Python plugin for statistical analysis of landslide susceptibility over wide areas

Paola Salmona[1], Rossella Bovolenta[1], Bianca Federici[1], Ilaria Ferrando[1]

[1] Department of Civil, Chemical and Environmental Engineering, University of Genoa, via Montallegro 1, 16145 Genoa, Italy
(paola.salmona, rossella.bovolenta, bianca.federici)@unige.it; ilaria.ferrando@edu.unige.it

Keywords: GRASS GIS, Ground Instability, Open Data, Logistic Regression, Multivariate Statistics, Nationwide geospatial data.

Abstract

As part of the RETURN project, a model for assessing landslide susceptibility through logistic regression within a GIS environment has been developed, aimed at supporting public authorities and professionals in managing ground instability risks. The model utilizes freely accessible national-scale datasets to ensure high transferability and transparency of results. The analysis is implemented in Python and integrated into GRASS GIS, with the objective of automating the workflow and making the procedure accessible even to non-expert users. The methodology was tested in the Province of Savona (Liguria, Italy), using eight predisposing factors and landslide data from the IFFI inventory. The results demonstrated reliability exceeding 75% in most cases. The resulting susceptibility maps are reclassified into three qualitative categories—low, medium, and high susceptibility—to improve interpretability. The case study highlighted both the strengths and limitations of the approach, notably the need to standardize data and procedures to ensure applicability at broader scales. Ongoing development efforts are focused on enhancing the identification of relevant factors and minimizing subjectivity in data preparation. The automation of the model paves the way for extensive testing across different areas and geomorphological settings, contributing to the development of a robust tool for landslide risk management.

1. Introduction

Within the Piano Nazionale di Ripresa e Resilienza (PNRR, literally National Recovery and Resilience Plan) Extended Partnership (EP) "Multi-Risk sciEnce for resilienT commUnities undeR a changiNg climate" (RETURN), the research group of the Department of Civil, Chemical and Environmental Engineering (DICCA) of the University of Genoa is developing a system for processing landslide susceptibility maps in Geographic Information System (GIS) environment. The expected result is a tool meant to be used by administrations and local authorities for ground instability assessment and management. Therefore, high usability is required, that implies free and easily available input data, readily interpretable results, and clearly explained limitations of use and reliability levels.
To accomplish this objective, some stakes have been established:

- flows, slides, areas subject to diffuse shallow landslides and those landslides classified as "undetermined" and "complex" were considered for the processing of susceptibility maps. It does not apply to falls, topples and areas subject to diffuse falls and topples and Deep-seated Gravitational Slope Deformations.
- the analysis method is the logistic regression, which is widely used and documented in scientific literature (Chang et al., 2019; Chowdhury et al., 2024; Goyes-Peñafiel et al., 2021; Kadavi et al., 2019; Lin et al., 2017; Süzen et al, 2011; Zhao et al., 2019);
- the proposed tool should be scalable and transferable;
- minimum "guaranteed" reliability threshold is imposed. Tests are undergoing in several areas to verify that a reliability of at least 75% is achieved in the calibration phase. Alternatively, the conditions contributing to lower outcomes should be identified, and potential corrective measures should be evaluated;
- to facilitate the usability of the procedure for users with limited GIS experience and to ensure the accurate implementation of the planned operations, the entire procedure is currently being developed in Python (Python Software Foundation, 2025), initially within the GRASS GIS environment (GRASS Development Team, 2025), and potentially as a standalone tool in the future.

2. Context: the province of Savona (Italy) case study

The work builds upon an already tested methodology of statistical analysis within a GIS environment, based on logistic regression (Bovolenta et al., 2016; Marzocchi et al. 2015). The aim of this study is to adapt it to the current version of GRASS GIS (ver. 8.4), to streamline and standardize the individual steps, and to optimize the methodology for large-scale applications.
The procedure was tested on the province of Savona (Italy), an area of about 1000 km^2, considering the pixel as the minimum spatial unit, with a nominal scale of 1:100,000 and a raster resolution of 20 m.
Eight predisposing factors were chosen: elevation, slope, exposure, water accumulation, land use/land cover, lithology and rainfall influence. From the literature review, several other factors have been considered in logistic regression procedures (Chaohai et al., 2025; Liu et al., 2025; Salmona et al., 2025; Wu et al., 2023; Yang et al., 2024). However, the decision was made to start with the mentioned variables, as they are independent from one another and openly available across the entire Italian territory, either directly at the national scale or provided by individual Regions.
The first phase consisted of pre-processing and discretization of the base data, allowing general data quality checks and to reduce the possible combinations of factors to a manageable number. The eight considered factors were then brought into raster format at the set resolution and divided into qualitative (e.g., land use type) or ordinal (e.g., elevation intervals, from 0 to maximum elevation) classes.
The resulting maps were compared in a bivariate analysis with the Inventory of Landslide Phenomena in Italy (IFFI) (ISPRA, 2025a), and the classes of each factor were reordered based on conditional probability. Then the factors were related to each other and to the presence of landslides in a multivariate analysis by logistic regression, defining the probability values.

This procedure was applied both by considering all landslide areas as the statistical sample and by analyzing each landslide mechanism individually. The tests conducted to date have shown that each type of landslide is affected differently by the considered factors, and the model is more reliable if the different types of slope failure mechanisms are treated separately.
The procedure described above resulted in an Area Under the Curve (AUC) in calibration generally above the preset threshold of 75%. Diffuse shallow landslides and complex landslides remain below the threshold, although they still exceed 73%.
The resulting maps report probability values aggregated into three qualitative classes to indicate the high, medium and low levels of landslide susceptibility.

2.1 Lessons learnt from the case study

The procedure proved to be relatively straightforward and suitable for general screening purposes; however, it is undoubtedly time-consuming and prone to coarse errors, making it poorly suited for large-scale applications in its current form.
The area selected for the case study, identified on an administrative basis, proved to be adequate in terms of spatial extent, but it encompasses two environments that differ significantly in geomorphological, climatic, and anthropogenic characteristics. A delineation based on physical boundaries could result in more homogeneous conditions and a less noisy context; however, it could also present challenges in acquiring base data, as these may fall under the jurisdiction of different administrative entities.
While the quality of input data remains crucial, a minimum dataset must be defined to ensure the broad applicability of the proposed procedure and facilitating comparisons across diverse spatial and temporal contexts. Such data should be open, of high quality (or at least possess a clearly defined level of reliability), and available not only for a single case study but, at a minimum, on a national scale, as required by the RETURN project.
Moreover, even when identical datasets and clear guidelines for applying the proposed methodology are provided, multiple stages remain at which a non-specialist operator may face technical challenges or be compelled to make subjective decisions. In this regard, the procedure should be considered a continuous work in progress, and it must be accompanied by detailed documentation, allowing for ongoing integration of improvements or adjustments as new needs arise.

2.2 The Python script

In response to this issue, the mode, initially developed within the GRASS GIS environment, is being rewritten as a Python script, not only to enable more efficient application, thus reducing time consumption and minimizing the risk of gross errors, but also to establish a standardized procedure that ensures accessibility for users outside the research team and facilitates the production of comparable results. The presented tool includes a predominantly automated "basic version", optimized for the use of a minimum dataset consisting of openly available data covering the entire Italian territory, to describe predisposing factors, and using landslide areas reported in the IFFI inventory as the statistical sample.
Alternatively, the tool can be customized by introducing additional factors not included in the initial tests or by employing site-specific datasets (e.g., mapped landslides or land use derived from remote sensing). In this case, the initial phases of the procedure, namely, the data preparation, are not automated but are left to the operator, following more stringent specifications.

2.3 Considerations about data

The required data are listed in the themes outlined in Annexes 1, 2, and 3 of INSPIRE directive (European Parliament, 2007). However, an analysis of the nationally available datasets revealed issues of insufficient quality; consequently, the same types of data were sourced from regional authorities.
In the following, some recommendations for the selection of reference data are reported.

2.3.1 Study area extent: It defines the boundary of the domain under analysis. It is advisable to start with a vector dataset, which will be used as a mask for all subsequent raster maps. While using an administrative boundary may be simpler for obtaining data distributed by local authorities, it is preferable to work with an area that does not exhibit highly divergent features (e.g., two slopes separated by a main watershed or contrasting regions such as alpine areas and plains). In this regard, a physical delineation, such as a hydrological basin or a large slope, may be preferable.
From the experiments conducted to date, a scale of 1:100,000 with a 20 m resolution for raster data resulted to be a good compromise, considering areas around 1000 km².

2.3.2 Digital Terrain Model (DTM): It represents the elevation of an area. Based on the resolution of the DTM, the resolution of all other maps is defined, including those derived directly from it (slope, aspect, and water accumulation), as well as those created independently. Regardless of its origin, it is important that the DTM is seamless and that it does not result from an upscaling operation, meaning it should not be resampled to increase its resolution.

2.3.3 Lithology: For the Italian context, Digital Lithological Map of Italy at 1:100,000 scale (Bucci et al., 2021; Bucci et al., 2022) was chosen. It classifies the different lithological types present in the Italian territory into 19 classes, based on compositional and geo-mechanical criteria. These classes are identified by a number and a brief description. 19 classes are many, but by clipping the map to the study area, some are usually excluded, and in the bivariate statistics section, others report 0 landslides and are therefore grouped as a single class. It is also possible to use other geological and lithological maps with different grouping methods. The simplest method, based on the origin of geological formations, results in 5 classes, but with noticeably less accurate results.

2.3.4 Land Use: Any land use/land cover map may be utilized; however, it is recommended that, in the case of large-scale and highly detailed nominal maps, land units be grouped according to soil cover rather than "economic" use. For instance, an olive grove may be classified as "agricultural" in terms of land use, but as "tree-covered" in terms of land cover. A reliable reference is the ESA WorldCoverMap v200 of 2021 (Zanaga et al., 2022), with a 10 m resolution, which identifies 11 classes. For instance, in the case study, the Land Use map of the Liguria Region (Regione Liguria 2019a, Regione Liguria 2019b), originally with 82 classes, was grouped into 8 classes according to the ESA WorldCoverMap v200 legend. Each class should also be assigned a progressive number, which will be used in the statistical analyses.

2.3.5 Distance from the road network: This factor is intended to evaluate the impact of slope modifications and road presence on the occurrence of landslide phenomena.

It is derived from the road and railway networks (excluding tunnels and bridge/viaduct sections) by applying progressively widening buffers around each segment, based on the assumption that beyond 200 meters, the influence of slope cuts diminishes substantially (Brenning et al., 2015). In the case of Italy, although national-level data are available, it may be preferable to use region-specific data if they are more up-to-date and accurate. If the data are at a more detailed scale (1:5000 or 1:10000), this does not significantly affect the results, as small curves and details are lost in the rasterization and buffering process, ultimately reaching a similar level of detail. It is worth emphasizing that, since the analysis approach takes the road network into account, the resulting mapping may be useful for the monitoring of these infrastructures.

2.3.6 Rainfall influence: It is quantified through Climatic Aggressiveness, an indicator that characterizes the degree to which an area is exposed to extreme rainfall events by relating average monthly and annual precipitation over an extended multi-year period. Climatic aggressiveness is calculated using the following equation (Arnoldus, 1980):

$$AC = \sum_{i=1}^{12} \frac{p_i^2}{P} \tag{1}$$

where p_i is the average rainfall for each month and P is the average annual rainfall. This parameter describes how the presence of drought periods and extreme rainfall events influence ground instability. To incorporate this factor into the model, monthly cumulative precipitation data for the study area over a minimum period of 30 years is required; if such data are unavailable, the longest possible time series should be used (21 years in the presented case study). In Italy, the "National System for the Collection, Processing, and Dissemination of Climate Data – SCIA" (ISPRA, 2025b) provides point-specific climate data collected by a network of monitoring stations distributed throughout the country. Alternatively, rainfall data collected by local operators (Regional authorities, municipalities, associations, private institutions, etc.) or continuous data (raster maps) from global sources can be used. In all cases, preprocessing is required. If the rainfall data is in point form, it must be interpolated to create continuous maps at the specified resolution. In the case of large-scale data, resampling to the desired resolution is likely necessary.

2.3.7 Landslides: In Italy, at the national level, the most comprehensive and up-to-date dataset regarding landslide phenomena is the IFFI inventory, which provides delineation and identification points for landslide events collected by various River Basin Authorities at a 1:25,000 scale. Each landslide event is associated with the type of landslide mechanism (as a number and a brief description) and, when available, other descriptive data. The data are organized by Italian Regions and are readily accessible and downloadable as shapefiles, even for non-expert users. Therefore, the data from this dataset were selected as a basic statistical sample, and the statistical analysis procedure was optimized based on their defining characteristics. Alternative datasets may also be employed, for instance, when locally collected data are available or for areas outside of Italy. In such cases, it is essential that the data be structured in a manner consistent with the IFFI dataset.

2.4 Workflow

To sequence the GRASS commands used in the procedure, the *grass.script* package was employed. Additionally, the *os*, *pandas*, *re* and *csv* packages were utilized.The work consists of eight parts, initially developed separately and subsequently integrated into a sequential workflow. The portion completed so far represents the simplest scenario, which can be customized, for instance, by using different datasets, and extended by incorporating additional predisposing factors or by applying alternative methods for input data classification. The code is available on: https://github.com/LabGeomatica/SALSA (Salmona, 2025). The subsequent Figure 1 and Figure 2 represent graphically the eight steps of the workflow, that are more extensively described in sections 2.4.1-2.4.8.

2.4.1 Working environment setting: The settings defined during this initial stage, specifically the working area, spatial resolution, and a designated folder for storing all files generated and used in subsequent phases, will be retained throughout the entire workflow. Once created a Location in a projected coordinates system, the first required input consists of a vector map of the study area and the specified spatial resolution. Based on this input, a mask is generated with the *r.mask* module, which is applied throughout the procedure to ensure that all input and output maps are consistently aligned with the same working area and resolution. Unfortunately, it is not always possible to clip maps representing the various predisposing factors with perfect spatial alignment, which may result in the presence of empty or undefined areas. This issue is particularly common along coastlines or administrative boundaries, where some maps do not precisely match the outline of the study area, leading to small gaps.

2.4.2 Predisposing factors maps: The maps representing the various predisposing factors are imported and clipped using the mask. To ensure consistency across input data, it is required that the extent and resolution of the input datasets match with the predefined working area and resolution. For raster data, a control script is implemented to enforce this requirement: by comparing the dataset extent (calculated with *r.info*) with the current region extent (calculated with *g.region*), the script prevents the import of data with a smaller extent or resolution, generating an error message and skipping the map. If the resolution of the input is higher than the specified one, the map is resampled accordingly. Elevation, slope, aspect, and water accumulation are all derived from the DTM, which is therefore the only required input for these factors. Processing is performed at this stage using the *r.slope.aspect* and *r.watershed* commands, ensuring that border cells are also included in the computation. Regarding the lithology layer, it is based on the Digital Lithological Map of Italy for the present version of the procedure, which is applicable exclusively to areas within the Italian territory. For land use/land cover vector maps, once it has been verified that the bounding box is at least as large as the defined working region (using *r.info* and *g.region*), the features are rasterized directly within the mask and at the specified resolution. As no default dataset has been identified so far, users are expected to prepare the map using available data, in accordance with the specifications outlined above. Concerning the proximity to surface transportation networks, the required data are vector graphs representing ground-level segments of both the road and railway networks. Once imported, these are rasterized, merged, and used to generate buffer zones.

https://doi.org/10.5194/isprs-archives-XLVIII-4-W13-2025-209-2025 | © Author(s) 2025. CC BY 4.0 License.

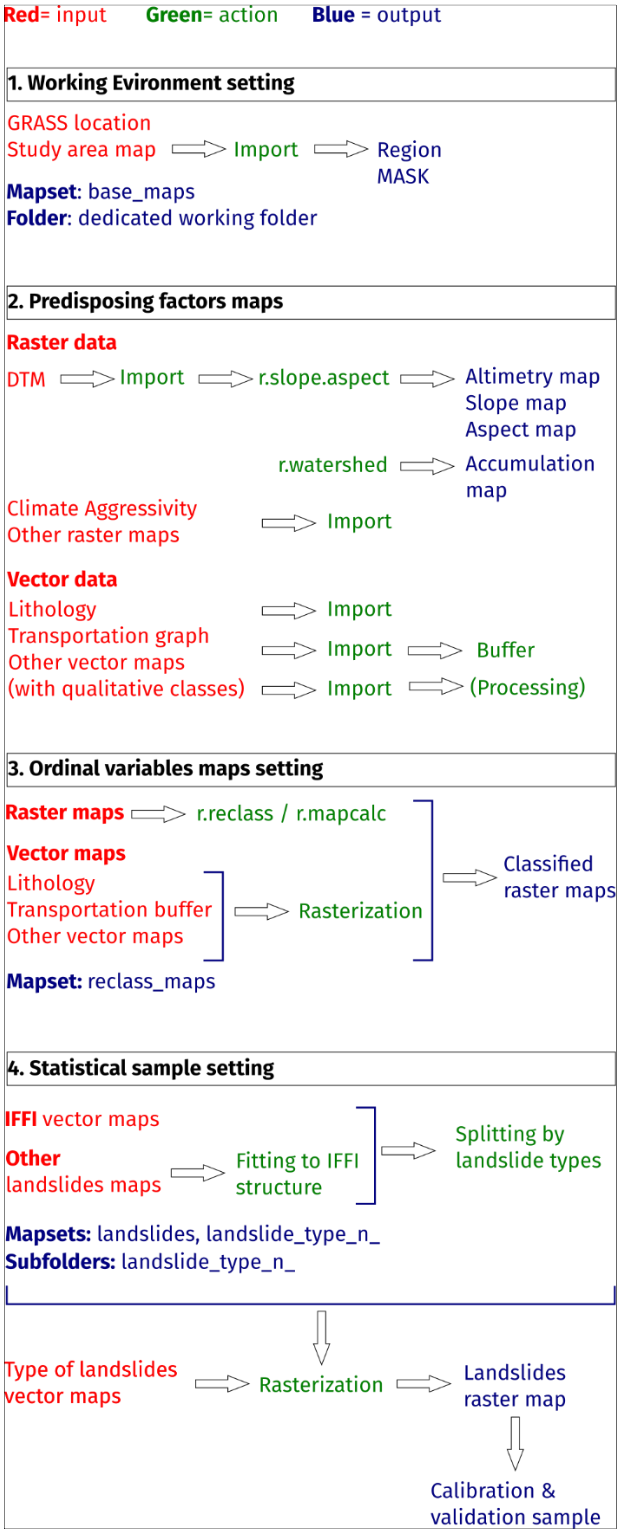

Figure 1. First phase of the workflow: (1) working environment setting, (2) import/creation of predisposing factors maps, (3) maps classification, (4) statistical sample setting.

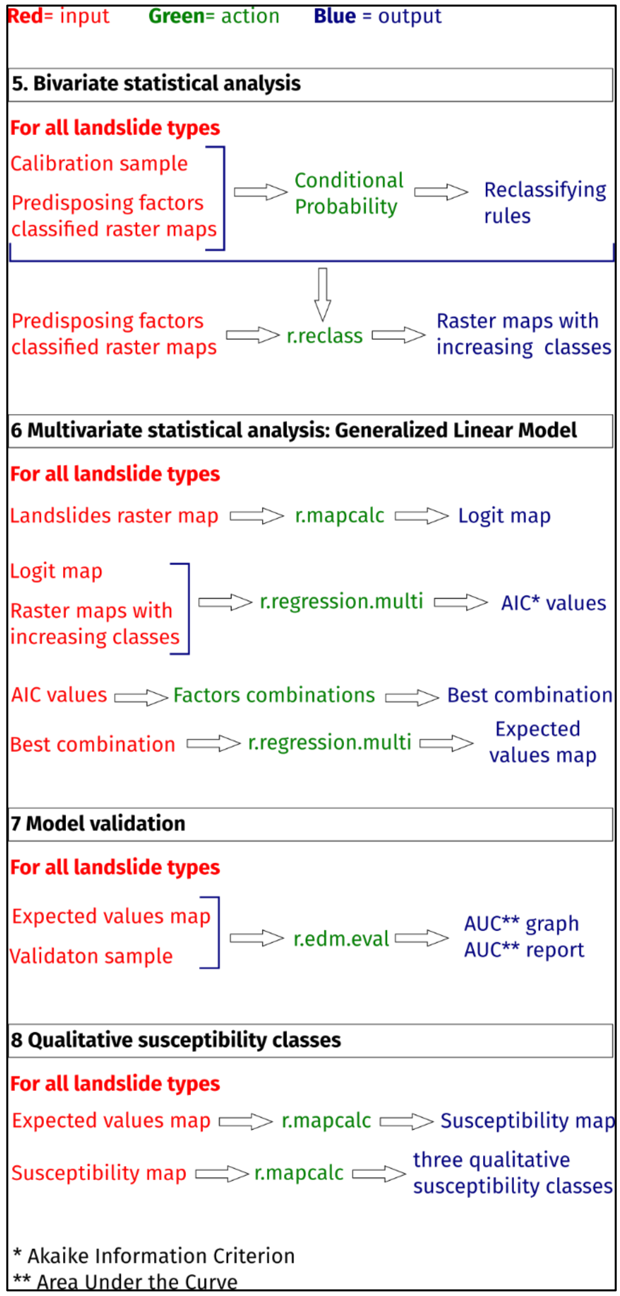

Figure 2. Second phase of the workflow: (5) bivariate statistical analysis, (6) multivariate statistical analysis, (7) validation; (8) susceptibility map.

2.4.3 Ordinal variables maps setting: Factors represented by vector maps are already categorized into qualitative classes, each assigned a consecutive integer value starting from 1. In contrast, raster-based maps contain continuous values, which are reclassified into ordinal classes, likewise assigned consecutive integer values beginning with 1. For the elevation map, the classification begins with the minimum value of the DTM, and the amplitude of the intervals can be defined as needed. For aspect, accumulation, and slope maps, a fixed classification has currently been adopted; however, it would be desirable to allow for customizable parameters, for example, the number and orientation of aspect classes or the degree intervals between slope classes.

2.4.4 Statistical sample setting: The statistical sample refers to the area affected by landslide phenomena. This step handles the import of landslide data in the form of vector polygon files. In the IFFI database, landslides are divided into two layers, the first one consisting of polygon for individual landslides and another representing polygon for areas affected by widespread landslides. In both layers, landslide types are identified by numerical values listed in a column of the attribute table. Alternative datasets can also be used, provided that the associated attribute table includes a column containing a unique numeric identifier for each landslide type represented. After a preliminary check to ensure the dataset fully covers the working area (by comparing the dataset extent using *v.info* with the region extent calculated using *g.region*), different types of soil landslides are extracted and loaded into GRASS as raster layers, where landslide areas are assigned a value of 1 and non-landslide areas a value of 0. For each landslide type, a dedicated mapset and a corresponding subfolder within the working directory are created, where the results of subsequent processing steps will be stored. To assess the reliability of the procedure and the influence of the various factors on each landslide type, 100% of the landslide polygons were initially considered. For a workflow that includes model calibration and validation, 80% of the polygons for each landslide type are randomly selected and used for calibration, while the remaining 20% are reserved for validation. The subsequent steps are conducted in parallel for each selected landslide type, and the resulting maps are stored in their respective mapsets.

2.4.5 Bivariate analysis: The statistical sample is compared with each predisposing factor to identify the classes in which the ratio between landslide area and class area is highest, according to the conditional probability formula:

$$P_{cond} = P(f|k) = \frac{P(f \cap k)}{P(k)} \tag{2}$$

where $P(f \cap k)$ is the probability that a pixel is inside a landslide area, and $P(k)$ is the probability that a pixel belongs to a certain class. For each factor, a comparison is made between the landslide map and the classified factor map using the *r.stats* command, producing a *.csv* file that reports the number of cells affected and unaffected by landslides for each class. These values are used to calculate the conditional probability for each class. The classes are then re-ranked: a value of 1 is assigned to the class with the lowest conditional probability, with increasing values assigned to the other classes. In cases where two or more classes have the same probability, they are assigned the same value. In the experimental setup, the entire study area is considered potentially susceptible to landslides, even in areas where the probability is extremely low, such as flat terrains. Consequently, even when a class contains no landslide cells (i.e., conditional probability equal to zero), it is still assigned the lowest value. Moreover, when using the Digital Lithological Map of Italy, which comprises 19 classes, it is common for several classes to contain no landslide occurrences and thus be assigned a value of 1, or to exhibit significantly lower probability values compared to most other classes. In such cases, it may be advisable to implement a script that groups these low-probability classes into a single category.

2.4.6 Multivariate statistics: Logistic regression is part of the family of Generalized Linear Models (GLMs). It can be regarded, in practical terms, as a form of linear regression of the type:

$$z = \beta_0 + \beta_1(x_1) + \beta_2(x_2) + \ldots + \beta_i(x_i) \tag{3}$$

where $x_1, x_2, ..., x_i$ are the classified and reordered maps of the different predisposing factors, β_0 is the intercept value, $\beta_1, \beta_2, ..., \beta_i$ are the weights assigned to the different factors.
The expected value z_i can be expressed as:

$$\text{logit}(Y) = z = \ln\frac{P(Y=1)}{1 - P(Y=1)} \tag{4}$$

By the properties of logarithms, (4) can be rewritten as:

$$P(Y_i = 1) = \frac{e^{z_i}}{1 + e^{z_i}} \tag{5}$$

The first step is to calculate the logit from the landslide area map using the *r.mapcalc* function:

$$\begin{aligned} \text{logit} = &\ln\left(\text{landslide\_map} + 10^{-8}\right) + \\ &- \ln\left(1 - \text{landslide\_map} + 10^{-8}\right) \end{aligned} \tag{6}$$

where the value 10^{-8} is added to avoid the possibility of having ln(0).
The GRASS command *r.regression.multi* establishes the relationships between the predisposing factors and the logit, producing a map of the expected values and a report containing various indices to assess the contribution of each factor. Even if the predisposing factors have been chosen as known to be related to landslide development and reciprocally independent, under specific environmental conditions, the influence of each one may vary. Among the indexes calculated by *r.regression.multi*, the Akaike Information Criterion (AIC) is particularly suitable to compare different models (Chakrabarti and Ghosh, 2011). Considering the models deriving from all the possible factors combinations, the lower the reported AIC value, the better is the model. Therefore, the factors combination that according to the AIC best fits the data is selected and it is used to process a map of the expected values.

2.4.7 Model validation: AIC values tell which model is the best among the possible ones, but do not supply any information about the actual reliability of the model itself. Therefore, it was decided to use the *r.edm.eval* addon (Van Breugel and GRASS Development Team, 2025), which offers a general evaluation of model reliability as a percentage, along with the Receiver Operating Characteristic (ROC) curve.

2.4.8 Identification of susceptibility classes: The map of expected values is converted into probability values, representing the actual susceptibility to landslides, using the inverse Logit formula (5). The values shown in this map are extremely small and difficult to interpret for land planning and management purposes. Therefore, it is useful to identify a few classes that qualitatively indicate the predisposition of each portion of the study area to landslide occurrence. The susceptibility map is compared with the actual landslide inventory, and the probability values within the landslide areas are analyzed. The mean probability value of these areas is used as the lower threshold for the “high” susceptibility class, while the remaining values are divided into two equal intervals corresponding to the “medium” and “low” susceptibility classes. Figure 3 reports an example of a resulting susceptibility map for the fast flows landslide type. However, this classification approach poses certain challenges, as it is sensitive to extreme values or outliers. As a result, alternative classification methods are currently under evaluation.

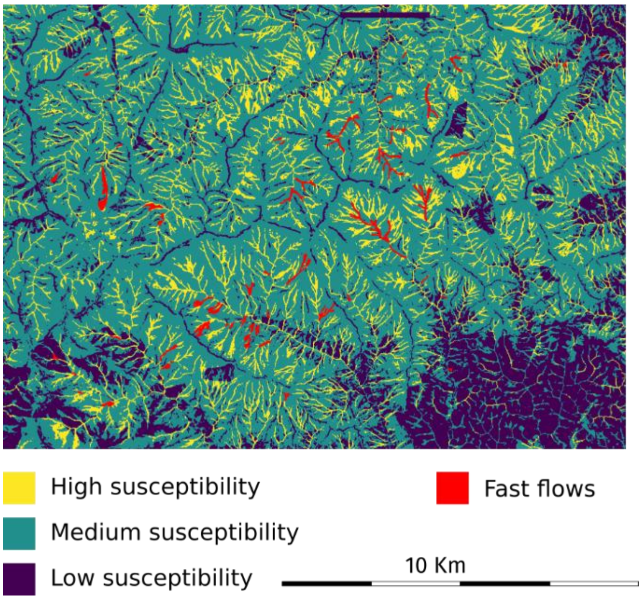

Figure 3. Fast flows susceptibility classes.

3. Conclusions and future developments

A model for assessing landslide susceptibility through logistic regression within a GIS environment has been developed, aimed at supporting public authorities and professionals in managing slope instability risk. The model focuses on flows, slides, and areas subject to diffuse shallow landslides and utilizes freely accessible national-scale datasets to ensure high transferability and transparency of results.

The analysis is implemented in Python and integrated into GRASS GIS, with the objective of automating the workflow and making the procedure accessible even to non-expert users. The code is available here: https://github.com/LabGeomatica/SALSA

The methodology has been applied in several areas of the Liguria region (Italy), directly providing the procedure and the input data to those who tested the method in other areas, yielding results consistent with those obtained for the province of Savona. Through automation, it will be possible to rapidly conduct analyses across multiple regions, thereby generating a sufficient volume of results to comprehensively evaluate the strengths and limitations of employing logistic regression for landslide susceptibility analysis. As part of the automation process, efforts are also underway to develop ancillary functions aimed at addressing issues encountered during experimentation.

The definition of import procedures and specifications for the basic data has certainly accelerated the preliminary work and helped to standardize the results. However, since the preparation of land use/land cover and climatic aggressiveness maps remains the responsibility of the operator, the use of the described procedure still requires a certain level of expertise. Moreover, both factors introduce elements of subjectivity because the procedures required for their preparation vary from case to case.

One potential approach currently under investigation involves utilizing satellite imagery in place of land use maps, from which different land cover types could be extracted via an unsupervised classification procedure. This method would enable the assessment of temporal changes in landslide susceptibility as a function of varying land cover.

As regards the processing of climatic aggressiveness, it remains complex due to the frequent incompleteness of rainfall data series, the involvement of multiple administrative bodies, and the variability in calculation methods, which can range in complexity depending on the availability and quality of input data. On the other hand, rainfall is undoubtedly an important factor in the development of landslide phenomena (Brunetti et al., 2025; Passalacqua et al., 2016; Troncone et al., 2021; Zhang et al., 2024) and the processing of time series data could help to better understand the relationship between extreme events and instability phenomena.

With regard to the definition of the statistical sample, i.e., the areas actually affected by landslides, the perimeters provided in the IFFI inventory include, for each landslide, both the detachment area, from which the landslide originated, and the accumulation area, representing the portion of the territory impacted by the landslide effects within which the displaced material lies above the original ground surface. Moreover, in the case of complex landslides and areas subject to widespread shallow landslides, the total extent of all the landslide phenomena that make them up is considered, including in the statistical sample areas that are not affected by landslides or are subject to different types of landslide mechanisms.

These situations certainly introduce noise into the model and reduce its accuracy. A preliminary solution is the development, currently in progress, of a script that separates the detachment area from the accumulation area based on elevation.

What has been achieved so far represents only the foundational framework; however, the procedure can be expanded by incorporating additional data, whether generally relevant or site-specific. It is essential that any new data introduced is independent of one another and of the variables included in the base model. Further customization may also involve implementing additional checks to verify the independence of the base data and to assess the influence of each factor on susceptibility. Moreover, in the case of large-scale landslides, post-event slope characteristics often differ substantially from those that contributed to the landslide initiation. In this regard, integrating the temporal dimension, such as time series on land cover and records of major rainfall events, could provide valuable insights.

Acknowledgements

This study was carried out within the RETURN Extended Partnership and received funding from the European Union Next-GenerationEU (National Recovery and Resilience Plan – NRRP, Mission 4, Component 2, Investment 1.3 – D.D. 1243 2/8/2022, PE0000005).

References

Arnoldus, H. M. J., 1980: An approximation of the rainfall factor in the universal soil loss equation. In: Assessment of Erosion, De Boodt, M. and Gabriels, D. (Eds.), John Wiley and Sons, New York, 127–132

Bovolenta, R., Passalacqua, R., Federici, B., Sguerso, D., 2016: LAMP—LAndslide Monitoring and Predicting for the analysis of landslide susceptibility triggered by rainfall events. *Landslides and Engineered Slopes. Experience, Theory and Practice*, Aversa, S., Cascini, L., Picarelli, L. and Scavia, C. (Eds.), 511–516. https://doi.org/10.1201/9781315375007

Brenning, A., Schwinn, M., Ruiz-Páez, A. P., Muenchow, J., 2015: Landslide susceptibility near highways is increased by 1 order of magnitude in the Andes of southern Ecuador, Loja province. *Natural Hazards Earth System Sciences*, 15, 45–57, https://doi.org/10.5194/nhess-15-45-2015

Brunetti, M. T., Gariano, S. L., Melillo, M., Rossi, M., Peruccacci, S., 2025: An enhanced rainfall-induced landslide catalogue in Italy. *Scientific Data*, 12. https://doi.org/10.1038/s41597-025-04551-6

Bucci, F., Santangelo, M., Fongo, L., Alvioli, M., Cardinali, M., Melelli, L., Marchesini, I., 2021: A new digital lithological Map of Italy at 1:100000 scale [dataset]. https://doi.org/10.1594/PANGAEA.935673

Bucci, F., Santangelo, M., Fongo, L., Alvioli, M., Cardinali, M., Melelli, L., Marchesini, I., 2022: A new digital lithological map of Italy at the 1:100 000 scale for geomechanical modelling. *Earth System Science Data*, 14(9), 4129–4151. https://doi.org/10.5194/essd-14-4129-2022

Chang, K. T., Merghadi, A., Yunus, A. P., Pham, B. T., Dou, J., 2019: Evaluating scale effects of topographic variables in landslide susceptibility models using GIS-based machine learning techniques. *Scientific Reports*, 9. https://doi.org/10.1038/s41598-019-48773-2

Chaohai, L., Yuan, R., Ying, W., 2025: Distribution laws of landslides and theirs influencing factors in the Qiaojia segment of Jinsha River, China. *Natural Hazards Research*, 5(1), 48–60. https://doi.org/10.1016/j.nhres.2024.06.002

Chakrabarti A., Ghosh J. K., 2011: AIC, BIC and Recent Advances in Model Selection. *Philosophy of Statistics*, 7, 583–605. https://doi.org/10.1016/B978-0-444-51862-0.50018-6

Chowdhury, M. S., Rahman, M. N., Sheikh, M. S., Sayeid, M. A., Mahmud, K. H., Hafsa, B., 2024: GIS-based landslide susceptibility mapping using logistic regression, random forest and decision and regression tree models in Chattogram District, Bangladesh. *Heliyon*, 10(1). https://doi.org/10.1016/j.heliyon.2023.e23424

European Parliament and of the Council of the European Union, 2007: Directive 2007/2/EC of the European Parliament and of the Council of 14 March 2007 establishing an Infrastructure for Spatial Information in the European Community (INSPIRE). https://eur-lex.europa.eu/eli/dir/2007/2/oj/eng

Goyes-Peñafiel, P., Hernandez-Rojas, A., 2021: Landslide susceptibility index based on the integration of logistic regression and weights of evidence: A case study in Popayan, Colombia. *Engineering Geology*, 280. https://doi.org/10.1016/j.enggeo.2020.105958

GRASS Development Team, 2025. Geographic Resources Analysis Support System (GRASS) Software, ver. 8.4. Open-Source Geospatial Foundation. grass.osgeo.org (15 May 2025)

ISPRA Istituto Superiore per la protezione e la Ricerca Ambientale, 2025a. IFFI – Inventario dei Fenomeni Franosi in Italia. https://www.progettoiffi.isprambiente.it/ (15 May 2025)

ISPRA Istituto Superiore per la protezione e la Ricerca Ambientale, 2025b. SCIA Sistema nazionale per l'elaborazione e diffusione di dati climatici. https://scia.isprambiente.it/dati-e-indicatori (22 April 2025)

Kadavi, P. R., Lee, C. W., Lee, S., 2019: Landslide susceptibility mapping in Gangwon-do, South Korea, using logistic regression and decision tree models. *Environmental Earth Sciences*, 78. https://doi.org/10.1007/s12665-019-8119-1

Lin, G. F., Chang, M. J., Huang, Y. C., Ho, J. Y., 2017: Assessment of susceptibility to rainfall-induced landslides using improved self-organizing linear output map, support vector machine, and logistic regression. *Engineering Geology*, 224, 62–74. https://doi.org/10.1016/j.enggeo.2017.05.009

Marzocchi, R., Rovegno, A., Federici, B., Bovolenta, R. e Berardi, R. 2015. Applicazione della regressione logistica per la zonazione della suscettibilità da frana in ambiente GIS. *Bollettino della società italiana di fotogrammetria e topografia*. 4 (giu. 2015), 39–47

Passalacqua, R., Bovolenta, R., Federici, B., Balestrero, D., 2016: A physical model to assess landslide susceptibility on large areas: recent developments and next improvements. *Procedia Engineering*, 158, 487–492. https://doi.org/10.1016/j.proeng.2016.08.477

Python Software Foundation, 2025. Python Language Reference, ver. 3.10.12. http://www.python.org (6 May 2025)

Regione Liguria, 2019a. Uso del Suolo sc. 1:10000 - ed. 2019 https://geoportal.regione.liguria.it/catalogo/mappe.html (17 April 2025)

Regione Liguria, 2019b. Uso del suolo - Allegato interpretativo alla carta dell'uso del Suolo della Regione Liguria. https://srvcarto.regione.liguria.it/repertoriocartografico/documentazione/Fotoatlante_Uso_suolo_Liguria.pdf?type=DS (17 April 2025)

Salmona, P., 2025: https://github.com/LabGeomatica/SALSA

Salmona, P., Bovolenta, R., Federici, B., Ferrando, I., 2025. Influence of data preprocessing and optimization in multivariate statistical analysis of landslide susceptibility. *Communications in Computer and Information Science*. Borgogno-Mondino, E., and Zamperlin, P., (Eds.). (in press)

Süzen, M. L., Kaya, B. Ş., 2011: Evaluation of environmental parameters in logistic regression models for landslide susceptibility mapping. *International Journal of Digital Earth*, *5*(4), 338–355. https://doi.org/10.1080/17538947.2011.586443

Troncone, A., Pugliese, L., Lamanna, G., Conte, E., 2021: Prediction of rainfall-induced landslide movements in the presence of stabilizing piles. *Engineering Geology*, 288. https://doi.org/10.1016/j.enggeo.2021.106143

Van Breugel, P., GRASS Development Team, 2025: Addon r.edm.eval. Geographic resources analysis support system (GRASS) software, ver. 8.4. Open-Source Geospatial Foundation. https://grass.osgeo.org/grass-stable/manuals/addons/r.edm.eval.html (20 May 2025)

Wu, J., Zhang, Y., Yang, L., Zhang, Y., Lei, J., Zhi, M., Ma, G., 2023: Identifying the essential influencing factors of landslide susceptibility models based on hybrid-optimized machine learning with different grid resolutions: a case of Sino-Pakistani Karakorum Highway. *Environmental Science and Pollution Research*, 30, 100675–100700. https://doi.org/10.1007/s11356-023-29234-w

Yang, Y., Peng, S., Huang, B., Xu, D., Yin, Y., Li, T., Zhang, R., 2024: Multi-scale analysis of the susceptibility of different landslide types and identification of the main controlling factors.

Ecological Indicators, 168. https://doi.org/10.1016/j.ecolind.2024.112797

Zanaga, D., Van De Kerchove, R., Daems, D., De Keersmaecker, W., Brockmann, C., Kirches, G., Wevers, J., Cartus, O., Santoro, M., Fritz, S., Lesiv, M., Herold, M., Tsendbazar, N. E., Xu, P., Ramoino, F., Arino, O., 2022: ESA WorldCover 10 m 2021 v200 [dataset]. https://doi.org/10.5281/zenodo.7254221

Zhang, Q., Shen, D., 2024: Rainfall-induced landslides: influencing, modelling and hazard assessment. *Water*, 16. https://doi.org/10.3390/w16233384

Zhao, Y., Wang, R., Jiang, Y., Liu, H., Wei, Z., 2019: GIS-based logistic regression for rainfall-induced landslide susceptibility mapping under different grid sizes in Yueqing, Southeastern China. *Engineering Geology*, 259. https://doi.org/10.1016/j.enggeo.2019.105147