices have been active at different times, essentially controlled by the main NNW–SSE and N–S regional tectonic trends. The 776 AD Monte Pilato eruption represents the youngest volcanic activity (Forni et al. 2013; Lucchi 2013), which occurred on the island that nowadays is in a quiescent stage with a few active low-temperature fumaroles and hot springs, mostly located in the eastern sector of the island close to Bagni Termali di San Calogero and Bagno Secco and the most inhabited areas of the western sector. As a result of its volcanic origin, the island is covered with pumice and cineritic pyroclastic rocks that have caused translational, rotational landslides and debris flows. The structural pattern observed on Lipari is dominated by major faults and alignments of eruptive centres that follow the NNW–SSE direction of the TL Fault System. Four volcano-tectonic collapse structures have been recognized on Lipari from subvertical escarpments with curved geometries, whose rims are invariably marked by high-angle unconformities between the pre-collapse and in-filling volcanic products post-collapse (Forni et al. 2013).

The complete geological evolution, stratigraphy and volcanological history of both Vulcano and Lipari are available on the 1:10,000 scale geological maps and in associated articles which have recently been published: De De Astis et al. (2013) and Lucchi (2013), respectively.

In general, the geological structures and landforms, such as the ancient rims of craters, collapses and faults with high slopes and fractured rocks, are areas prone to landslides.

3. Susceptibility zoning

The results of a susceptibility assessment depend on the variables or conditioning factors taken into account, and the classes into which they are divided. Accordingly, in this research, specific processes in the classification of numerical factors and lithology (nominal data) were addressed with the aim of not limiting the classification to a single grouping of data or a single approach. Thus, more than one result could be compared and the most suitable classification for each factor selected. Similarly, to determine the best set of variables, multiple factor combinations were introduced into the GMM and their resulting susceptibility distributions validated. In addition, taking into account the lack of an inventory for the area of the island of Vulcano, the general GMM was modified, enabling its application on an area without a landslide inventory.

3.1 Definition of conditioning factors

A susceptibility assessment depends directly on different factors, each one favoring to an extent the possibility of one type of landslide occurring in a specific zone. Owing to the complex nature of these factors, the variety of their characteristics and their own spatial distribution, selecting a simple combination of conditioning factors is not trivial. Furthermore, the availability of the information related to these factors can prevent the use of more than one variable in the susceptibility assessment. This fact leads experts to decide on the specific set of variables selected and try to validate different combinations of factors or scenarios (Meten et al. 2015; Van Westen et al. 2003). In this paper, the susceptibility analysis was carried out based on the available information from which seven factors were considered. Nonetheless, only the land use, obtained in a vector format (.shp) could directly be used. On the contrary, the lithology map and its corresponding database were digitalized from the scanned cartography, and the remaining factors (elevation, aspect, slope and LiDAR intensity) were generated from the LiDAR raw data.

Below, the landslide inventory and conditioning factors used in this work are described:

- A landslide inventory generated for the island of Lipari, supporte by the interpretation of two sets of aerial photographs at a scale of 1:36,000 and taken in 1964 and 2005, was used. However, a comprehensive inventory for the island of Vulcano was not available.

- The data acquired by the airborne LiDAR (Light Detection and Ranging) scanning technique and ALS (Aerial Laser Scanning), during a survey carried out in 2008 by the Italian Ministry for Environment, were used to generate additional variables. This ALS dataset consists of a dense Bpoint cloud^ recorded during a LiDAR survey, represented by a set of 3D coordinates (X, Y and Z). Another important parameter is the reflected intensity (I) of the laser beam (Marsella et al. 2015b). This parameter is a function of the distance to the object, the angle of incidence of the laser beam and the reflectance properties of the object (D'Aranno et al. 2015). The usefulness of this data source has been proven when performing vulnerability, susceptibility or hazard assessments (Abdulwahid and Pradhan 2017; Gorsevski et al. 2016) as various factors can be extracted from it. In addition to that, it is well known that water saturation is a very important variable as a conditioning and triggering factor, while the vegetation is a significant factor that controls the ground water infiltration and accumulation (Xiao et al. 2017). Applications of LiDAR intensity to differentiate the vertical layers of the vegetation can be found, despite its limitations. Other than its own measurement platform characteristics (altitude and incidence angle), the canopy structure is dispersed within the LiDAR footprint and its complex composition of elements with different sizes, orientations and reflectance values makes it difficult to associate the reflected intensity with the different types of plants (Morsdorf et al. 2010; Jason and Bork 2007; Kim et al. 2009). Similarly, studies have been carried out to differentiate the land-cover objects. In this case, despite the fact that in practice, LiDAR intensity

is not easily correlated with the theoretical reflectance of materials, the relative reflectance allows for the separation of different classes of materials by filtering or normalizing the intensity values (Song et al. 2002). This paper is not intended to provide a detailed land use or vegetation classification via LiDAR intensity. Instead, a simple normalization and classification of LiDAR intensity values (in the range 0–1) was performed (Fig. 2). Thus, the higher intensity values have been assumed to correspond with less scattering materials like bare rock or artificial targets; in general, with well-defined planes and therefore with less scatter effects. They are also assumed to be more prone to mass movements. On the contrary, low-intensity classes will be assumed to be more correlated with vegetated areas with different species and densities, preventing in some proportion the water infiltration. In addition, the high-resolution LiDAR point cloud, with a mean spacing of 2 m, enabled the building of high-quality digital elevation models (DEMs) of both islands. From these DEMs, the elevation map, the aspect map and the slope map were generated (Fig. 2). Elevation is considered as an indirect expression of some environmental settings such as vegetation types and rainfall, while slope aspect is related to the illumination time and thus, to the soil moisture. Slope or slope angle will have a direct impact on the gravitational forces that favor the downwards movement and so, determine the equilibrium state of a landslide.

- Another continuous variable was derived by computing the distances from steep terrain features that are common in volcanic areas. These are geological structures integrated to craters and collapse escarpments, whose higher traces were digitalized and distances to them from the lower terrain were computed and represented in a layer named distances to geological structures (Fig. 3). These structures and their proximities are characterized by altered rocks as the meteoritic products of the high degassing rates (volcanic gases) and hydrothermal alteration. Furthermore, the hillside in the surroundings of these structures is of high to nearly vertical slopes, favoring the generation of different types of landslides, like debris flows, rock falls and toppling. Moreover, the high slopes favor the long runoff tracks of rock fragments.

The land-use map (Fig. 3) provided by Sicily Regional Administration (Italy) was also included in this analysis which contains eight well defined classes. Land use may be related to deliberate water supply, slope changes and modification of the soil and regolith thickness, which can also affect the equilibrium state of the slope mass.

The numerical factors derived from LiDAR data (LiDAR intensity, DEM, slope and aspect) and the distance to geological structures were subdivided into classes using the natural-breaks method (Jenks 1967). However, to determine the effect of classification on the correlation between independent variables and the dependent variable (landslide occurrence), 9 different classifications were applied, using classes ranging from 2 to 10 for each of these continuous variables.

Considering that the rheology properties of rock and soil layers will affect the strength of the slope mass, the lithology layer was also included in this study. As for the lack of geotechnical parameters to be mapped, spatial dedication was put in the lithology classification. Thus, two approaches were used to group the original units (each single polygon of the geodatabase) from the geological maps of Lipari and Vulcano (Tranne et al. 2002; De Astis et al. 2006). First, the minor units were grouped into major units through a heuristic classification based on expert criterion. This approach consisted of reviewing the different lithological units of the geological map and their common properties to group them into major units. To help carry out this classification, the descriptive characteristics defined by Protodyakonov (1962) for the estimation of the relative hardness of rock and soils were utilized. As a result, all the minor units were classified into 17 classes (numbered in Fig. 1 as follows): (1) altered lava dome, (2) altered lava flow, (3) altered pyroclastic unit, (4) clastogenic lava flow, (5) colluvial and alluvial deposit, (6) conglomerate, (7) dike, (8) lava dome, (9) lava flow, (10) lava flow with pyroclastic layers, (11) littoral deposit, (12) locally altered lava flow, (13) obsidian rich lava flow, (14) pumice rich pyroclastic rock, (15) pyroclastic rock, (16) urbanized area and (17) volcaniclastic deposit. On the contrary, the second

approach consisted of a semiquantitative classification of the lithological units based on data mining techniques was addressed in two major phases by using Statistica software (TIBCO 2017) (Fig. 4). After applying both methods, the impact of each classification on the final susceptibility assessment performance was compared.

This approach was carried out in the following order:

1. A text mining processing was used to prepare the data mining input. This process receives the lithological information stored in the database field of unit description. From this information, through text mining processing, linguistic terms are summarized and converted into quantitative variables. This technique is based on statistical natural language processing, and here it is only used to organize text for each case. The process is briefly described below, while a sound description of methods for processing unstructured (or textual) data can be found in Manning and Schütze (1999). Firstly, common terms like pronouns, adverbs and verbs are filtered by comparing the original text describing the lithological units with a text database. At the same time, the terms with the same root and synonyms are grouped together and then supervised to link those with the same meaning. Additionally, other synonyms are selected and grouped into a single term (e.g. layer and unit). At the end of this step, only the interesting terms remain (aa-type, alternate, ashy, blocky, brecciated, coarse, coherent, detrital, etc.). Secondly, the program automatically searches and calculates the frequency for every variable or term that appears in each case (lithology polygon). As the purpose of this research is to group these variables describing lithological features into a lower dimensional space, it is only necessary to know whether these terms appear or not in each case. Therefore, the frequency values are substituted by 0 (if the frequency is 0) or 1 (if the frequency is different from 0).

2. Once the categorical variables are converted into numerical values (0 and 1), the next phase consists of applying a SOM classifier for the unsupervised clustering of lithological units. All the ANN's (artificial neural networks) are built with basic units called neural network processing units (henceforth, NPU). The simplest model of an artificial neuron or perceptron proposed by McCulloch and Pitts (1943) consists of a single NPU that attempts to simulate the behaviour of the biological neural network in a human brain (Nedjah and de Macedo Mourelle 2007). These NPUs receive incoming signals (data inputs) of a n-dimensional Euclidean space (Rn) represented by vectors of the type ai = (x1,x2,…,xn), and uses a function of these inputs, f(x), to generate the outputs or activity states.

3.2 Classification methods

In an artificial neuron, their synapses (connection with inputs or other neurons) are represented by weights (Wi = w1, w2, …,wn) with exciting (positive) or inhibitory (negative) values, all of which are applied to the input data. The total entry of a NPU is then calculated at the sum of its weighted inputs (Fig. 5a). A simple NPU can learn and map a linear function, but to accommodate some functions, for example when every input and its corresponding weight are 0 but the predicted output must be different from 0, a constant or "bias" is needed. Therefore, a biased weighted sum is used in this case (Eq. 1):

$$\sum_{i}^{n} (w_n \cdot a_n) + c$$

In the following step, the biased weighted sum is analyzed using the activation function (or transfer function) of the neuron to decide its activation state: activated (1) or deactivated (0). The most popular activation function is the step (threshold) function, which results in 1 when the input sum is equal or greater than 0, or 0 in other cases. When more than one neuron is run in a neural network (NN), each NPU can learn or approach a segment of a non-linear function. A NN is defined by a topology formed by different neuron layers, with known weights and transfer functions for all NPU's. However, both weights

and the bias parameter have to be iteratively adjusted during the training phase so that the input data is correctly mapped to one of the expected results. The most common training method is that which is based on a feed-forward network (or perceptron), where all the inputs are processed from one layer to the following layer without cycles or feedback loops that use outputs as new input in previous layers.

In this research, a SOM (Kohonen 1998, 2001) also known as a self-organizing feature map (SFOM) is used. Thus, the Euclidea distance between two points (a, b) can be assessed through Eq. 2:

$$d = \|a-b\| = \sqrt{\sum_{i=1}^{n} (a_i-b_i)^2} = \sqrt{(a_1-b_1)^2 + \ldots + (a_n-b_n)^2}$$

SOM consists of two layers, one input layer and one output layer. The output layer is a competitive layer in which a single neuron (representative of a cluster or class) is activated for an input vector. This network is characterized by using unsupervised training for weight adjustment, so that each one can recognize or link an input vector with characteristic features without needing a training set. It also does not contain activation functions, as it only needs to determine which the winning neuron is to continue with the self-organizing process.

The training algorithm consists of three principles:

- Competition: the input vector is compared with every weight vector of each output neuron (j): Xi against Wj. The winning neuron is that with the closest Wj to the input Xi, in terms of Euclidean distance.
- Cooperation: the winning neuron strengthens its neighbours in the output layer through a neighbourhood function (F). Usually, this neighbourhood function is a Gaussian function (Fig. 5b, Eq. 3):

  $$h_{\text{winner on } j}(d) = e^{\frac{-d^2}{\sigma^2}}$$

- Adaptation: once an input Xi has been processed, all the weights (Wj) are adjusted to make them closer to the input. To recalculate weights, Eq. 4 is used:

  $$W_j' = W_j + \alpha h\left[X_i - W_j\right]$$

  where α is the learning rate coefficient. FromEq. 4, it follows that the amount of change depends on the neighbourhood (h), i.e. the distances between the neighbouring neurons and the winning neuron. However, h decreases with time.

The performance of unsupervised learning in clustering techniques can be evaluated by using the sum of squared errors (SSE) within clusters (Eq. 5):

$$\text{SSE} = \sum_{i=1}^{k} \sum_{p \in C_j} (p-m_j)^2$$

where k is the number of clusters, C the set of cases in a cluster and m the center point of a cluster. Thus, the lower the SSE, the lower the network error, which depends on the distances between each cluster object (p) and the cluster centre (mj).

In addition to the adjusted weights, the topology is an important factor due to its performance as a SOMneural network (Ermini et al. 2005; Melchiorre et al. 2008). Accordingly, in this research, different topologies were tested so as to compare the clustering error. The SOM structure was iteratively changed by altering the number of the NPUs in the lattice of the output layer, as indicated in Table 1.

Finally, both nominal variables, land use and lithology, were converted to numerical classes (ordered integers) before being utilized in the GMM model.

3.3 Susceptibility assessment

This susceptibility assessment involves a modification of the GMM, which is based on a cross analysis of determinant factor maps and the spatial frequency of landslides. This method is not capable of predicting the susceptibility to slope movements with absolute probability; however, it enables the assessment of the potential relative instability in a broad region by using a series of classified factors. The GMM requires an inventory of landslides and the selection of the most significant determinant factors to be included in the analysis. In addition, to ensure the effectiveness of thismethod, the bivariate analysis of factor classifications, as well as the susceptibility map validation, plays an important role. Consequently, this method can be divided into three major phases (Fig. 4): contingency analysis, generation of the susceptibility matrix and susceptibility map validation.

3.3.1 Contingency analysis

After generating the multiple classifications of continuous factors through the natural breaks method and the two classifications of lithological units, a contingency analysis was carried out. The Pearson coefficient (C) was used as the measurement of the degree of association between conditioning factors and the landslide spatial distribution. It is a very common parameter used to assess the strength of association between independent and dependent variables (C = 0 indicates no dependence and C = 1 indicates the maximum dependence). However, this parameter is related to the data independence, but it was also affected by sample size. This fact can prevent a pair of variables from reaching higher C values. Thus, the percentage of the adjusted C (%C/Cmax), based on the maximum C that can be reached, was also calculated here.

3.3.2 GIS matrix method and validation

When applying the GMM, three major stages are carried out:

1. All the possible geographical intersections between determinant factor classes are obtained and the mobilized area (training landslide inventory) associated with each combination is calculated and saved in the landslide matrix (LM).
2. The total area covered by each combination of factor classes is determined in the total surface matrix (TSM).
3. The susceptibility matrix (SM) is calculated by dividing LM by TSM. This matrix contains the percentage of mobilized area linked to each factor class combination.

The entire process for applying the GMM and the validation of the susceptibility map has been implemented in different cases, providing successful results (Irigaray 1995; Irigaray et al. 2007; Jiménez-erálvarez et al. 2011), and has been used in areas with an existing inventory. However, given that the only landslide inventory used here was that of the island of Lipari, two modifications were integrated within the three previous stages of the common GMM.

i. First, the geographical intersection of factor classes that are common to both islands was calculated. With this step, a first vector file was generated from the factor class intersection, which was clipped to only contain the intersection corresponding to the Lipari area. This second file was used to calculate the LM, the TSM and the SM by using the Lipari landslide inventory. Similarly, the validation step was applied only to the analysis area of Lipari.
ii. Once the different tests of the GMM with different conditioning factor combinations were validated and the best model selected, this model was applied again over the entire area covered by both islands (application area).

After applying the contingency analysis and selecting the best classifications of conditioning factors to be included in the susceptibility model, the intersection of factor classes was generated (step i), and the SM was calculated for the island of Lipari. However, to improve the performance of the GMM, multiple combinations of the conditioning factors were tested by omitting the factors with the lowest adjusted C. Each attempt was validated by using the degree of fit (DOF) (Eq. 6) as a measurement of the degree of association between the test inventory (30% of themapped landslides) and the landslide susceptibility map. Thus, the effectiveness of each susceptibility map was assessed by using a spatial autocorrelation analysis regarding the Goodchild equation (Goodchild 1986):

$$\mathrm{DOF}_i = \frac{Z_i/S_i}{\sum Z_i/S_i}$$

Where

- DOFi represents the percentage of mobilized area that matches each susceptibility class (i).
- Zi represents the mobilized area in each susceptibility class.
- Si represents the total area covered by each susceptibility class.

After checking the validation results, certain errors can appear which are linked to the low and very low susceptibility classes (relative error), while DOF increases in the high or very high susceptibility classes (relative accuracy) (Fernández et al. 2003; Irigaray et al. 2007).

The modified GMM was implemented using ArcGIS (ESRI 2018) by creating new tools that can be found in Palenzuela (2018) (Fig. 6).

4. Results

Prior to the susceptibility assessment, the results of the contingency analysis were reviewed. Table 2 contains two cases linked to lithology, since both heuristic and semiquantitative techniques for lithology classification were compared. For unsupervised clustering, the input data was split at 70%. Therefore, 70% of the dataset was used, first to train the NN, and then, the error (Eq. 5) was calculated for the test dataset. After plotting the SOM error, a relative minimum error is observed when running a topology of 10 neurons (2 × 5) (Fig. 7). From this point, the error progressively decreases as the number of neuron increases. Nonetheless, by overfitting, this can lead to the increase in neural network noise, resulting in a loss of generalization capability (Melchiorre et al. 2008) as the number of neurons increases. Taking these aspects into account, the topology of 10 neurons (10 classes) was used for the purpose of comparison. Since the urbanized area class was well identified, it was directly added to the class set resulting from the unsupervised clustering, making a total of 11 classes.

When comparing the adjusted C for both lithology classifications (Table 2), the highest value for the heuristic classification based on the estimated material strength was observed. In addition, by directly reviewing the automatic classification (SOM classification) of lithotypes, misclassifications were observed with regard to genesis and major characteristics that make each lithological unit clearly distinguishable. For instance, rocks (andesite, basalt, obsidian and other) and loose deposits (littoral, colluvial and alluvial) were joined together in some classes. Accordingly, the heuristic classification was included in the GMM model. However, it is worth noting that the unsupervised clustering based on text and data mining techniques showed the capacity to reduce the dimensionality (number of classes) by 35.3% (11 classes against 17), while the adjusted C only decreased by 3.3% (13.1 against 9.8) (Table 3).

For every continuous variable (LiDAR intensity, DEM, slope and aspect, distance to collapses and craters), the classification falls at the inflection point of the adjusted C, from which the trend becomes more pronounced (Fig. 8), was selected.

Finally, the classifications selected for all the conditioning factors were tabulated and ordered (Table 3) to take into account the importance of these variables for the susceptibility analysis. In general, the strength of the associations between explanatory factors and the landslide spatial distribution is low, giving the highest values for the variables of lithology, elevation and slope.

The previous coefficient considers each separated factor. However, to evaluate the effectiveness of the susceptibility analysis, 14 tests of the GMM were performed and validated by using 14 combinations of the conditioning factors (Table 4). The first test (test 1) included all the factors. The second one omitted the factor with the lowest C. In the following tests of the GMM, pairs of factors were formed with one of the variables with the lowest contingency coefficient (LiDAR intensity and distances to collapses and craters), together with one of the remaining factors were progressively omitted.

To compare the effectiveness of every combination, the degree of fit was calculated for every GMM test (Table 5). To better observe the separation between the area covered by the lowest susceptibility levels—Very Low and Low—and the highest values—High and Very High—the difference between both pairs of values was assessed. When reviewing the validation results, the best match between the test inventory and the susceptibility classes was found in the second test, where only LiDAR intensity was omitted in the GMM model. The difference between the highest and lowest values for this test resulted in 62.49%.

By selecting the input variables of the second test, the GMM was applied to the factor combinations common to both islands. Finally, the resulting values were reclassified into five susceptibility levels by using the natural-breaks method. These classes are “Very Low”, “Low”, “Moderate”, “High” and “Very High” and they express the susceptibility map of Fig. 9.

5. Discussion and conclusions

Susceptibility analysis is a very important component in landslide hazard assessment, and hence, a necessary element for the mitigation of risk to infrastructure and human lives. Accordingly, the major advancement of this paper is on the adaptation of the GMM to be applied to an area of terrain without a landslide inventory but with similar settings to those of another area with an existing inventory. This was not covered by previous models of the GMM, only being acceptable to be used in areas with an existing assignation of stable and unstable zones or a landslide inventory (e.g. Jiménez-Perálvarez et al. 2011; Irigaray et al. 2007). Thus, this adapted methodology enables the generation of a preliminary susceptibility zoning. This susceptibility assessment can be started from the only use of causative factors that are first combined in the analysis area (classified in stable and unstable zones) and then the susceptibility matrix (SM) is transferred to the application area (both areas with or without a landslide inventory). In this research, the area without a landslide inventory coincides with the entire expanse of Vulcano Island, while the analysis area was established for the island of Lipari sharing similar geomorphological and geological settings.

In the field of susceptibility analysis, some works deal with the validation of more than one causative factor combination (e.g. Van Westen et al. 2003), which permits to select the data set with the most suitable factors that better explain the landslide susceptibility levels. Besides that, in this research, an analysis and comparison of the internal classification of the conditioning factors were also considered to improve the quality of the results. Thus, each quantitative variable was divided into a different number of classes and the contingency tables were then used to select the classification that best explains the independent variable (landslide susceptibility). Furthermore, a bimodal (heuristic and unsupervised) approach to classify lithology when no geotechnical parameters are directly known was applied, and the Pearson coefficient (C) was also used here to decide the best classification method. The heuristic classification depends on the order of lithologies based on their relative strength that is given in relation to their characteristics. This can lead to some subjective judgement and the number of classes will coincide with the number of lithologies that is registered in the existing cartography carried out by the expert. To try minimizing this subjectivity, an innovation in the classification process was introduced by

applying an advanced data mining methodology. This is done by combining text mining- and artificial neural network-based techniques; with the aim of the semiautomatic processing and clustering of categorical variables within a certain number of classes. The unsupervised training of a self-organizing map, the network topology, the wrong variables and noisy data are factors affecting network performance (Ermini et al. 2005). Thus, a text mining pre-process was performed for the selection of the most important terms (nominal variables) associated with the characteristics of the lithological units. In addition, nine topologies were tried, and the internal error of the clusters was examined to select a classification that maintains generalization and avoids noisy learning. After addressing the contingency analysis, the highest Pearson C and the misclassifications observed in the neural network classification were considered, so that the heuristic classification could be selected. Nonetheless, it is worth noting that the SOM network reduced the number of classes considerably by 35.3% and only decreased the adjusted Pearson coefficient by 3.3%. This case, in which a qualitative method relying on expert rules to give a better result, is consistent with previous works (Bourenane et al. 2015). However, it is known that unsupervised clustering is a very suitable technique for discovering data structures where previous classes are not defined (Melchiorre et al. 2008). Therefore, one cannot exclude that improvements in the classification can be made by adding other inputs (features of the objects to be classified) and datasets to the NN at the training stage.

Similarly, to raise the objectivity, the GMM was modified and implemented in ArcGIS® to apply the trained matrix to an area lacking in a comprehensive landslide inventory. The modified GMM allowed for the generation of the susceptibility matrix for the island of Lipari (with an existing landslide inventory), and the application, such as the susceptibility matrix to the entire complex formed by both islands (Lipari and Vulcano). From the findings of this research, the modified GMM represents a useful contribution to identify spatial distribution of potential unstable areas without a landslide inventory. Nonetheless, before applying this methodology it must be assumed that landslides will occur under the same geological, geomorphological, hydrogeological and climatic conditions as in the past (Aleotti and Chowdhury 1999; Jiménez-Perálvarez et al. 2011).

Despite a multi-classification approach being used for the conditioning variables, and up to 14 combinations of these factors being tested, a more reliable and precise susceptibility mapping could be generated by incorporating additional improvements. Among them, an updated landslide inventory and different training and test datasets (Melchiorre et al. 2008) can be introduced into the model being validated. Finally, considering that the present method is based on a recursive process, improvements can be made by collecting additional factors and performing new tests of the GMM on their classifications. This is how the bivariate statistics (contingency analysis) and the validation of each GMM test can be applied to distinguish the factors that will best explain the landslide susceptibility levels.

Funding information
This work was supported by the DPC-INGV Project V3 on the island of Vulcano (http://sites.google.com/site/progettivulcanologici), founded by the Italian National Institute of Geophysics and Volcanology and by the Italian Civil Protection Department. The 2008 ALS DTM was provided by the Italian Ministry for Environment.

References

Abdulwahid WM, Pradhan B (2017) Landslide vulnerability and risk assessment for multihazard scenarios using airborne laser scanning data (LiDAR). Landslides 14(3):1057–1076. https://doi.org/10.1007/s10346-016-0744-0

Aleotti P, Chowdhury R (1999) Landslide hazard assessment: summary review and new perspectives. Bull Eng Geol Env 58(1):21–44 Ayalew L, Yamagishi H (2005) The application of GIS-based logistic

regression for landslide susceptibility mapping in the Kakuda-Yahiko Mountains, Central Japan. Geomorphology 65(1):15–31. https://doi.org/10.1016/j.geomorph.2004.06.010

Barberi F, Neri G, Valenza M, Villari L (1991) 1987-1990 unrest at Vulcano. Acta Vulcanol 1:95–106 Bonforte A, Guglielmino F (2008) Transpressive strain on the Lipari–Vulcano volcanic complex and dynamics of the BLa Fossa^ cone (Aeolian Islands, Sicily) revealed by GPS surveys on a dense network. Tectonophysics 457(1):64–70. https://doi.org/10.1016/j.tecto.2008.05.016

Bourenane H, Bouhadad Y, Guettouche MS, Braham M (2015) GIS-based landslide susceptibility zonation using bivariate statistical and expert approaches in the city of Constantine (Northeast Algeria). Bull Eng Geol Environ 74(2):337–355

Chacón J, Irigaray C, Fernandez T, El Hamdouni R (2006) Engineering geology maps: landslides and geographical information systems. Bull Eng Geol Environ 65(4):341– 411. https://doi.org/10.1007/s10064-006-0064-z

Continisio R, Ferrucci F, Gaudiosi G, Lo Bascio D, Ventura G (1997) Malta escarpment and Mt. Etna: early stages of an asymmetric rifting process? Evidences from geophysical and geological data. Acta Vulcanol 9:39–47

D'Aranno PJV, Marsella M, Scifoni S, Scutti M, Sonnessa A, Bonano M (2015) Advanced DInSAR analysis for building damage assessment in large urban areas: an application to the city of Roma, Italy. Dissertation, Proceedings of SPIE - The International Society for Optical Engineering 9642. Doi https://doi.org/10.1117/12.2194808

De Astis G, Ventura G, Vilardo G (2003) Geodynamic significance of the Aeolian volcanism (southern Tyrrhenian Sea, Italy) in light of structural, seismological, and geochemical data. Tectonics 22(4). Doi https://doi.org/10.1029/2003TC001506

De Astis G, Dellino P, La Volpe L, Lucchi F, Tranne CA (2006) Geological map of the island of Vulcano (Aeolian Islands). University of Bari, University of Bologna and INGV. L.A.C, Firenze

De Astis G, Lucchi F, Dellino P, La Volpe L, Tranne CA, Frezzotti ML, Peccerillo A (2013) Chapter 11 Geology, volcanic history and petrology of Vulcano (central Aeolian Archipelago). In: Lucchi F, Peccerillo A, Keller J, Tranne CA, Rossi PL (eds) The Aeolian Islands volcanoes, Memoirs 37. Geological Society, London, pp 281–349. https://doi.org/10.1144/m37.1

DeGraff J, Romesburg C (1980) Regional landslide susceptibility assessment for wildland management: a matrix approach. In: Coats CR, Vitek J (eds) Thresholds in geomorphology, chapter 19. Allen and Unwin, London, pp 401–414

Ermini L, Catani F, Casagli N (2005) Artificial neural networks applied to landslide susceptibility assessment. Geomorphology 66(1):327–343. https://doi.org/10.1016/j.geomorph.2004.09.025

ESRI (2018) About ArcGIS The mapping and analytics platform. https://www.esri.com/ en-us/arcgis/about-arcgis/overview.

Fernández T, Irigaray C, El Hamdouni R, Chacón J (2003) Methodology for landslide susceptibility mapping by means of a GIS. Application to the contraviesa area (Granada, Spain). Nat Hazards 30(3):297–308. https://doi.org/10.1023/B%3ANHAZ.0000007092.51910.3f

Forni F, Lucchi F, Peccerillo A, Tranne CA, Rossi PL, Frezzotti ML (2013) Chapter 10 stratigraphy and geological evolution of the Lipari volcanic complex (central Aeolian archipelago). In: Lucchi F, Peccerillo A, Keller J, Tranne CA, Rossi PL (eds) The Aeolian Islands volcanoes, Memoir 37. Geological, Society, London, pp 213–279. https://doi.org/10.1144/m37.10

Frazzetta G, Lanzafame G, Villari L (1982) Deformazioni e tettonica attiva a Liapri e Vulcano (Eolie). Mem Soc Geol It 24:293–297

Goodchild MF (1986) Spatial autocorrelation. Catmog 47. Geo Books, Norwich, p 56

Gorsevski PV, Brown MK, Panter K, Onasch CM, Simic A, Snyder J (2016) Landslide detection and susceptibility mapping using LiDAR and an artificial neural network approach: a case study in the Cuyahoga Valley National Park, Ohio. Landslides 13(3):467–484. https://doi.org/10.1007/s10346-015-0587-0

Guzzetti F, Carrara A, Cardinali M, Reichenbach P (1999) Landslide hazard evaluation: a review of current techniques and their application in a multi-scale study, Central Italy. Geomorphology 31(1–4):181–216. https://doi.org/10.1016/S0169-555X(99)00078-1

Irigaray C (1995) Movimientos de Ladera: Inventario, Análisis y Cartografía de la Susceptibilidad Mediante un Sistema de Información Geográfica. Aplicación a las Zonas de Colmenar (Málaga), Rute (Córdoba) y Montefrío (Granada). Dissertation, University of Granada, Spain, p. 578

Irigaray C, Fernandez T, El Hamdouni R, Chacon J (2007) Evaluation and validation of landslide-susceptibility maps obtained by a GIS matrix method: examples from the Betic Cordillera (southern Spain). Nat Hazards 41(1):61–79. https://doi.org/10.1007/s11069-006-9027-8

Jason GS, Bork E (2007) Characterization of diverse plant communities in Aspen Parkland rangeland using LIDAR data. Appl Veg Sci 10:407–416 Jenks GF (1967) The data model concept in statistical mapping. Int Yearb Cartogr 7:186–190

Jiménez-Perálvarez JD, Irigaray C, El Hamdouni R, Chacón J (2011) Landslide susceptibility mapping in a semi-arid mountain environment: an example from the southern slopes of Sierra Nevada (Granada, Spain). Bull Eng Geol Environ 70(2):265–277. https://doi.org/10.1007/s10064-010-0332-9

Kim S, McGaughey RJ, Andersen HE, Schreuder G (2009) Tree species differentiation using intensity data derived from leaf-on and leaf-off airborne laser scanner data. Remote Sens Environ 113(8):1575–1586. https://doi.org/10.1016/j.rse.2009.03.017

Kohonen T (1998) The self-organizing map. Neurocomputing 21(1):1–6. https://doi.org/10.1016/S0925-2312(98)00030-7

Kohonen T (2001) Self-organizing maps. In: Huang TS, Kohonen T, Schroeder MR, Manfred R (eds) Springer series in information sciences. Berlin Heidelberg, Sec 20:502. doi https://doi.org/10.1007/978-3-642-56927-2

Lacasse S, Nadim F (2009) Landslide risk assessment and mitigation strategy. In: Sassa K, Canuti P (eds) Landslides – disaster risk reduction, vol 3. Springer Berlin, Heidelberg, pp 31–61. https://doi.org/10.1007/978-3-540-69970-5_3

Lanzafame G, Bousquet JC (1997) The Maltese escarpment and its extension from Mt. Etna to the Aeolian Islands (Sicily): importance and evolution of a lithosphere discontinuity. Acta Vulcanol 9:113–120

Locardi E, Nappi G (1979) Tettonica e vulcanismo recente nell'isola di Lipari (implicazioni geodinamiche). Boll Soc Geol It 98:447–456

Lucchi F (2013) Chapter 5 Stratigraphic methodology for the geological mapping of volcanic areas: insights from the Aeolian Archipelago (southern Italy). In: Lucchi F, Peccerillo A, Keller J, Tranne CA, Rossi PL (eds) The Aeolian Islands volcanoes, Memoir 37. Geological Society, London, pp 37–53. https://doi.org/10.1144/m37.5

Malamud BD, Turcotte DL, Guzzetti F, Reichenbach P (2004) Landslide inventories and their statistical properties. Earth Surf Proc Land 29(6):687–711. https://doi.org/10.1002/esp.1064

Manning CD, Schütze H (1999) Foundations of statistical natural language processing, Massachusetts Institute of Technology. The MIT Press, Cambridge

Marsella MA, Scifoni S, Coltelli M, Proietti C (2011) Quantitative analysis of the 1981 and 2001 Etna flank eruptions: a contribution for future hazard evaluation and mitigation. Ann Geophys 54(5):492–498. https://doi.org/10.4401/ag-5334

Marsella M., Salino A., Scifoni S., Sonnessa A., Tommasi P. (2013) Stability Conditions and Evaluation of the Runout of a Potential Landslide at the Northern Flank of La Fossa Active Volcano, Italy. In: Margottini C., Canuti P., Sassa K. (eds) Landslide Science and Practice. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-31310-3_4

Marsella M, D'Aranno PJV, Scifoni S, Sonnessa A, Corsetti M (2015a) Terrestrial laser scanning survey in support of unstable slopes analysis: the case of Vulcano Island (Italy). Nat Hazards 78(1):443–459. https://doi.org/10.1007/s11069-015-1729-3

Marsella MD, Aranno PJV, Scifoni S, Sonnessa A, Corsetti M (2015b) Terrestrial laser scanning survey in support of unstable slopes analysis: the case of Vulcano Island (Italy). Nat Hazards 78:443–459. https://doi.org/10.1007/s11069-015-1729-3

Mazzuoli R, Tortorici L, Ventura G (1995) Oblique rifting in Salina, Lipari and Vulcano Islands (Aeolian Islands, southern Tyrrhenian Sea, Italy). Terra Nova 7:444–452

McCulloch WS, Pitts W (1943) A logical calculus of the ideas immanent in nervous activity. Bull Math Biophys 5(4):115–133. https://doi.org/10.1007/BF02478259

Melchiorre C, Matteucci M, Azzoni A, Zanchi A (2008) Artificial neural networks and cluster analysis in landslide susceptibility zonation. Geomorphology 94(3):379–400. https://doi.org/10.1016/j.geomorph.2006.10.035

Meten M, PrakashBhandary N, Yatabe R (2015) Effect of landslide factor combinations on the prediction accuracy of landslide susceptibility maps in the Blue Nile Gorge of Central Ethiopia. Geoenviron Disasters 2(1):9. https://doi.org/10.1186/s40677-015- 0016-7

Morsdorf F, Mårell A, Koetz B, Cassagne N, Pimont F, Rigolot E, Allgöwer B (2010) Discrimination of vegetation strata in a multi-layered Mediterranean forest ecosystem using height and intensity information derived from airborne laser scanning. Remote Sens Environ 114(7):1403–1415. https://doi.org/10.1016/j.rse.2010.01.023

Nedjah N, de Macedo Mourelle L (2007) Reconfigurable hardware for neural networks: binary versus stochastic. Neural Comput & Applic 16(3):249–255. https://doi.org/10.1007/s00521-007-0086-x

Palenzuela JA (2018) Landslide-susceptibility assessment in an area with similar geoenvironmental conditions to that of the analysis area- implementing computer tools. http://www.ugr.es/~jpalbae/Apps.html.

Panizza M (1996) 3 Geomorphological hazard. In: Panizza M (ed) Developments in earth surface processes. Elsevier, Berlin, Vol. 4, pp 75–76. https://doi.org/10.1016/S0928-2025(96)80020-4

Pavel M, Fannin RJ, Nelson JD (2008) Replication of a terrain stability mapping using an artificial neural network. Geomorphology 97(3):356–373. https://doi.org/10.1016/j.geomorph.2007.08.012

Pavel M, Nelson JD, Jonathan Fannin R (2011) An analysis of landslide susceptibility zonation using a subjective geomorphic mapping and existing landslides. Comput Geosci 37(4):554–566. https://doi.org/10.1016/j.cageo.2010.10.006

Protodyakonov MM (1962) Mechanical properties and drillability of rocks. Dissertation, Symposium on Rock Mech., Minnessota. Univ 103–118 Rahmati O, Haghizadeh A, Pourghasemi HR, Noormohamadi F (2016) Gully erosion susceptibility mapping: the role of GIS-based bivariate statistical models and their comparison. Nat Hazards 82(2):1231–1258. https://doi.org/10.1007/s11069-016-2239-7

Scifoni S, Palenzuela Baena JA, Marsella M, Pepe S, Sansosti E, Solaro G, Tizzani P (2015) An integrated remote sensing approach for landslide susceptibly mapping at the volcanic islands of Vulcano and Lipari (Eolian Island, Italy). In Notarnicola C, Paloscia S, Pierdicca Nazzareno (eds) SPIE Proceedings Vol. 9642, SAR Image Analysis, Modeling, and Techniques XV, 96420H. 21-24 September 2015. Toulouse, France. https://doi.org/10.1117/12.2195060

Soeters R, van Westen CJ (1996) Slope instability recognition, analysis and zonation. In: Turner AK, Schuster RL (eds) Landslides investigation and mitigation. TRB Special Report 247. National Academy Press, Washington D.C, pp 129–177

Song JH, Han SH, Yu K, Kim YL (2002) Assessing the possibility of land-cover classification using LIDAR intensity data. In Kalliany R, Leberl F, Fraundorfer F (eds) ISPRS Commission III Proceedings, Photogrammetric Computer Vision. 9-13 September 2002, Graz, Austria. 34(3B):259–262

Thouret JC (1999) Volcanic geomorphology—an overview. Earth Sci Rev 47(1):95–131. https://doi.org/10.1016/S0012-8252(99)00014-8

TIBCO (2017) TIBCO Statistica TM. https://www.tibco.com/es/products/tibco-statistica.

Tinti S, Bortolucci E, Armigliato A (1999) Numerical simulation of the landslide-induced tsunami of 1988 on Vulcano Island, Italy. Bull Volcanol 61(1):121–137. https://doi.org/10.1007/s004450050

Tommasi P, Boldini D, Cignitti F, Graziani A, Lombardi A, Rotonda T (2007) Geomechanical analysis of the instability phenomena at Stromboli volcano. In: Eberhardt E, Stead D, Morrison T (eds) Dissertation, Proceedings of the 1st Canada-US Rock Mechanics Symposium - Rock Mechanics Meeting Society's Challenges and Demands. 2:933–941

Tranne CA, Lucchi FC, Calanchi N, Lanzafame G, Rossi PL (2002) Geological map of the island of Lipari (Aeolian Islands). University of Bologna and INGV. L.A.C, Firenze Van Westen CJ, Rengers N, Soeters R (2003) Use of geomorphological information in indirect landslide susceptibility assessment. Nat Hazards 30(3):399–419. https://doi.org/10.1023/B:NHAZ.0000007097.42735.9e

Varnes DJ (1984) Landslide hazard zonation: a review of principles and practice. Nat Hazards 3:61. Commision on Landslides of IAEG, UNESCO. United Nations Educational, Scientific and Cultural Organization. Paris

Ventura G (1994) Tectonics, structural evolution and caldera formation on Vulcano Island (Aeolian Archipelago, southern Tyrrhenian Sea). J Volcanol Geotherm Res 60(3):207–224. https://doi.org/10.1016/0377-0273(94)90052-3

Ventura G (1995) Relationships between tectonics and volcanism in the central and eastern sectors of the Aeolian Islands. Dissertation, Proceedings of the National Congress. Gruppo Naz. Geofis. Terra Solida-Cons. Naz. delle Ric. Rome

Ventura G (2013) Chapter 2 Kinematics of the aeolian volcanism (southern Tyrrhenian sea) from geophysical and geological data. Geol Soc Mem 37:3–11. https://doi.org/10.1144/m37.2

Ventura G, Vilardo G, Milano G, Pino NA (1999) Relationships among crustal structure, volcanism and strike–slip tectonics in the Lipari–Vulcano volcanic complex (Aeolian Islands, southern Tyrrhenian Sea, Italy). Phys Earth Planet Int 116(1):31–52. https://doi.org/10.1016/S0031-9201(99)00117-X

Wan S (2009) A spatial decision support system for extracting the core factors and thresholds for landslide susceptibility map. Eng Geol 108(3):237–251. https://doi.org/10.1016/j.enggeo.2009.06.014

Xiao H, Huang J, Ma Q, Wan J, Li L, Peng Q, Rezaeimalek S (2017) Experimental study on the soil mixture to promote vegetation for slope protection and landslide prevention. Landslides 14(1):287–297. https://doi.org/10.1007/s10346-015-0634-x

List of tables:

List of figures:

Table 1

N	6	9	10	12	16	18	20	21	24
$(H \times W)$	2×3	3×3	2×5	3×4	4×4	3×6	4×5	3×7	3×8

Table 2

Classes	χ^2 Pearson	C	Cmax.	Adjusted C (% C/Cmax)
11	182,699.47	0.07	0.71	9.8
17	320,580.58	0.09	0.71	13.1

Table 3

Factor	Classes	χ^2 Pearson	C	Cmax.	Adjusted C (% C/Cmax)
LiDAR intensity	4	13,854.17	0,02	0.71	2.7.
Distance to C.C.	4	75,809.71	0,04	0.71	6.4
Aspect	6	97,210.16	0,05	0.71	7.2
Land use	8	131,642.53	0,06	0.71	8.5
Lithology	17	320,580.58	0,09	0.71	13.0
Elevation	5	537,698.83	0,12	0.71	16.9
Slope	5	537,698.83	0,12	0.71	16.8

Table 4

Try	Included factors Slope	Elevation	Lithology	Land use	Aspect	C.C. dist	LiDAR int.
1	X	X	X	X	X	X	X
2	X	X	X	X	X	X	
3	X	X	X	X	X		
4	X	X	X	X		X	
5	X	X	X		X	X	
6	X	X		X	X	X	
7	X		X	X	X	X	
8		X	X	X	X	X	
9	X	X	X	X	X		X
10	X	X	X	X			X
11	X	X	X		X		X
12	X	X		X	X		X
13	X		X	X	X		X
14		X	X	X	X		X

Table 5

Try	DOF % Very Low	Low	Moderate	High	Very High	1. Very Low and Low	2. High and Very High	2–1
1	6.25	8.59	9.47	19.05	56.64	14.84	75.69	60.85
2	5.67	8.72	8.72	3.88	73.01	14.39	76.89	6.49
3	5.68	11.49	20.98	17.86	43.99	17.17	61.85	44.68
4	10.45	11.68	18.71	28.38	30.78	22.13	59.16	37.03
5	7.35	14.18	15.18	8.99	54.29	21.53	63.28	41.75
6	10.06	15.21	19.23	14.21	41.29	25.27	55.50	30.23
7	7.45	12.03	14.61	14.91	51.01	19.48	65.92	46.44
8	6.38	7.62	10.13	7.13	68.75	13.99	75.88	61.89
9	5.66	11.80	14.83	22.17	45.54	17.45	67.71	50.26
10	6.05	12.78	17.10	15.86	48.21	18.83	64.07	45.23
11	4.33	9.25	15.39	17.38	53.65	13.58	71.03	57.45
12	5.66	11.80	14.83	22.17	45.54	17.45	67.71	50.26
13	4.94	13.00	20.01	23.56	38.48	17.94	62.05	44.10
14	6.06	10.69	12.93	16.84	53.47	16.75	70.32	53.57

Figure 1

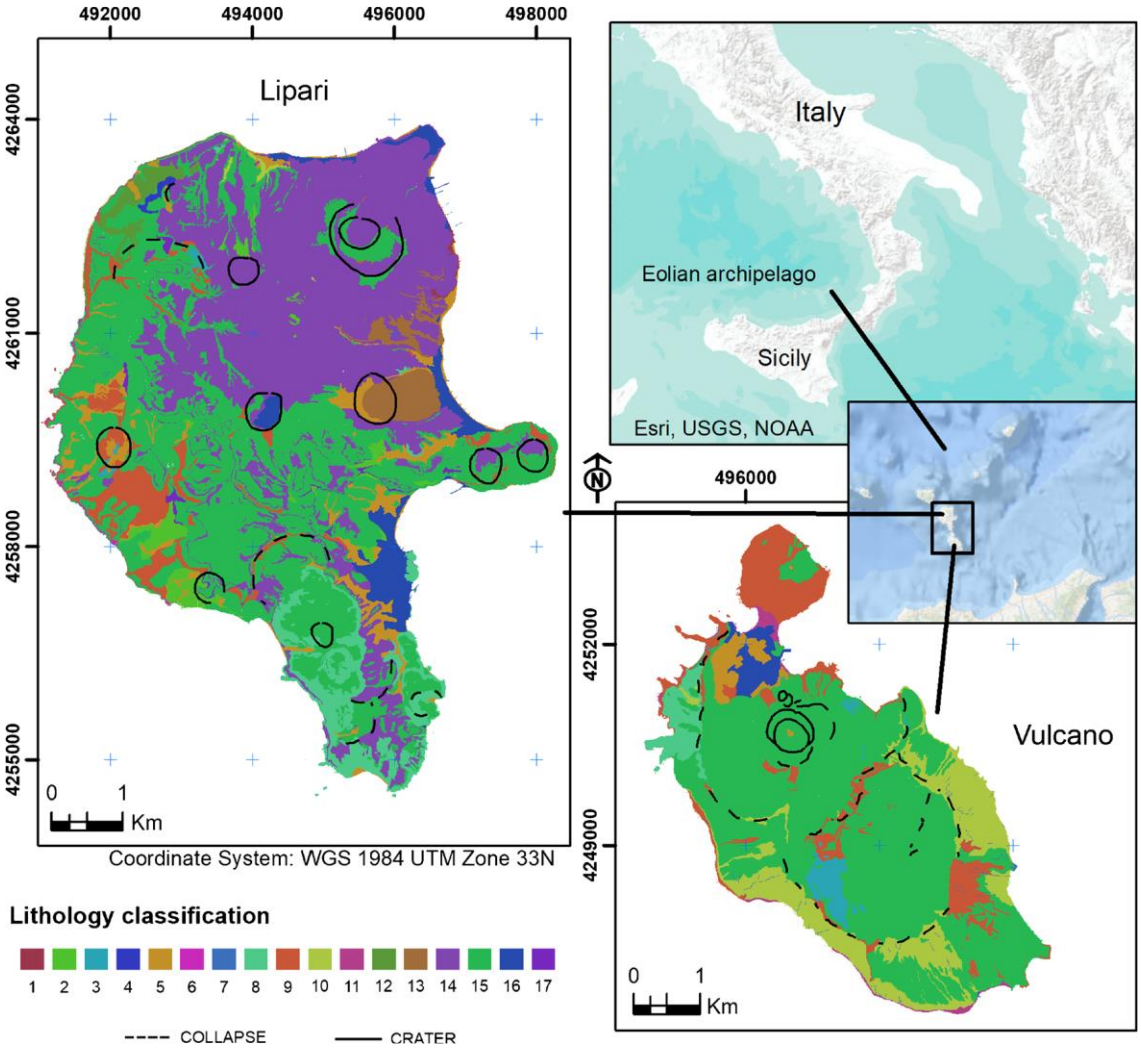

492000
494000
496000
498000
Lipari
4264000
4261000
4258000
4255000
0
1
Km
Coordinate System: WGS 1984 UTM Zone 33N
Italy
Eolian archipelago
Sicily
Esri, USGS, NOAA
N
496000
4252000
4249000
Vulcano
0
1
Km
Lithology classification
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17
COLLAPSE
CRATER

Figure 2

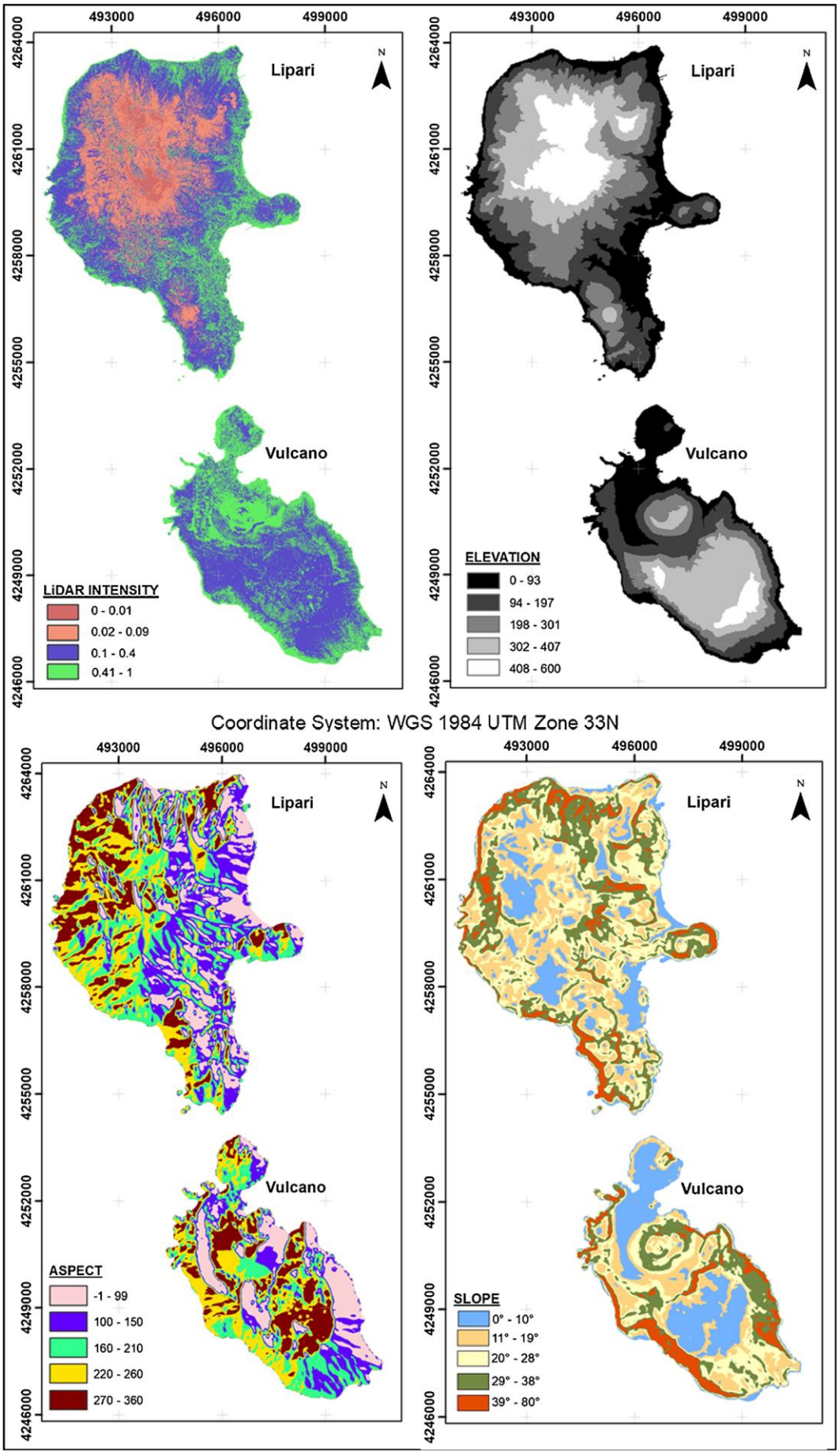

493000
496000
499000
4264000
4261000
4258000
4255000
4252000
4249000
4246000
N
Lipari
Vulcano
LiDAR INTENSITY
0 - 0.01
0.02 - 0.09
0.1 - 0.4
0.41 - 1
ELEVATION
0 - 93
94 - 197
198 - 301
302 - 407
408 - 600
Coordinate System: WGS 1984 UTM Zone 33N
ASPECT
-1 - 99
100 - 150
160 - 210
220 - 260
270 - 360
SLOPE
0° - 10°
11° - 19°
20° - 28°
29° - 38°
39° - 80°

Figure 3

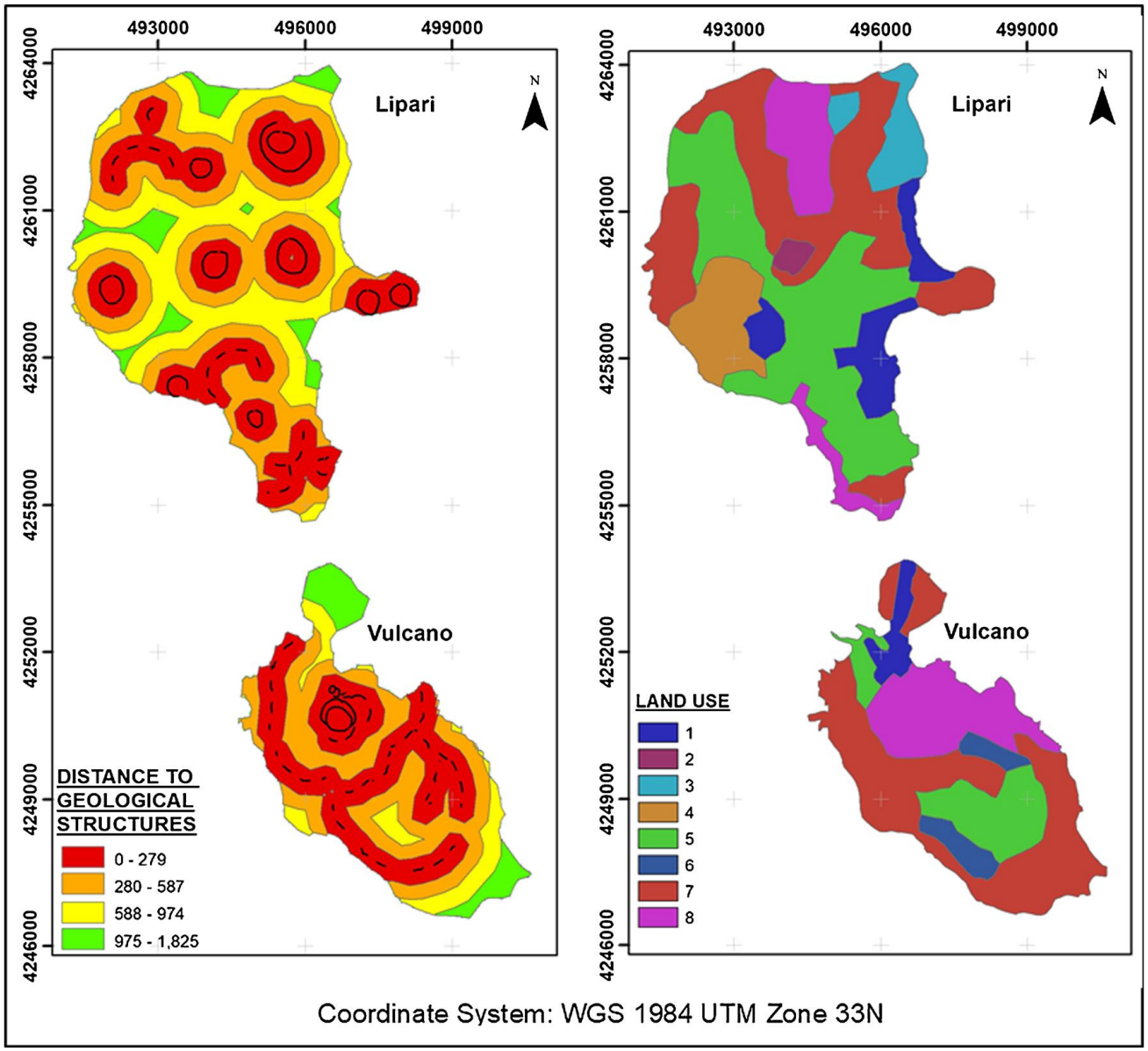

493000
496000
499000
4264000
4261000
4258000
4255000
4252000
4249000
4246000
N
Lipari
Vulcano
DISTANCE TO GEOLOGICAL STRUCTURES
0 - 279
280 - 587
588 - 974
975 - 1,825
LAND USE
1
2
3
4
5
6
7
8
Coordinate System: WGS 1984 UTM Zone 33N

Figure 4

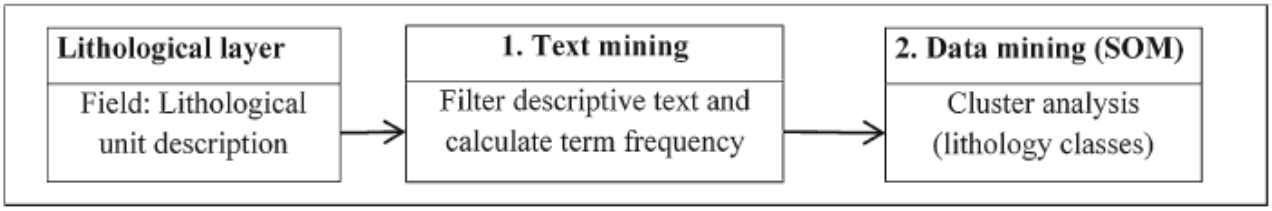

Lithological layer
Field: Lithological unit description
1. Text mining
Filter descriptive text and calculate term frequency
2. Data mining (SOM)
Cluster analysis (lithology classes)

Figure 5

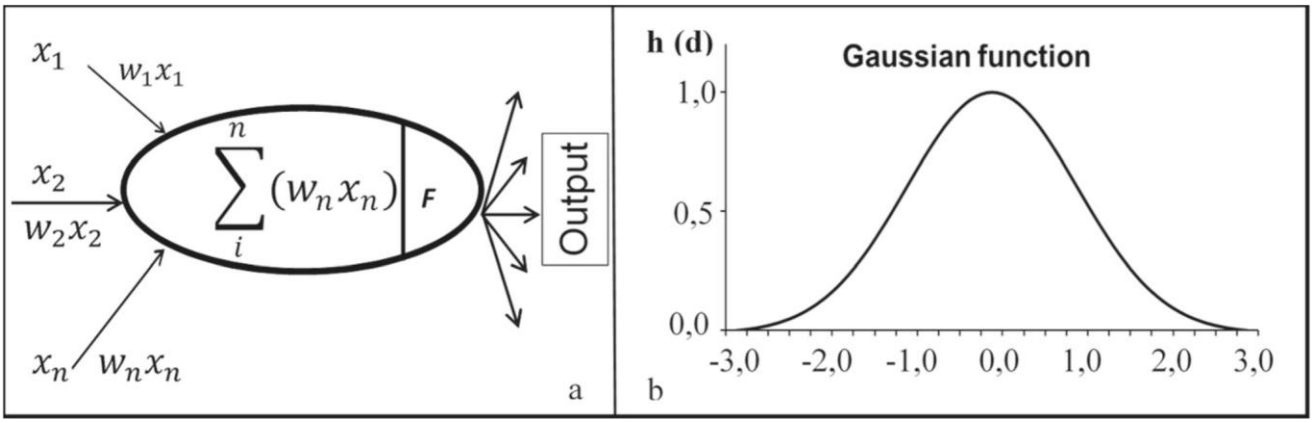

x_1
w_1x_1
x_2
w_2x_2
x_n
w_nx_n
$\sum_i^n (w_nx_n)$
F
Output
a
h (d)
Gaussian function
1,0
0,5
0,0
-3,0
-2,0
-1,0
0,0
1,0
2,0
3,0
b

Figure 6

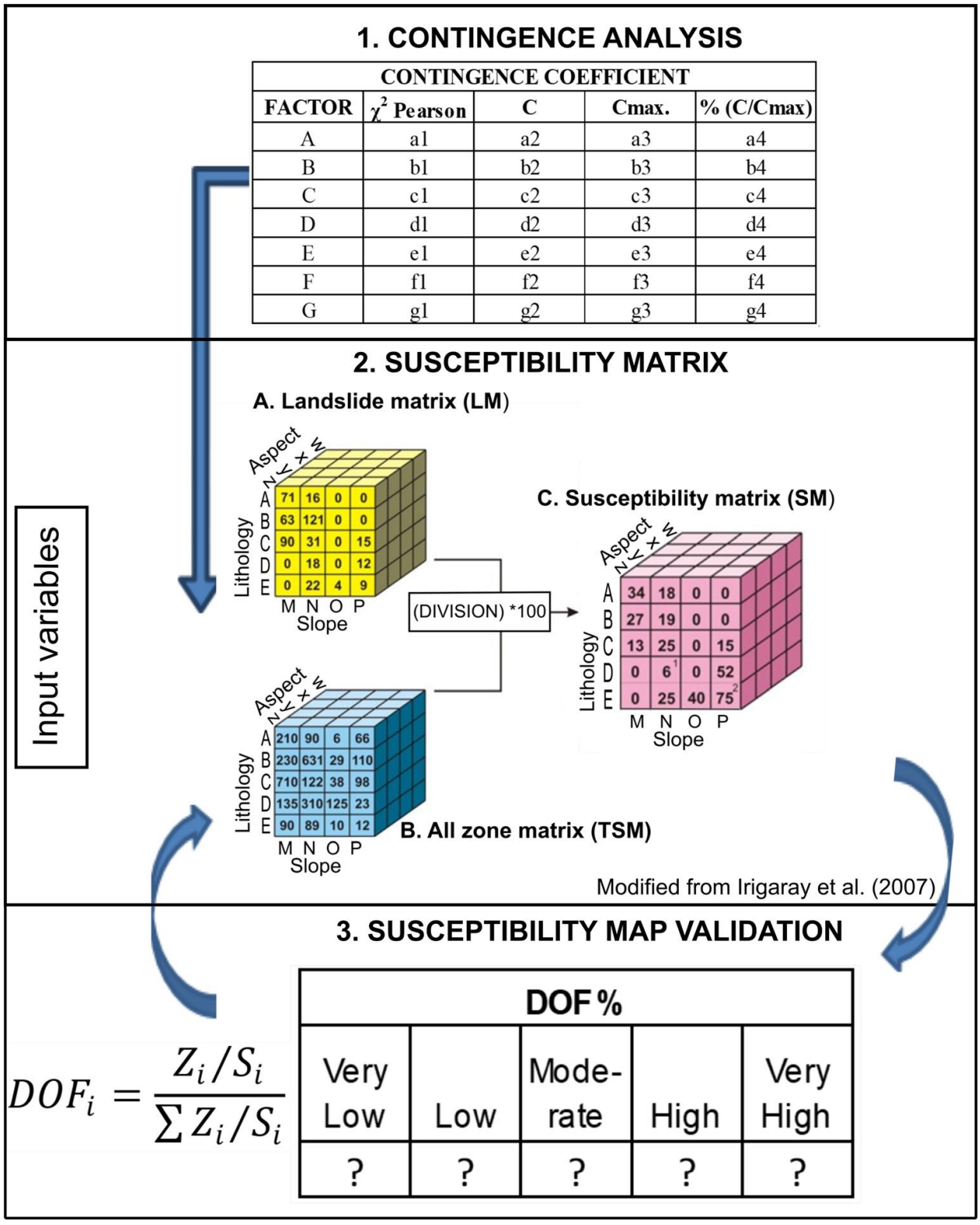

1. CONTINGENCE ANALYSIS
CONTINGENCE COEFFICIENT
FACTOR | χ² Pearson | C | Cmax. | % (C/Cmax)
A a1 a2 a3 a4
B b1 b2 b3 b4
C c1 c2 c3 c4
D d1 d2 d3 d4
E e1 e2 e3 e4
F f1 f2 f3 f4
G g1 g2 g3 g4
2. SUSCEPTIBILITY MATRIX
A. Landslide matrix (LM)
Aspect
N NW W SW S SE E NE
Lithology
A 71 16 0 0
B 63 121 0 0
C 90 31 0 15
D 0 18 0 12
E 0 22 4 9
M N O P
Slope
(DIVISION) *100
C. Susceptibility matrix (SM)
A 34 18 0 0
B 27 19 0 0
C 13 25 0 15
D 0 6 0 52
E 0 25 40 75
B. All zone matrix (TSM)
A 210 90 6 66
B 230 631 29 110
C 710 122 38 98
D 135 310 125 23
E 90 89 10 12
Input variables
Modified from Irigaray et al. (2007)
3. SUSCEPTIBILITY MAP VALIDATION
$DOF_i = \frac{Z_i/S_i}{\sum Z_i/S_i}$
DOF %
Very Low | Low | Mode-rate | High | Very High
? ? ? ? ?

Figure 7

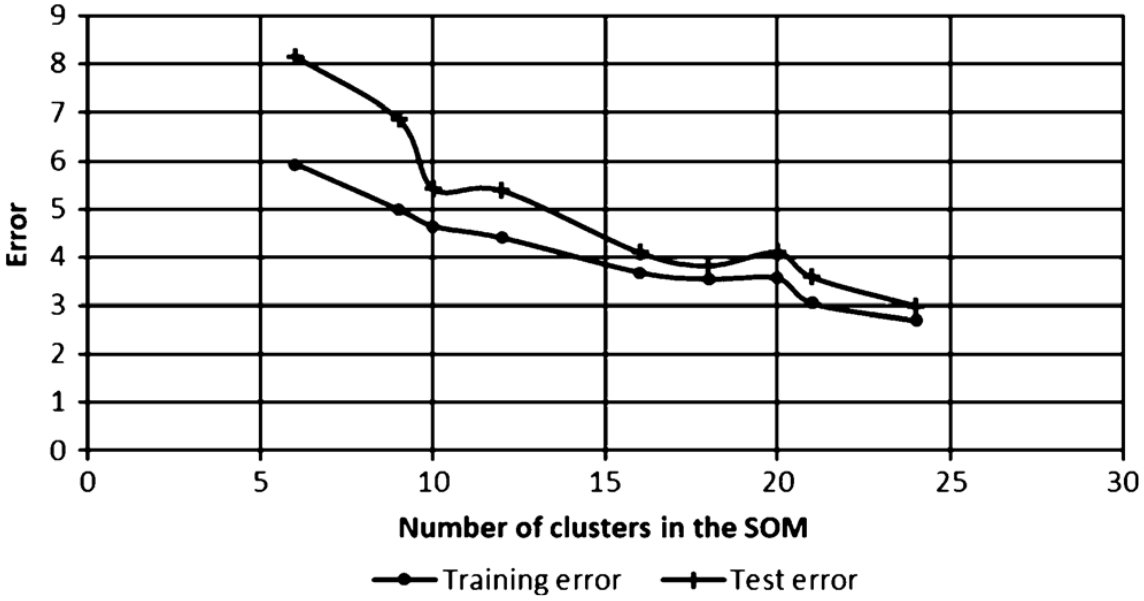

Error
9
8
7
6
5
4
3
2
1
0
0
5
10
15
20
25
30
Number of clusters in the SOM
Training error
Test error

Figure 8

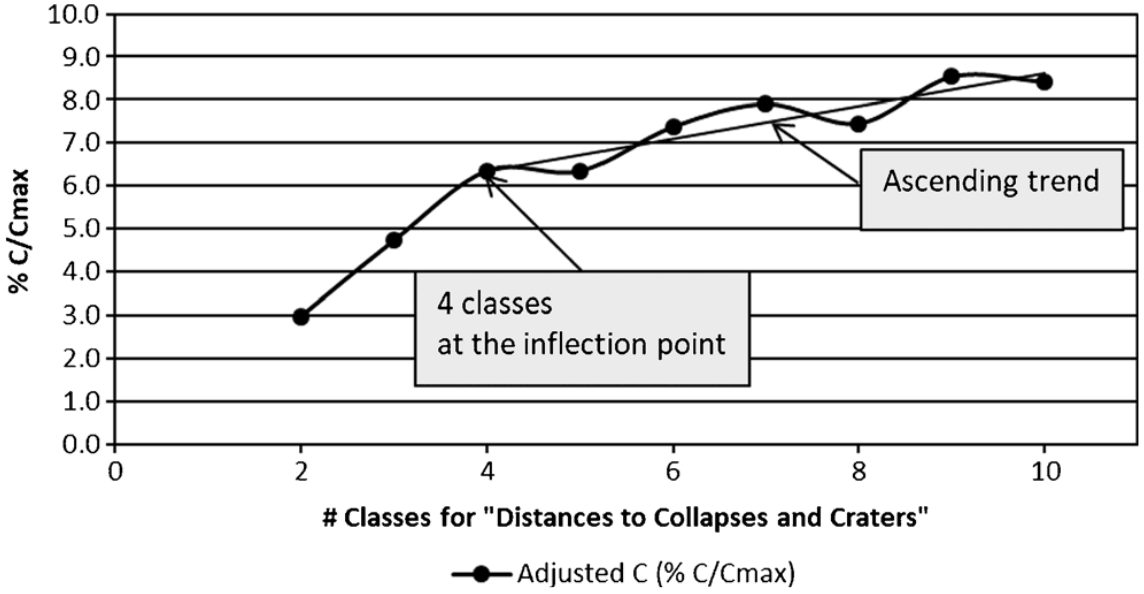
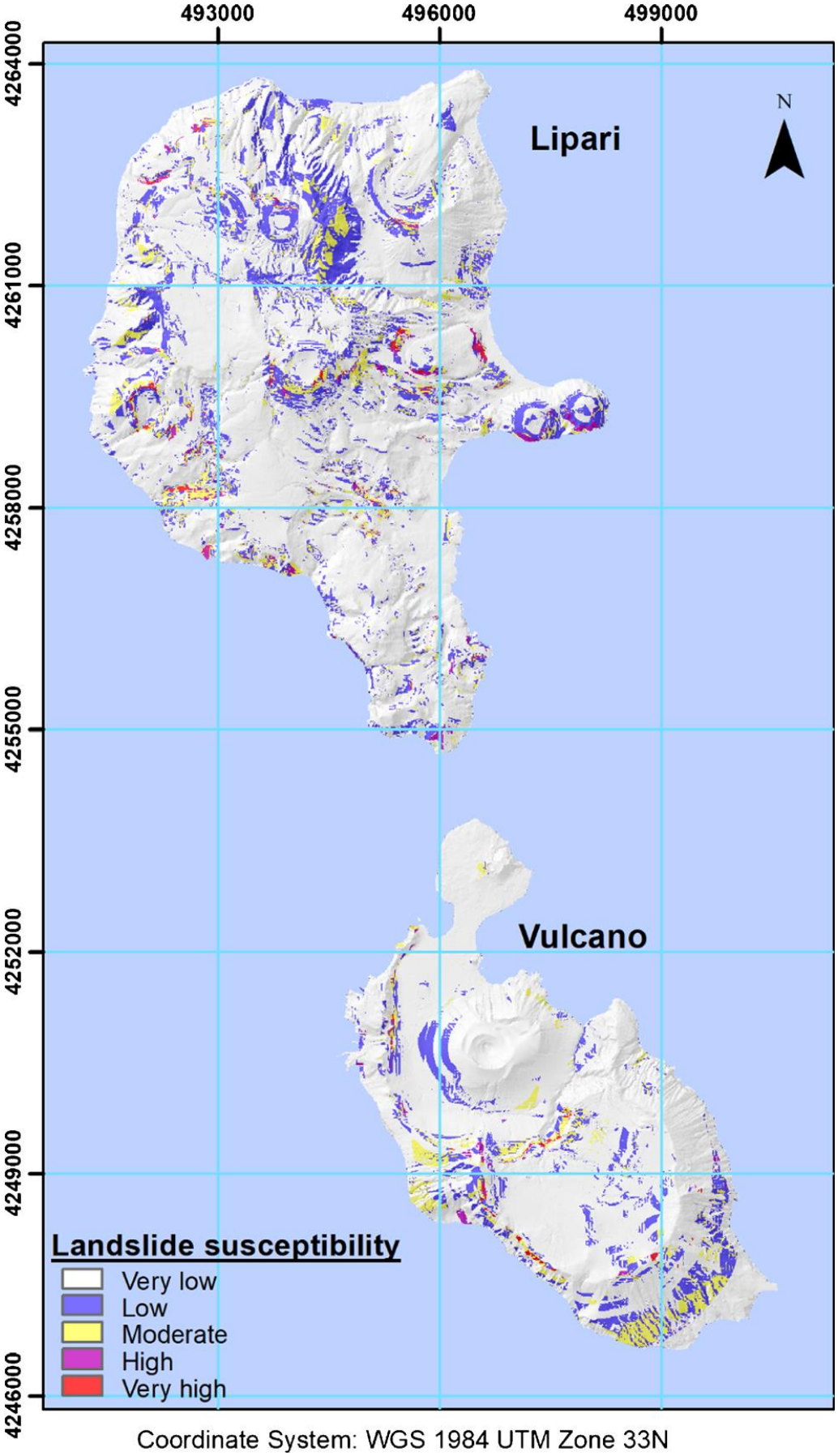

10.0
9.0
8.0
7.0
6.0
5.0
4.0
3.0
2.0
1.0
0.0
% C/Cmax
0
2
4
6
8
10
Classes for "Distances to Collapses and Craters"
Adjusted C (% C/Cmax)
Ascending trend
4 classes
at the inflection point

Figure 9


493000
496000
499000
4264000
4261000
4258000
4255000
4252000
4249000
4246000
N
Lipari
Vulcano
Landslide susceptibility
Very low
Low
Moderate
High
Very high
Coordinate System: WGS 1984 UTM Zone 33N