CHAPTER 1

INTRODUCTION

Mass movements in mountainous terrain are natural degradational processes. Under the influence of a variety of causal factors, and triggered by events such as earthquakes or extreme rainfall, most of the terrain in mountainous areas has been subjected to slope failure at least once (Naranjo and van Westen, 1994).

In recent years, growing population and expansion of settlement and life-lines have largely increased the impact of natural hazards both in industrialized and developing countries. In many countries, the economic losses and causalities due to landslide are greater then commonly recognized and generate a loss of property larger that from any other natural hazards, including earthquakes, floods and windstorms.
Landslides are considered the second most significant natural hazard among those identified by the United Nations Development Program (UNEP, 1997). The full awareness of the effects produced by natural hazards led the United Nations, in 1989, to sponsor a resolution that declared the years 1990-2000 the "International Decade for Natural Disaster Reduction".
Damage caused by catastrophic events is too costly even for industrialized societies. In other words, natural catastrophes occur with higher frequency than our resilience or ability to recover from previous events. The recent trend is towards the development of warning systems and land utilization regulations aimed at minimizing the loss of lives and property damage.

The best method for any landslide mitigation project is the zonation of landslide **hazard**. It can be considered one of the most powerful tools to improve land-use planning and to avoid the development of threatened areas, the most efficient and economic way to reduce future damage and loss of lives (Cascini, 2002; Cascini, 2005). Hence, it should supply planners and decision makers with adequate and understandable information.

Many methods and techniques have been proposed in literature to evaluate the landslide hazard and produce map portraying its spatial distribution (landslide hazard zonation). The first term means "the probability of occurrence within a specific period of time and within a given area of potentially damaging phenomena" e.g, a landslide. The second term refers to the division of the land surface into homogenous areas or domains and their ranking according to different degrees of hazard due to mass-movement. According to this definition, hazard maps should both display the location of actual and potential slope-failure, and provide information on the time or probability of their future occurrence (return period).
However, on a regional scale the temporal dimension of landsliding is essentially a function of the triggering mechanisms which are climatic (due to extreme rainfall) or geodynamic (earthquakes) in nature (Dewitte et al., 2006). The timing of such triggers cannot readily be linked to a model of spatial instability which is essentially founded upon the geomorphological and geological features of a region.
Hence, most of the current hazard maps aim to predict where failures are most likely to occur without any clear indication of when they are likely to take place.
These should be better defined as landslide **susceptibility** maps (Brabb, 1984).

Landslide susceptibility is defined as the proneness of the terrain to produce slope failures. Susceptibility is usually expressed in a cartographic way. A landslide susceptibility map depicts areas likely to have landslides in future by correlating some of the principal factors that contribute to landsliding with the past distribution of slope failure (Brabb, 1984).
Because many factors can play a role in the occurrence of mass movements, the analysis is complex. It requires not only a large number of input variables, but techniques of analysis may be very costly and time-consuming.
Consequently, attention was given to seeking and developing methods and techniques to enable a faster and more efficient acquisition and processing of those geological-geomorphological data which are both relevant in assessing landslide susceptibility and mappable at effective cost over wide regions (Carrara et al., 1988).

Moreover, during the last decades, the increasing availability of computers has created opportunities for more detailed and rapid analyses.
It is proved that the development of Geographic Information Systems (GIS) has enhanced the capabilities for susceptibility assessment over large region (van Westen, 1993). The performance of neighborhood operations with the GIS allows extraction of morphological and hydrological parameters from Digital Elevation Models (DEM), that otherwise would be difficult to obtain. The main goal is the automatic capture of most of the parameters in relation to the occurrence of slope failures.

1.1 OBJECTIVES

The aim of this study is the assessment of landslide susceptibility at a regional scale. On this scale, statistical techniques are considered the most appropriate approach for landslide susceptibility zonation. In particular, among the different statistical techniques developed in literature during the years, the multivariate approach was tested.
This work research born within the Erasmus Project: preliminary studies have been developed in Italy at the University of Salerno (UNISA) and the main analyses were carried out at the Technical University of Catalonia (UPC) in Barcelona, Spain. The experience gained from the application of statistical multivariate analysis to assess shallow landslide susceptibility will be considered and applied to an Italian case.

1.2 CONTENTS

In particular, after a brief introduction in the Chapter 2 about landslides classifications developed in literature, their main triggering factors and the relevance of the working scale in landslide studies, the Chapter 3 focuses on the importance of susceptibility zoning in landslide risk management process, showing different methods available at different working scales and for different purposes, different zoning levels and different mapping units.

Then, the Chapter 4 describes the adopted procedure to analyze landslide susceptibility. The proposed approach is the multivariate statistical analysis, based on the relationship between instability factors and the past and present distribution of landslides. The statistical tool selected is the discriminant analysis, used to classify individuals or objects into mutually exclusive and exhaustive groups on the basis of a set of independent variables.
A notable example in literature is also defined in this chapter (Baeza&Corominas, 2001). It refers to the application of statistical procedure to assess shallow landslides susceptibility in the Spanish Eastern Pyrenees, showing in particular all the steps followed in order to create a

reliable susceptibility map.

The Chapter 5 analyzes two attempts in applying the same statistical procedure in the Italian Benevento Province, in order to study large landslide susceptibility. The first, aimed at distinguishing active and dormant phenomena, gives irrelevant results and highlights the importance of an appropriate completion of the landslide inventories. The second is an attempt to analyze the actual distribution of landslides by multivariate statistical techniques.

Finally, concluding remarks are developed from the experience gained from the application of the statistical procedure in the different cases described.

CHAPTER 2

LANDSLIDES AND SCALES

2.1 LANDSLIDES AND THEIR CLASSIFICATIONS

2.1.1 Typologies of movement

The most common definition of a landslide used in literature is that given by Cruden, 1991: *"the movement of a mass of rock, earth or debris down a slope"*.
Different classifications of landslides have been developed in literature during the years. Among these the most noted belong to Varnes (1978), Hutchinson (1988), Cruden and Varnes (1996), Leroueil (1996) and Hungr (2001).
The classification used here refers to the one proposed by Varnes and modified by Cruden and Varnes and it is based on the typology of movement and the material involved.
Regarding the type of movement, the classification distinguishes:

- Falls. Movement of materials in free fall and subsequent movement, for jumps and/or bounce, of the fragments of rock.

- Topples. Movements similar to falls, characterized by the frontal overturning of the material around a point situated below the centre of gravity of the mass.

- Slides. Movements that occur when achieving the shear resistance available along one or more surfaces of neo-formation or pre-existing surfaces. These movements can be distinguished in rotational and translational, according to the shape of the slip surface.

- Lateral spreads. When a fractured mass of rock is overlapped to a plastic terrain, the mobilization of the overlooking rigid blocks may cause distortion in the underlying plastic masses.

- Flows. Movements characterized by plastic deformation of the involved materials.

- Complex phenomena. Phenomena whose movement is the result of the combination of two or more types of landslides described.

As in the landslide definition, the classification distinguishes three typologies of the material involved:

- Rock. It is a hard or firm mass that was intact and in its natural place before the initiation of the movement.

- Earth. It describes material in which 80% or more of the particles are smaller than 2mm, the upper limit of sand sized particles.

- Debris. Contain a significant proportion of coarse material: between 20% and 80% are larger than 2mm, and the remainder are smaller than 2mm.

By the intersection of the definitions given regarding the types of movement and the types of material, the resulting table (Table I) individuates different terminologies for landslides.

Table I. Cruden and Varnes landslide classification (1996)

TYPE OF MOVEMENT		TYPE OF MATERIAL		
		BEDROCK	ENGINEERING SOILS	
			Predominantly coarse	Predominantly fine
FALLS (A)		Rock fall	Debris fall	Earth fall
TOPPLES (B)		Rock topple	Debris topple	Earth topple
SLIDES (C)	ROTATIONAL	Rock slide	Debris slide	Earth slide
	TRANSLATIONAL			
LATERAL SPREADS (E)		Rock spread	Debris spread	Earth spread
FLOWS (D)		Rock flow (deep creep)	Debris flow (soil creep)	Earth flow
COMPLEX Combination of two or more principal types of movement				

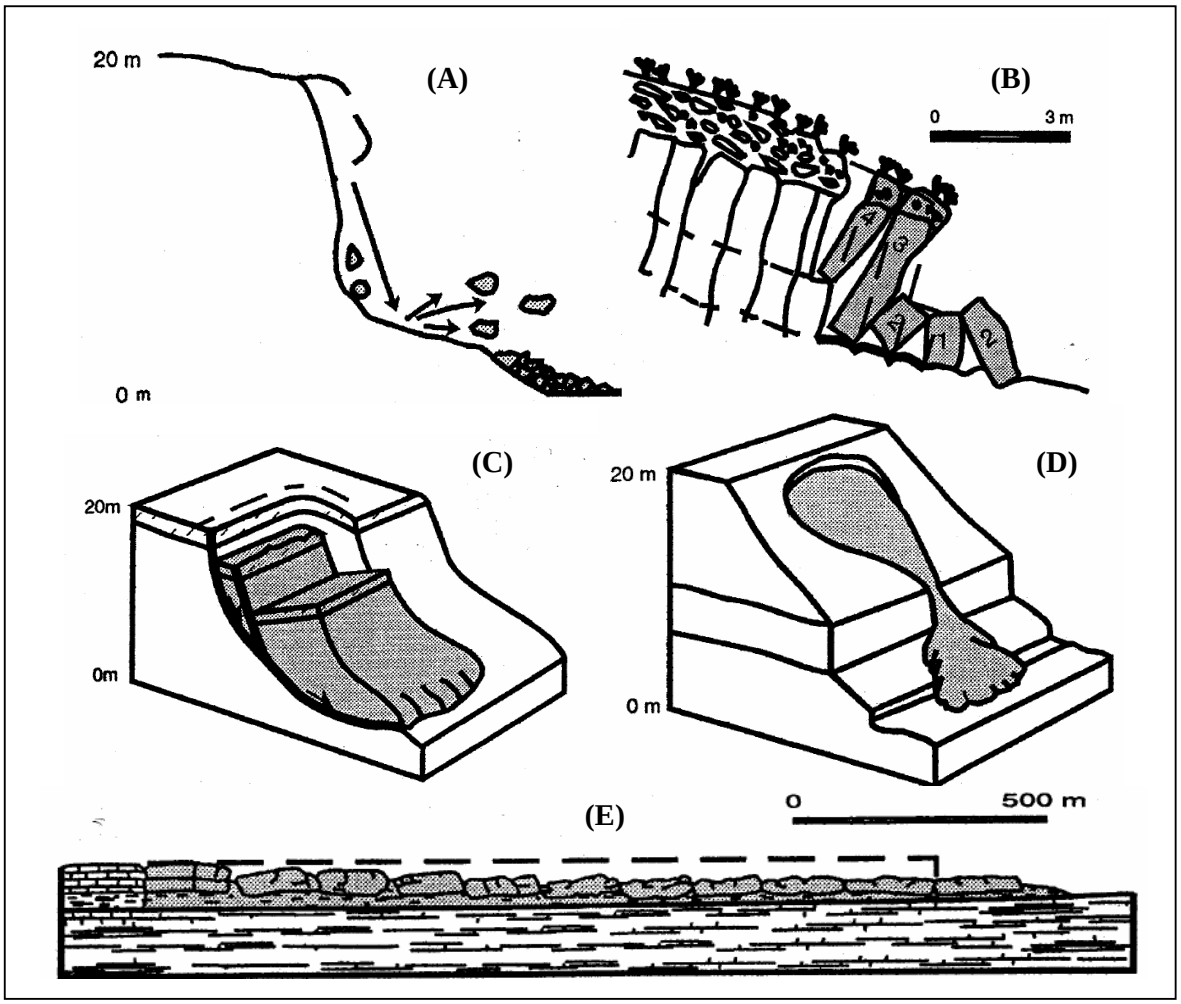


Figure 1. Examples of landslides

Regarding the activity of landslides, the subject is more complex. Actually, under activity broad aspects of landslides are described, those aspects that should focus the initial reconnaissance of movements before more detailed examination of the materials displaced.
Taking in account the World Landslide Inventory suggested by the Working Party (WP/WLI, 1994) as a contribution to the International Decade for Natural Disaster Reduction (1990-2000), we found standard terminology for describing landslides. In particularly we focus on the **activity** definitions.
The terms Varnes (1978) considered relating to age and state of activity with some of the terms from sequence or repetition of movement have been regrouped under three headings (table II):

- *State of Activity*
  Describes what is known about the timing of movements.

- *Distribution of Activity*
  Describes broadly where the landslide is moving.

- *Style of Activity*
  Indicates how different movements within the landslide contribute to its overall movement.

Table II. Activity definition (Varnes, 1978; Cruden and Varnes, 1994)

State of activity	Distribution of activity	Style of activity
Active	Retrogressing	Complex
Reactivated	Advancing	Composite
Suspended	Widening	Multiple
Inactive : Dormant	Confined	Successive
: Abandoned	Enlarging	Single
: Stabilized	Diminishing	
: Relict	Moving	

2.1.2State of Activity

Active landslides are those that are currently moving. Landslides which have moved within the last annual cycle of seasons but which are not moving at present were described by Varnes (1978) as *suspended*.
Inactive landslides are those which last moved more than one annual cycle of season ago. This state may be subdivided. If the causes of movement apparently remain, then the landslides are *dormant*. Perhaps, however, the river which had been eroding the toe of the moving slope has itself changed course and the landslide is *abandoned*. If the toe of the slope has been protected against erosion by bank armoring or other artificial remedial measures have stopped the movement, the landslide can be described as *stabilized*. Landslides often remain visible in the landscape for thousands of years after they have initially moved. Landslides which have clearly developed under different geomorphological or climatic conditions have been called *relict*.
A landslide which is again active after being inactive may be called *reactivated*. Slides that are reactivated generally move on pre-existing shears whose strength parameters approach residual (Skempton, 1970) or ultimate (Morgensten, 1979) values. They can be distinguished from first-time slides on whose rupture surfaces resistance to shear may initially approach peak values.

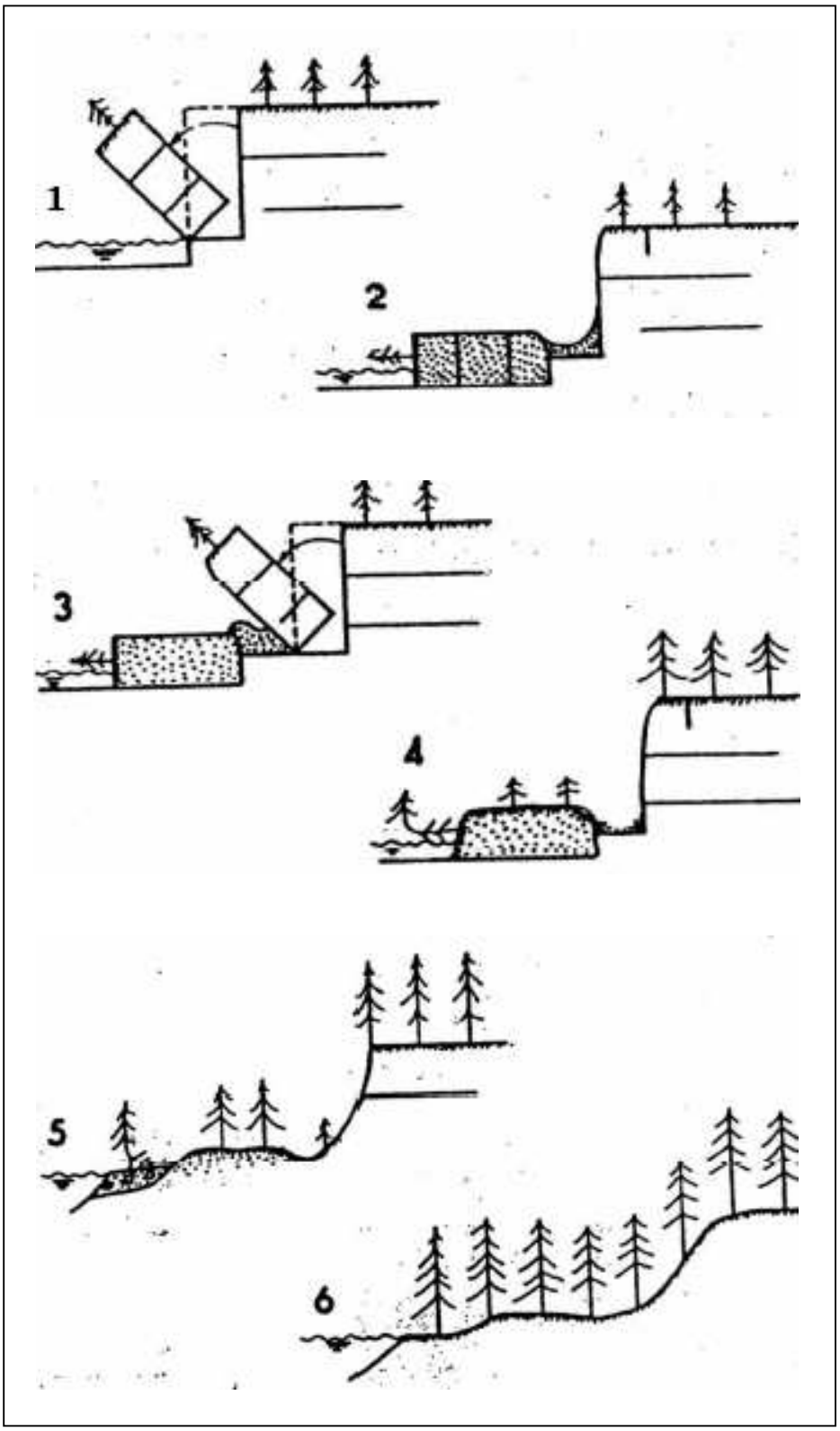


Figure 2. Sections trough topples in different states of activity: (1) *active*; (2) *suspended*; (3) *reactivated*; (4) *dormant*; (5) *stabilized*; (6) *relict* (WP/WPLI, 1994).

The definition of the various states of activity can also be shown on a graph of the displacing material against time. This graph might be created by plotting differences in the position of a target on the displacing material against time. Such graphs are particular suited to portraying the behavior of slow-moving landslides.

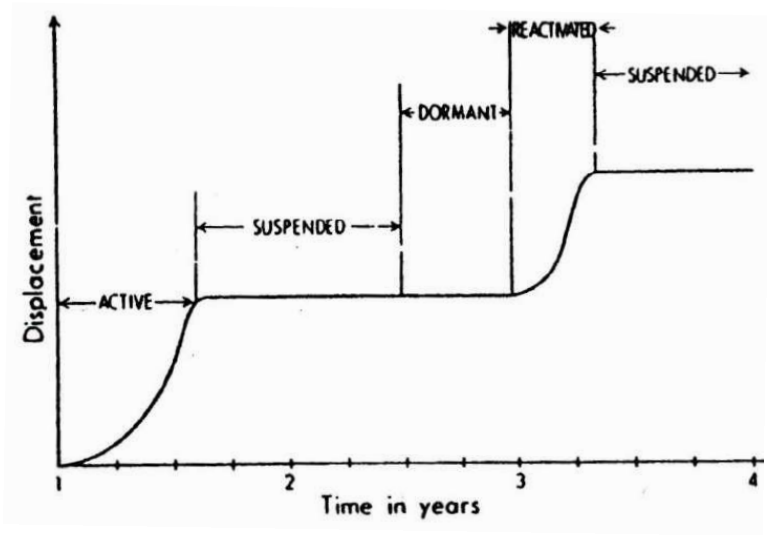


Figure 3. Graph showing the displacing material against time (WP/WPLI, 1994).

2.1.3 Distribution of Activity

Varnes (1978) defined a number of terms that can be used to describe the activity distribution of a landslide. Movements may be limited to the displacing material or the rupture surface may be extending, continually adding to the volume of the displacing material. If the rupture surface is extending in the direction opposite to the movement of the displaced material, the landslide is called *retrogressing*. If the rupture surface is extending in the direction of the movement, the landslide is *advancing*. If the rupture surface is extending at one or both lateral margins, the landslide is *widening*.

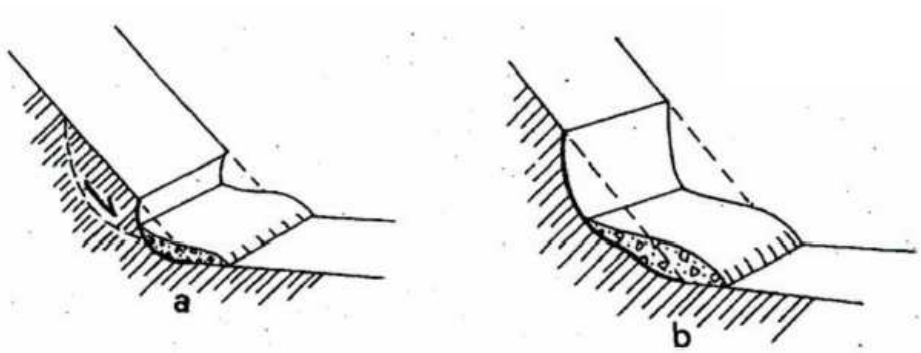


Figure 4. Distribution of activity – Retrogressing (WP/WPLI, 1994).

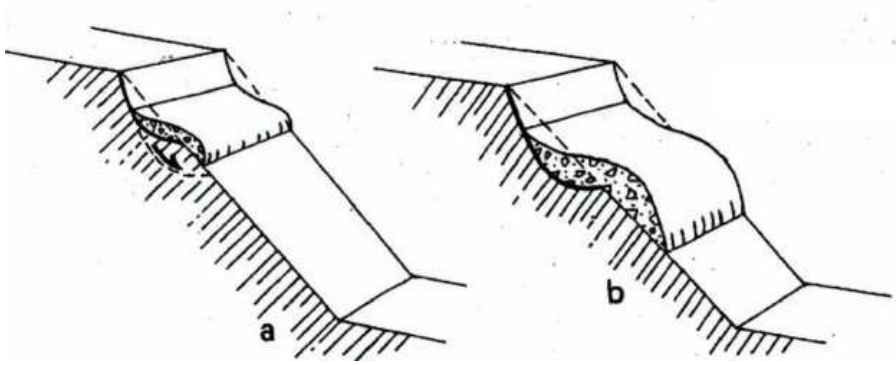


Figure 5. Distribution of activity – Advancing (WP/WPLI, 1994).

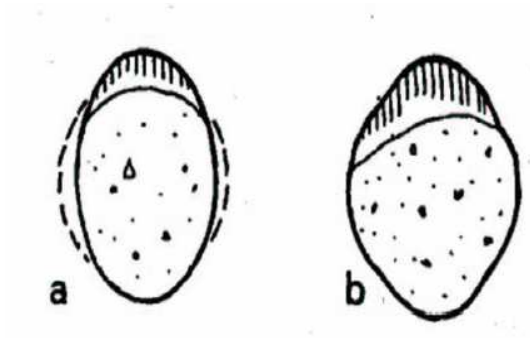


Figure 6. Distribution of activity – Widening (WP/WPLI, 1994).

Hutchinson (1988) has drawn attention to *confined* movements, which have a scarp but no rupture surface visible in the foot of the displaced mass. He suggested that displacements in the head of the displaced mass are taken up by compression and slight bulging in the foot of the mass. If the rupture surface of the landslide is enlarging in two or more directions, Varnes (1978) suggested the term progressive for the landslide while noting this term had also been used for both advancing and retrogressive landslides. This term is also current for describing the process, progressive failure, by which the rupture surface in some slides extends. The possibility of confusion seems sufficient now to abandon the term progressive in favor of describing the landslide as *enlarging*.

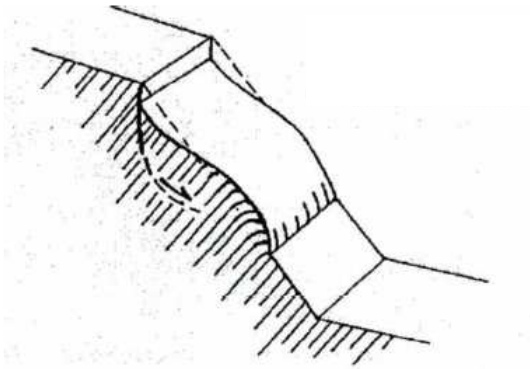


Figure 7. Distribution of activity – Confined (WP/WPLI, 1994).

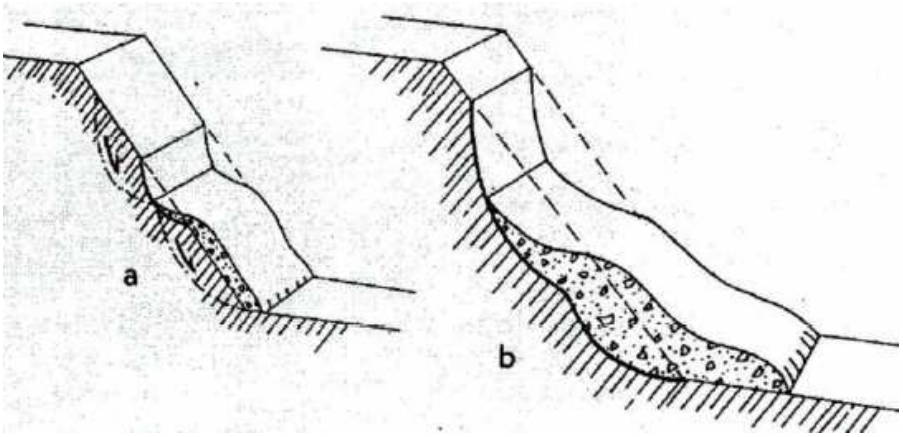


Figure 8. Distribution of activity – Enlarging (WP/WPLI, 1994).

The term *diminishing* is used to characterize landslide whose displacing material is decreasing in volume, while landslides whose displaced materials continue to move but whose rupture surfaces show no visible changes can be simply described as *moving*.

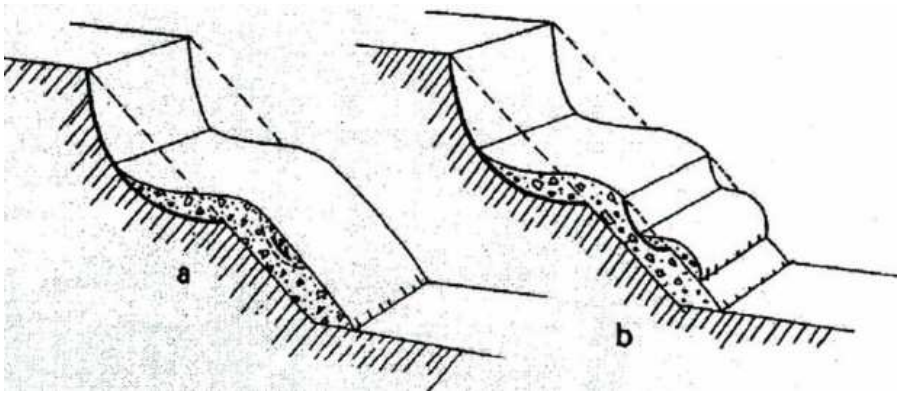


Figure 9. Distribution of activity – Diminishing (WP/WPLI, 1994).

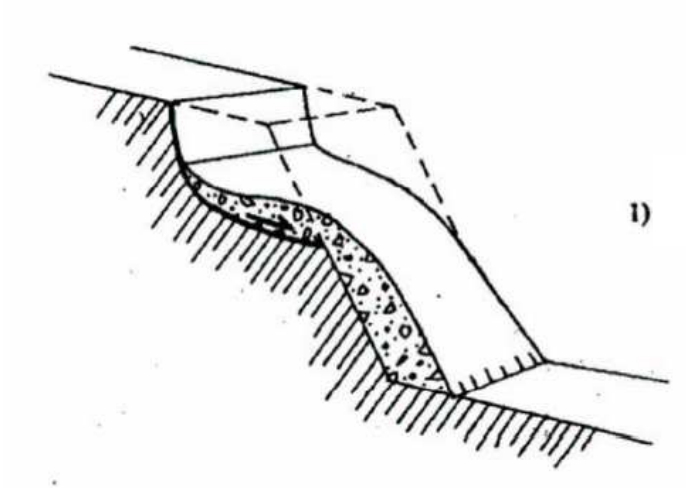


Figure 10. Distribution of activity – Moving (WP/WPLI, 1994).

2.1.4 Style of activity

The way in which different movements contribute to the landslide can be described by terms from Varnes (1978).

Complex landslides are defined as exhibiting at least two types of movements in which the types are in sequence.

The term *composite* can be used to distinguish landslides, previously classifies as complex, in which two different types of movements occur in different areas of the displaced mass.

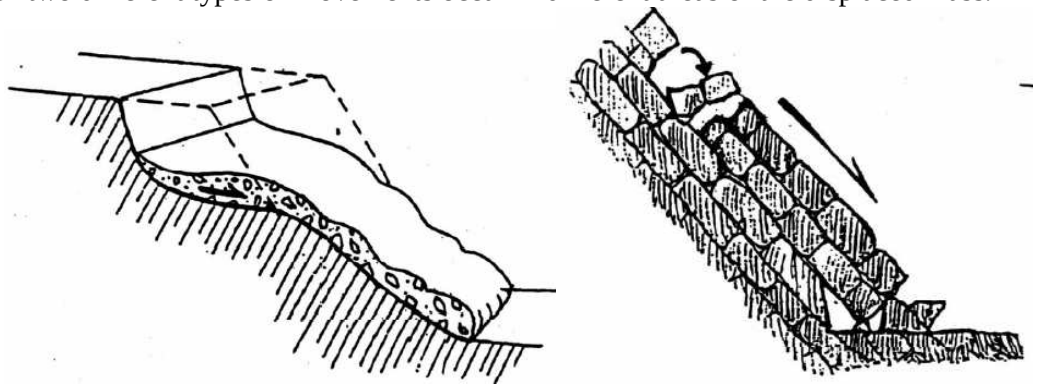


Figure 11 and 12. Style of activity - Complex and Composite (WP/WPLI, 1994).

Multiple movements (Hutchinson, 1968) are landslides with repeated development of the same type of movement. In a retrogressive, multiple, rotational slide, two or more blocks have each moved on curved rupture surfaces tangential to a common generally deep rupture surface.

A *successive* movement is identical in type to the earlier movement but in contrast to a multiple movement does not share displaced material or a rupture surface with it.

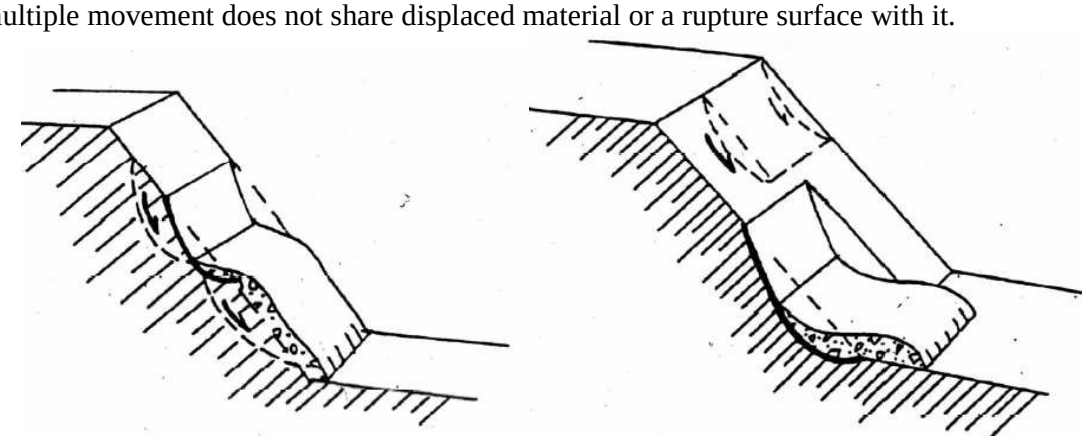


Figure 13 and 14. Style of activity – Multiple and Successive (WP/WPLI, 1994).

Single landslides consist of a single movement of displaced material often as an unbroken block. They are in contrast to the previous styles of movements which require disruption of the displaced mass or independent movements of portions of the mass.

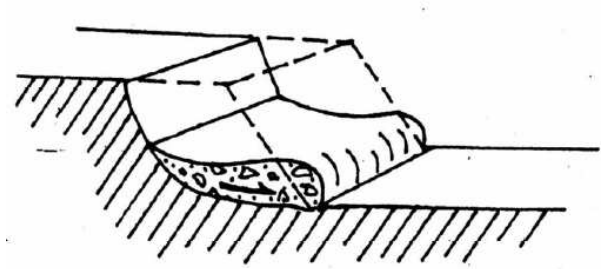


Figure 15. Style of activity – Single (WP/WPLI, 1994).

2.1.5 Landslide triggering factors

In general, a slope becomes unstable when the shear stress along a potential failure surface exceeds the available shear strength. Referring to the definition of the safety factor, it is represented by the ratio between the shear strength τ_f and shear stress τ_{eq} on the potential slip surface, as described below:

$$F = \frac{\tau_f}{\tau_{eq}} \qquad (1)$$

The triggering mechanisms for slides can be categorized into 2 groups:

- Factors that contribute to increase the shear stress on the potential slip surface.
- Factors that contribute to reduce the soil shear strength.

Among the factors that contribute to increase the shear stress, it is possible to include:

(c) Removal of lateral slope support

- *toe erosion by waves or tidal currents*
- *man-made cuts at foot of slope*
- *altered levels of lakes or reservoirs*

(b)Added surcharge (loads)

- *weight due to saturation of dry slope*
- *accumulation of sediments, talus, colluviums, debris, or volcanic material*
- *seepage forces*

(c) Earthquake shaking

- *inertial forces*

(d)Increased lateral pressures

- *water in cracks*
- *freezing of water in cracks*
- *swelling pressures as a result of hydration*

(e)Regional tilting and increase in slope angle

(f)Volcanic processes

- *changes in stress patterns*
- *fluctuation in lava-lake levels*
- *increase in tremors*

Among the factors that contribute to reduce shear strength, it is possible to enclose:

(a)Changes in ground water conditions due to natural or man-made processes

- *increased pore pressures and reduced effective stresses due to cyclic loading*
- *capillary tension (effective stresses) lost upon saturation*
- *erosion*
- *softening of fissured and cracked soils (e.g. OC-clays and shales)*
- *"collapse" of loess due to leaching out of clay cement bonding between silt particle*
- *loss of vegetation and root system which may have stabilized the slope*

(b)Changes due to weathering and other physical and chemical reactions

- *physical disintegration of granular rocks (e.g. granite, sandstone) under action of thermal expansion/frost*
- *base exchange in clays*
- *drying of clays and shale that results in cracks and subsequent water infiltration*
- *removal of cement bonding by solution*

(c) Creep deformations with time

- *weakening due to progressive creep (from peak strength to residual strength)*
- *loosening of slopes due to release of lateral restraint*
- *opening of fissures and cracks.*

2.2 SCALES

As already said, landslides are dangerous phenomena that may cause damage and loss of lives. The risk associated to a landslide depends on different factors. According to the definition given by Varnes (1994), the *risk* is 'the measure of the probability and severity of an adverse effect to life, health, property, or the environment'. It can be seen as the product of three components:

$$R = H \cdot E \cdot V \qquad (2)$$

where:

- ***H: Hazard.*** Probability that a particular danger (threat) occurs within a given period of time;
- ***E: Elements at risk.*** Population, buildings and engineering works, economic activities, public services utilities, other infrastructures and environmental values in the area potentially affected by the landslide hazard.
- ***V: Vulnerability.*** The degree of loss to a given element or set of elements within the area affected by the landslide hazard. It is expressed on a scale of 0 (no loss) to 1 (total loss).

According to the framework proposed by Fell et al. (2005), landslide risk management process includes three phases: the risk analysis, the risk assessment and the risk management.

- *Risk analysis* includes hazard analysis (landslide or danger characterization and analysis of frequency) and consequence analysis (characterization of consequence scenarios, analysis of probability and severity of consequence), with the identification and quantification of the elements at risk, their temporal-spatial probability and their vulnerability.
- *Risk assessment* takes the output from the risk analysis and assesses these against values judgments and risk acceptance criteria (policy-maker decisions regarding risk acceptability or treatment and priorities to be set according to a complex procedure that must consider both technical and socio-economic aspects).
- *Risk management* takes the output from the risk assessment and considers risk mitigation, including accepting the risk, reducing the likelihood and reducing the consequences, e.g. by developing monitoring, warning and evacuation plans.

The framework proposed by Fell et al. (2005) is reported below.

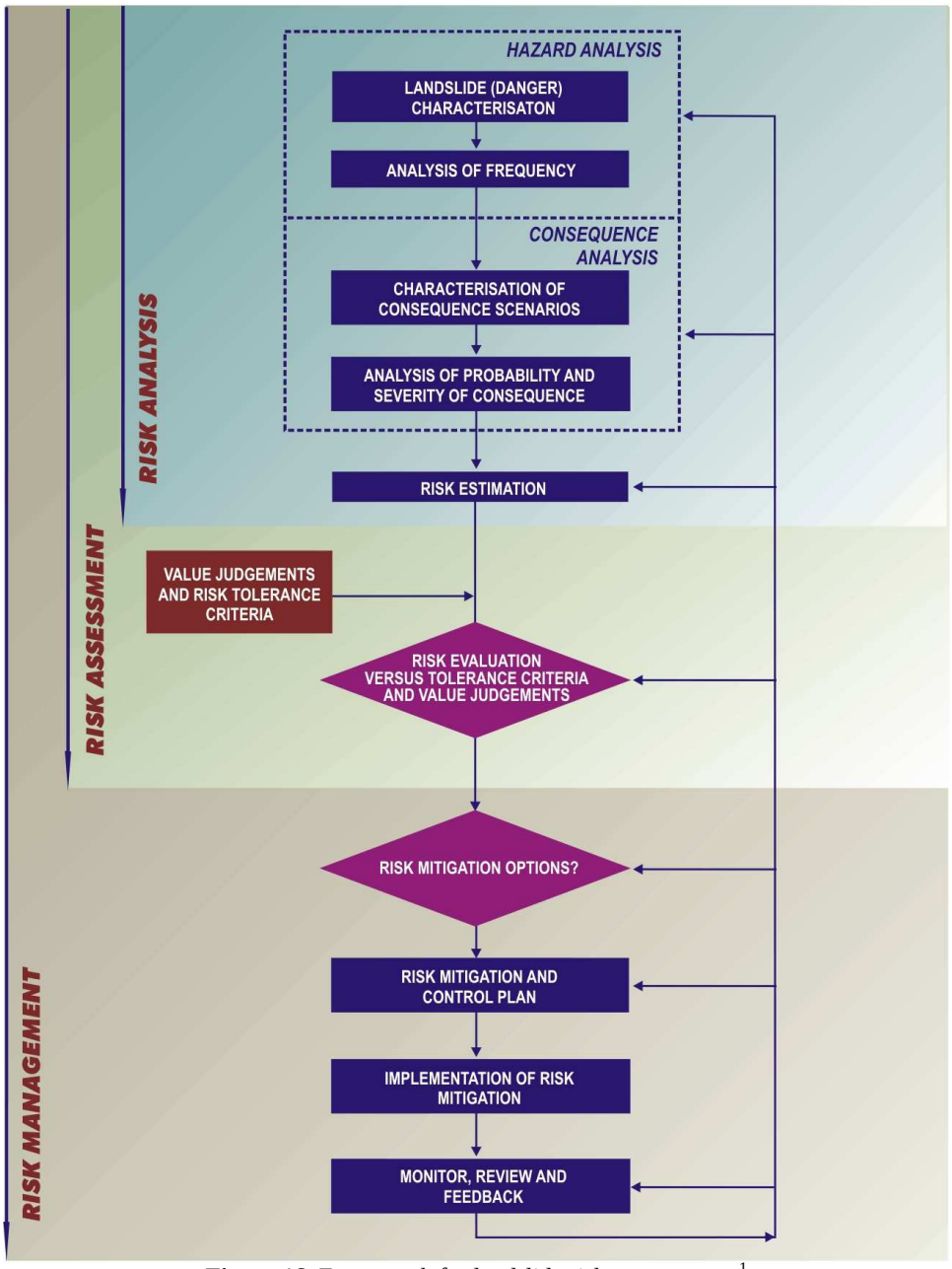


Figure 16. Framework for landslide risk management[1]

Within this framework, the rule exercised by the working scale is very important. Actually, landslides are phenomena with complex feedback varying in scale from the local to the regional. The complexity of the landslide phenomena, and in particular the role of rainfall infiltration and run-off, often require analyses at the level of the drainage area.

A correct investigation first requires an analysis at a small scale (1:100,000 – 1:50,000), creating a document useful for general planning purposes in which landslide only constitute one of the numerous planning constraints.

[1] This figure is taken from Fell et al. (2005) and represents a framework widely used internationally. It was the basis for the State of the Art papers and invited papers at the International Conference on Landslide Risk Management held on Vancouver in May 2005 (Hungr et al., 2005).

Then, in densely populated areas, investigations on landslide hazards have to be improved at an intermediate scale (1:25,000) in order to give more precise delimitations of the exposed zones. At this scale, valuable maps can be useful in implementing monitoring systems.
Finally, when risk analyses are carried out at the level of plots of land or individual buildings, large scale mapping is required (1:5,000 or larger), especially where the value of the land justifies exploiting any possibility of housing development in safe zones even if they are quite near to landslide zones.
Of course, it is important to adapt the quality of landslide investigations to both the required scale and the pursued aims. In particular, when large scale landslide maps may severely reduce the value of a plot of land, detailed in-depth data must be gathered by in-situ investigations (boreholes, inclinometers and other techniques) as well as by mathematical modeling which, in turn, can improve the monitoring system at a site scale.
Leroi (1996) introduces risk mapping as a problem at different scales, with each scale having a well defined meaning and aim (Figure 17).

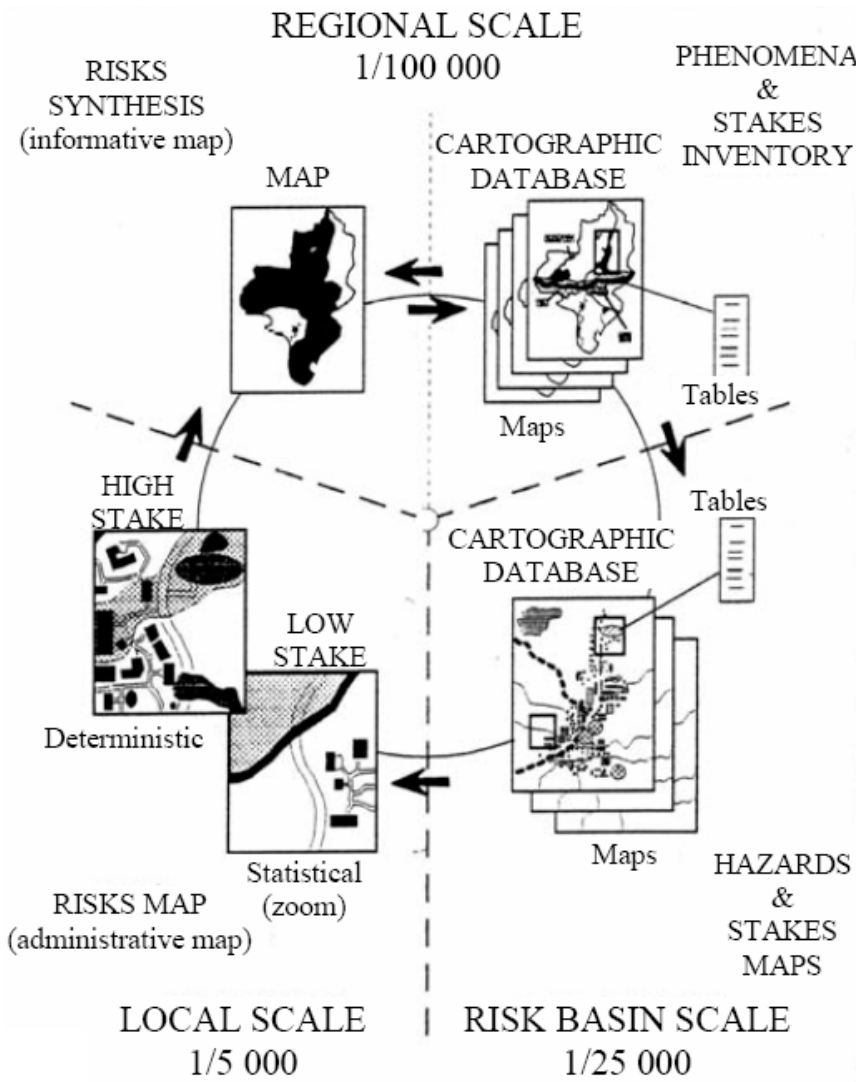


Figure 17. Risk mapping at different scales (Leroi, 1997)

CHAPTER 3

SUSCEPTIBILITY ZONING

Destitution and demographic pressure have led lots of people to live in areas prone to landslides. Poor land-use planning, environmental mismanagement and a lack of regulatory mechanism increase the risk and exacerbate the effects of disasters. Several cases of the negative role played by demographic pressure on the increasing number of disasters can be mentioned (Brand, 1988).
For example, in the case of very large dormant landslide zones, or that are generally affected by slow movements that may be permanent or occasional, old villages or new urban areas may extend onto these unstable areas, first because an active landslides zone generally presents a more gentle slope than adjacent stable zones, and thus is assessed as more favorable for settlements; then because the fast development of the suburbs of a city located in a valley may induce inhabitants to occupy unstable slopes in the vicinity of the city center, where stable areas are not available.

This example highlights that the major risks in urban areas derive from the unplanned development during centuries as well as from the growth of marginal housing in landslide-prone areas. Due to the dense occupation of such poor urban areas, the risk for life related to a sudden landslide event is very critical.
In similar cases, a reliable **risk** zoning for urban planning and development is an urgent need.
Within the landslide risk management framework proposed by Fell et al. (2005) (Figure 16), **hazard** zoning turns out to be a part of both risk analysis and risk assessment since the hazard distribution must be compared with the urban plan.

Landslide hazard zoning have been developed since the 1970's (Brabb et al., 1972; Nilsen et al., 1979; Kienholz, 1978) to face practical problems at different scales. The scientific literature highlights the extensive development during the last few decades of landslide hazard zoning, which nowadays is a powerful tool to improve land-use planning and to avoid the development of threatened areas. It represents the most efficient and economic way to reduce future damage and loss of lives (Cascini, 2002; Cascini, 2005).

Hazard maps should provide information about the location of actual and potential slope-failure and about the return period. However, on a regional scale the temporal dimension of landsliding is essentially a function of the triggering mechanisms which are climatic (due to extreme rainfall) or geodynamic (earthquakes) in nature (Dewitte et al., 2006). The timing of such triggers cannot readily be linked to a model of spatial instability which is essentially founded upon the geomorphological and geological features of a region.
Hence, most of the current hazard maps aim to predict where failures are most likely to occur without any clear indication of when they are likely to take place.
These should be better defined as landslide **susceptibility** maps (Brabb, 1984).
Actually, susceptibility, hazard and risk are often used interchangeably in landslide zoning maps. However, a review of the recent experience in landslide susceptibility and hazard zoning (Cascini et al. 2005) highlights the fact that these maps have different accuracy and reliability. The amount of information supplied and the degree of detail constitute the greatest difference between susceptibility, hazard and risk maps.

3.1 TYPES OF LANDSLIDE ZONING FOR LAND USE PLANNING

In order to avoid confusion about the terminology, some definitions have to be supplied. According to the International Guidelines currently being developed by JTC-1 (Joint Technical Committee on Landslides and Engineered Slopes), different types of landslides mapping/zoning are possible, respectively defined as:

- Landslide Inventory Mapping;
- Landslide Susceptibility Zoning;
- Landslide Hazard Zoning;
- Landslide Risk Zoning.

A description of each typology of zoning is presented below.

Landslide Inventory Mapping

Involves the location, classification, volume, activity, date of occurrence and other characteristics of landslides which exist or potentially may occur in an area.
Landslide inventories are essentially factual in nature. However in some cases there may be a degree of interpretation because they may be based on geomorphologic attributes seen on air photographs or mapped on the ground.

Landslide Susceptibility Zoning

Involves the classification, volume (or area) and spatial distribution of existing and potential landslides in the study area. It may also include a description of the travel distance, velocity and intensity of the existing or potential landsliding. Landslide susceptibility zoning usually involves developing an inventory of landslides which have occurred in the past together with an assessment of the areas with a potential to experience landsliding in the future, but with no assessment of the frequency (annual probability) of the occurrence of landslides. In some situations susceptibility zoning will need to be extended outside the study area being zoned for hazard and risk to cover areas from which landslides may travel on to or regress into the area being zoned. It will generally be necessary to assess independently the propensity of the slopes to fail and areas onto which landslides from the source landslides may travel or regress.
Landslide susceptibility zoning involves a degree of interpretation. Susceptibility zoning involves the spatial distribution and rating of the terrain units according to their propensity to produce landslides. This is dependent on the topography, geology, geotechnical properties, climate, vegetation and anthropogenic factors such as development and clearing of vegetation. It should consider all landsliding which can affect the study area and include landslides which are above the study area but may travel onto it, and landslides below the study area which may retrogressively fail up-slope into it.
The travel and regression of the landslides is dependent on different factors to those causing the landslides. Areas which may be affected by travel or regression of the landslides from the source will often be assessed independently.

There are some differences of viewpoint amongst experts in landslide zoning as to whether susceptibility zoning should include an assessment of the potential travel or

regression of landslides from their source. Some feel that this should be considered only in hazard zoning. However, in some situations it will be difficult to assess the frequency of landsliding and land use zoning may be carried out based on susceptibility zoning. In these cases the important matter of travel or regression would be lost. In view of this travel and regression should be considered in susceptibility zoning.

It should be recognized that the study area may be susceptible to more than one type of landslide e.g. rock fall and debris flows, and may have a different degree of susceptibility (and in turn hazard) for each of these. In these cases it will often be best to prepare separate susceptibility, and hazard zoning maps for each type of landslide and to combine them to obtain the global landslide hazard map of the area.

Landslide Hazard Zoning

Takes the outcomes of landslide susceptibility mapping and assigns an estimated frequency (annual probability) to the potential landslides. It should also consider all landsliding which can affect the study area including landslides which are above the study area but may travel onto it, and landslides below the study area which may retrogressively fail up-slope into it. The hazard may be expressed as the frequency of a particular type of landslide of a certain volume, or landslides of a particular type, volume and velocity (which may vary with distance from the landslide source), or in some cases as the frequency of landslides with a particular intensity, where intensity may be measures in kinetic energy terms. Intensity measures are most useful for rock falls and debris flows (e.g. depth x velocity).

Hazard zoning should be done for the area in its condition at the time of the zoning study. It should allow for the effects of existing development (such as roads) on the likelihood of landsliding. In some situations the planned development may increase or reduce the likelihood of landsliding. This can be assessed and a post-development hazard zoning map produced. Hazard zoning may be quantitative or qualitative. It is generally preferable to determine the frequency of landsliding in quantitative terms so the hazard from different sites can be compared, and the risk estimated consequently also in quantitative terms. However in some situations it may not be practical to assess frequencies sufficiently accurately to use quantitative hazard zoning and a qualitative system of describing hazard classes may be adopted. Usually, even for these cases, it will be possible to give some approximate guidance on the frequency of landslides in the zoning classes and this should be done.

Landslide Risk Zoning

Takes the outcomes of hazard mapping, and assesses the potential damage to persons (annual probability of loss of life), and/or to property (annual value of property loss) for the elements at risk, accounting for temporal and spatial probability and vulnerability.

Risk zoning depends on the elements at risk, their temporal spatial probability and vulnerability. For new developments, an assessment will have to be made of these factors. For areas with existing development it should be recognized that risks may change with additional development and thus risk maps should be updated on a regular basis. Several risk zoning maps may be developed for a single hazard zoning study to show the effects of different development plans on managing risk.

It can be observed that **the landslide inventory** is the basis for all the mapping, and it is important that this activity be done thoroughly. With this aim, the inventory should be mapped at a larger scale than the susceptibility and hazard zoning maps.

The type of zoning and level of detail associated depend on different factors:

- The stage of development of the land use zoning plan or engineering project. **Susceptibility and hazard** zoning are more likely to be used in preliminary stages of development, and **hazard and risk** zoning for more detailed stages. However the choice depends mostly on the intended purpose of the zoning in land use management and on the policies favored by governments and other interested stakeholders.
- The type of development**. Risk** zoning is more likely to be used for existing urban developments where the elements at risk are defined or for existing and planned road and railway developments where the elements at risk (the road or rail users) are readily predicted. However, the elements at risk often vary with time so risk zoning needs to be up-dated regularly.
- The classification, activity, volume or intensity of landsliding**. Risk zoning** is more likely to be required where the landslides are likely to travel rapidly and or have a high intensity as measured by the combination of volume and velocity (e.g. rock fall, debris flows, rock avalanches). For these situations life loss is more likely so it is useful to use risk zoning as this allows land use zoning to be determined using life loss risk criteria.
- While the purpose should determine the level of zoning and the scale of the maps, the funding available may be a practical constraint. Landslide **susceptibility** zoning is less costly than hazard zoning, and hazard zoning is somewhat less expensive than risk zoning, so land use planners may opt for a lesser type and level of mapping at least in a staged introduction of landslide land use planning.
- The amount and quality of available information. **Quantitative risk zoning** can not be performed where data on frequency of landslides either do not exist or are so uncertain as to not be relied on. In such a case, **susceptibility zoning** is recommended.
- The history of the area being zoned and its evolution in terms of land use must be carefully taken into account as human activities may modify the slope instability environment and modify the susceptibility to and likelihood of landsliding and hence the hazard.
- Qualitative methods are often used for s**usceptibility zoning**, and sometimes for **hazard zoning**. It is better to use quantitative methods for both susceptibility and hazard zoning. **Risk zoning** should be quantified. More effort is required to quantify the hazard and risk but there is not necessarily a great increase in cost compared to qualitative zoning.
- The required accuracy of the zoning boundaries. Where statutory land use planning constraints are proposed, large scale maps with appropriate levels of inputs should be used. In this regards it should be noted that State and Local governments may have different requirements. The largest scale required will determine the level and scale of landslide zoning.
- The use of complementary or linking processes such as planning schedules and development control plans whereby the landslide zoning initiates a more detailed assessment at site scale. In this case, the use of landslide susceptibility mapping or preliminary hazard mapping which defines a planning control area may be sufficient to identify where more detailed landslide risk assessment is needed.

The sequent chapter focuses on the applicability of landslide **susceptibility zoning**.

3.2 LANDSLIDE SUSCEPTIBILITY

As already said, the purpose of landslide susceptibility maps is the identification of areas threatened by present and potential slope instability. Their reliability depends mostly on the amount and quality of available data used as well as on the selection of the appropriate methodology or susceptibility and hazard assessment. The availability of data determines the type of analysis that can be performed.
On the other hand, the working scale also affects the quality of the results (Van Westen, 1994).
To be profitably used for urban planning and development, the maps must be performed at an appropriate scale, in order to avoid controversy in delivering building permits, expropriation and compensating measures (Leroi, 1996).
Actually, before starting a study, an earth scientist should be aware of the desired degree of detail of the map, given the requirements of the study. When a degree of detail and a working scale have been defined, the cost-effectiveness of obtaining input data must be considered (Naranjo et al., 1994).

3.2.1 Scales

In preparing a susceptibility map, the influence that a certain number of factors exercise on the slope instability must be evaluated. The more detailed is the map, the greater will be the number of factors to be considered. The working scale is one of the first points to define. Different choices depend on the working scale, such as the adopted methodology, the factors to be selected, the mapping unit, etc. Referring to Soeters and van Westen(1996) and to Cascini et al., (2005), landslide zoning scales are suggested as summarized in the table below.

Table III. Landslide zoning mapping scales (International Guidelines ; JTC-1, 2007)

Scale Description	Indicative Range of Scales	Typical Area of Zoning
Small	< 1:100,000	>10,000 km^2
Medium	1:100,000 to 1:25,000	1000 – 10,000 km^2
Large	1:25,000 to 1:5,000	10-1000 km^2
Detailed	> 1:5,000	Several hectares to tens of km^2

In practical terms the scale of mapping may be controlled by the scale of the available topographic maps. On the basis of the table, it can be observed that:

(a) The input data used to produce landslide zoning maps must have the appropriate resolution and quality. Generally speaking, the inputs to the zoning should be at larger scales than the zoning map, not smaller. Reliable zoning cannot be produced if, for instance, a landslide hazard zoning map prepared at a scale of 1:5,000 is based on a 1:25,000 geomorphological or topographic maps because the accuracy of boundaries will be potentially misleading.

(b) The use of larger scale zoning maps must be accompanied by a greater detail of input data and understanding of the slope processes involved.

(c) In practice, only limited detail can be shown on small, medium and even large scale maps. Most examples of municipal (local government) landslide hazard or risk zoning maps which assign a hazard or risk classification on an individual property level should be prepared at the detailed level on large scale landslide zoning maps. There are some who believe that even at detailed scale it is not technically or administratively defensible to make site specific decisions based on zoning maps, and that site specific assessment is necessary. Others believe it is possible, provided the zoning process includes ground inspection to define zoning boundaries, as was done

by Moon et al. (1992) for debris flow hazard zoning.

(d) The usefulness and reliability of small scale landslide zoning mapping is considered by some to be questionable, even for regional developmental planning.

At the intermediate scale (1:25000), landslide susceptible areas would show as input the classification, location, areal extent an possibly other geometric characteristics of each landslide, creeping zones and potential sliding; the activity classes of landslides; the areas onto which the potential sliding may travel with qualitative and/or quantitative information about the past events. Whenever possible, other information can be useful to improve the landslide characterization as those regarding the volume, the qualitative and quantitative estimation of the actual rate of movement, the data set on geotechnical aspects, triggering factors and so on. Unfortunately, many of such data are difficult to collect in a systematic way at an intermediate scale.

3.2.2 Mapping units

The evaluation of landslide susceptibility requires the preliminary selection of a suitable mapping unit. The term refers to a portion of land surface which contains a set of ground conditions which differs from the adjacent units across definable boundaries (Hansen, 1984). At the scale of the analysis, a mapping unit represents domain that maximizes internal homogeneity and between-units heterogeneity. Various methods have been proposed in literature to portion the landscape for landslide susceptibility assessment and mapping. We refer to the hypothesis adopted by Carrara, who distinguishes five different categories:

a. Grid cells;
b. Terrain units;
c. Unique-condition units;
d. Slope units;
e. *Topographic units.*

a. *Grid cells*, preferred by raster-based GIS users, divide the territory into regular squares of predefined size which become the mapping unit of reference. To each grid-cell is assigned a value of each factor taken into consideration.

b. *Terrain unit*, traditionally favored by geomorphologists, are based on the observation that in natural environments the interrelations between materials, forms and processes results in boundaries which frequently reflect geomorphological and geological differences.

c. *Unique-condition units* (Chung et al., 1995) imply the classification of each slope-instability factor into a few significant classes which are stored into a single map, or layer. By sequentially overlaying all the layers, homogenous domains (unique conditions) are singled out whose number, size and nature depend on the criteria used in classifying the input factors.

d. *Slope units*, automatically derived from high-quality DTMs, partition the territory into hydrological regions between drainage and divide lines. Depending on the type of instability to be investigated (deep-seated vs. shallow slides or complex slides vs. debris flows) the mapping unit may correspond either to the sub-basin or to the main slope-unit (right/left side of the sub-basin).

e. Slope units can be further subdivided into *topographic units*, defined by the intersections of contours and flow tube boundaries orthogonal to contours (O'Loughlin, 1986). For each topographic unit, local morphometric variables and the cumulative drainage area of all up-slope elements are computed.

The selection of the most appropriate mapping unit depends on different factors, namely the type of landslide studied; the scale of investigation; the quality, resolution, scale and type of the thematic information required; the availability of the adequate information management and analysis tools. The mapping unit is also related to the susceptibility model used.

3.2.3 Zoning methods

Susceptibility zoning maps can be developed using different **methods** available in literature. The differences between all the methods consist in the approximation done when selecting the factors which contribute to the slope instability. In general those factors arrange in order to define the susceptibility degree, expressed in a cartographic way through the susceptibility map. With the available data in place, various methods can be applied to establish levels of susceptibility. These methods can be grouped in three categories:

- *basic*
- *intermediate*
- *sophisticated.*

on the basis of the necessary activities to characterize the landslides. Details of some susceptibility zoning-activities required in order to characterize and evaluate the spatial distribution of potential landslides and their relationship to topography, geology and geomorphology are described in the table below.

Table IV. Landslide susceptibility zoning-activities required to characterize, determine the spatial distribution of potential landslides and their relationship to topography, geology and geomorphology (International Guidelines ; JTC-1, 2007)

Characterisation Method	Activities
Basic	Prepare a geomorphologic map (1)
	Prepare a landslide inventory (1)
	Prepare calculations of the % of the total landslide count for each susceptibility class, the % of the area affected by landslides for each class and the % of each class in comparison to the total study area.
	Correlate the incidence of landsliding with the geology and slope to delineate areas susceptible to landsliding.
	For regional zoning correlate the incidence of landsliding with annual rainfall or snowmelt, and / or seismic loading
	Prepare the landslide susceptibility zoning map superimposed on the topography with a suitable legend.
	Implement the data and the maps in a GIS (recommended)
Intermediate	**The same activities as basic plus**
	Obtain basic soil classifications and depths in the study area
	Classify more complex terrain units. Qualitative rating of the landslide susceptible areas based on overlapping techniques
	Develop quantitative ratings (often relative rating) of landslide susceptible areas based on data treatment techniques
	Implement the data and the maps in a GIS (recommended)
Sophisticated	**The same activities as Intermediate plus**
	Detailed mapping and geotechnical investigations to develop an understanding of the mechanics of landsliding, hydrogeology and stability analyses.
	Perform data treatment analysis (discriminate; neural networks; fuzzy logic; logistic regression; etc) and develop quantitative ratings to obtain susceptibility classes
	Perform stability analyses
	Implement the data and the maps in a GIS (recommended)

Note. (1) The landslide inventory and geomorphologic mapping should be carried out at intermediate and sophisticated levels for intermediate and sophisticated level susceptibility zoning.

Other classifications of the methods developed in literature to assess landslide susceptibility distinguish:

- *qualitative methods*
- *quantitative methods*

- *direct methods*
- *indirect methods*

Qualitative methods are subjective and portray the zoning in descriptive terms**. Quantitative** methods produce numerical values to define the susceptibility degree. The typology of adopted method is related also to the working scale. Concerning the intermediate scale (1:25000), present knowledge suggests that zoning must be produced using a qualitative approach. On the contrary, at a large scale (1:5000 or larger, as well as a site-scale) the quantitative approach must be preferred, where good and extensive knowledge is available.

Direct methods consist of the geomorphological mapping of the landslide susceptibility, based on the geomorphologist experience and knowledge of the terrain conditions (Duman et al., 2006). **Indirect** methods are based on information obtained from the interrelation between landscape factors and landslide distribution and are essentially stepwise: they require first the recognition and mapping of landslide distribution over an area, then the identification of a group of physical factors which are related to the slope instability, then the estimate of the relative contribution of the instability factors in generating slope-failures.
These **indirect methods** involve analysis techniques of various types (Soeters and van Westen, 1996):

(i) <u>*Heuristic Analysis*</u>
In heuristic methods the expert opinion of the person carrying out the zoning is used to assess the susceptibility. They are based on the a priori knowledge of all the causes and instability factors of landsliding in the area under investigation. The instability factors are ranked and weighted according to their assumed or expected importance in causing mass movements (Carrara et al, 1995). These methods combine the mapping of the landslides and their geomorphologic setting as the main input factors for assessing the hazard. Two main types of heuristic analysis can be distinguished: geomorphic analysis and qualitative map combination.
In *geomorphic analysis* the susceptibility is determined directly by the person carrying out the study based on individual experience and the use of reasoning by analogy. The decision rules are therefore difficult to formulate because they vary from place to place.
In *qualitative map combination* the person carrying out the study uses expert knowledge to assign weighting values to a series of input parameters. These are summed according to these weights, leading to susceptibility and hazard classes. These methods are common, but it is difficult to determine the weighting if the input parameters. The principal disadvantage of these methods is their dependence on how well and how much the investigator understands the geomorphological processes acting upon the terrain.

(ii) <u>*Knowledge based analysis.*</u>
Knowledge based analysis or heuristic ‘data mining’ is the science of computer modeling of a learning process (Quinlan, 1993). The data mining learning process extracts patterns from the databases of landslides (Flentje et al 2007). Pixels with

attributed characteristics (from the input data layers) matching those for known landslides are used to define classes of landslide zoning. The percentage distributions of landslides within the zones are then used to help define the zones.

(iii) *Statistical analysis.*

The statistical or probabilistic approach is based on the functional relationships between some of the main factor that contribute to the occurrence of slope failure, such as steep slope or presence of weak lithological units, and the past distribution of landslides (Carrara et al, 1995). This approach usually involves the mapping of the existing landslides, the mapping of a set of factors that are supposed be directly or indirectly linked to the stability of the slopes, and the establishment of the statistical relationships between these factors and the instability process.

A statistical model of slope instability is built on the assumption that the factors which caused slope-failure in a region are the same as those which will generate landslides in future.

Hence susceptibility zoning is conducted in a largely objective manner whereby factors and their interrelationships are evaluated on a statistical basis. Various methods exist for the development of the rules for and relationships between variables and these include bi-variate analysis (Brabb et al. 1972), multivariate analysis, particularly the discriminant analysis (Neuland, 1976; Carrara, 1983; Carrara et al. 1995), Boolean approaches using logistic regression (Atkinson and Massari, 1998; Dai and lee, 2001; Ayalew and Yamagishi, 2005), Bayesian methods using weights of evidence and neural networks (Gomez and Kavzoglu, 2005; Lee et al. 2006).

These methods can be used when a large amount of information is available, in form of both quantitative and qualitative data.

The principal advantage is the objectivity of the model, while the greater disadvantage is the cost of acquisition of some factors that are related to the slope instability. Limitations with such methods result from data quality such as errors in mapping, incomplete inventory and poor resolution of some data sets as the models are essentially data trained. In addition, the results of such models are not readily transferable from region to region. Moreover, all the statistical methods are very sensitive to the type and quality of the factors chosen for the susceptibility analysis.

(iv) *Deterministic Analysis.*

Deterministic methods are used in order to study the instability of a slope in a detailed study. They apply classical slope stability theory and principles such as infinite slope, limit equilibrium (e.g. Bishop, Sarma etc) and less commonly finite element and 3-D techniques and they take in account physical laws controlling slope instability; it is a great advantage respect to the other methods.

These models require standard soil parameter inputs such as soil thickness, soil strength, groundwater pressures, slope geometry etc. The resultant map details the average factor of safety and boundaries while susceptibility and hazard classes can be set according to factor of safety ranges (i.e. inunfailed<1.0, metaunfailed 1.0 to 1.1 etc). One-dimensional deterministic slope stability models have been used to calculate average safety factors of the slopes (van Westen and Terlien, 1996; Zhou et al. 2003) in which a hydrological model may also be incorporated (Montgomery and Dietrich, 1994; Montgomery et al. 1998). Three-dimensional deterministic models integrated in a GIS have been performed by Xie et al (2004). Deterministic distributed models require maps that give the spatial distribution of the input data. The variability of input data can be further used to calculate probability of failure in conjunction with return periods of triggers (Savage et al. 2004, and Baum et al. 2005).

The main advantage of these methods consists in their great reliability, if the input data are correct. However, the large amount of data that is required can only be afforded in the case of individual slopes or small areas, so they are not suitable when analyzing large areas.
Moreover, another problem with these methods is the oversimplification of the geological and geotechnical model, and difficulties in predicting groundwater pore pressures and their relationship to rainfall and/or snow melt.

As already said, the susceptibility models and the mapping units are conceptually and operationally interrelated (Carrara et al., 1995). According to the notable examples that can be found in literature, it is observed that grid-cells are preferred for heuristic (Mejìa-Navarro et al., 1994), statistical (Carrara, 1983; van Westen, 1994) and physical or simulation (Mark, 1992; Terlien et al., 1995) modeling. Unique-condition units have been applied to both heuristic (van Westen, 1993) and statistical (Carrara et al., 1995; Chung et al., 1995) methods. Slope-units and topographic units have been used in statistical (Carrara et al., 1991; 1995) and physically based (Montgomery and Dietrich, 1994) models.

The selection of the most appropriate zoning **method** depends on several factors. Table V summarizes the activities necessary to map the existing landslide and to assess the areas with a potential to experience landsliding in the future, by relating methods, input data and procedures.

Table V. Landslide zoning methods and procedures

Method	**Input** / **Procedure**	**Topography, Landslide Inventory, Geology, Geomorphology**	**Adding Soil classification and depth, Terrain units**	**Adding Hydrogeology and Geotechnics**
Basic	Heuristic or empirical methods	*		
Intermediate	Statistical analyses	*	*	
Sophisticated	Deterministic (physically based or geotecnical) models	*	*	*

For instance, methods using heuristic or empirical procedures (Brabb, 1984; Cascini et al., 2005; Evans and King, 1998; Hungr et al., 2005; Nilsen et al., 1979) to process essentially topographic, geological and geomorphological data are considered basic methods for inventory of existing landslides and characterization of potential landslides. The method can be defined as intermediate when further details on the input data and using procedures based on statistical analyses are added (Baynes and Lee, 1998; Carrara et al., 1995; Sanctana et al., 2003; van Westen, 1994). Finally, sophisticated methods necessarily need hydrogeological and geotechnical data, and deterministic or probabilistic procedures (Baum et al., 2005; Duncan, 1992; Goodman and Shi, 1985).

3.2.4 Zoning levels

On the basis of the adopted method, three different zoning **levels** can be obtained:

- preliminary
- intermediate
- advanced.

For instance, when using basic methods exclusively, only a preliminary zoning level can be obtained; while intermediate and sophisticated methods can allow the improvement of the zoning level according to the combination shown in the table VI.

Table VI. Levels of activity required for susceptibility, hazard and risk zoning levels (International Guidelines ; JTC-1, 2007)

Type of Zoning	**Susceptibility Zoning**		
	Inventory Mapping		
Zoning Level	**Inventory of existing landslides**	**Characterization of potential landslides**	**Travel distance and velocity**
Preliminary	Basic (1) (2)	Basic (1) (2)	Basic (1) Intermediate (2)
Intermediate	Intermediate	Intermediate	Intermediate
Advanced	Sophisticated	Sophisticated to Intermediate	Intermediate to Sophisticated

Notes: (1) For qualitative zoning
(2) For quantitative zoning

Table VI defines the levels of landslide inventory and susceptibility zoning in terms of geotechnical and other input data. It is important to match the level of the zoning to the required usage (purpose), the scale of mapping and in turn match these to the level of the input data.
It is not possible, for example, to produce a satisfactory advanced level hazard zoning without at least intermediate level assessment of frequency of landsliding.
If only a basic level assessment of frequency can be made, then the result will be no better than preliminary level, and there is no point in spending large resources getting the other inputs to an intermediate or, in particular, to a sophisticated level.
On the other hand, if a preliminary level hazard zoning is required then the inputs may be at the basic level.
The current practice shows that due to both the scarcity of available data and cost restrictions, only basic or intermediate inputs and methods are mostly used.

3.2.5 Purposes

Landslide zoning for land use planning can be carried out for different purposes.
It is most commonly required at the local government level for planning urban development, but may be required by state or federal governments for regional land use planning or disaster management planning. It may also be required by land developers, those managing recreational areas, or those developing major infrastructure such as highways and railways. It is the combination of having an area which is potentially subject to landsliding, and the scale and type of development of the area that will determine whether landslide zoning is needed for land use planning.
The type and level of detail of the zoning and the scale of the maps depend principally on the **purpose** to which the landslide zoning is to be applied (regional, local and site specific planning).

It will usually be appropriate to carry out landslide susceptibility zoning as a first stage in the development of landslide hazard or risk zoning for planning purposes. Staging will allow better control of the process and may reduce the costs of the zoning by limiting the more detailed zoning only to areas where it is necessary.
It should be noted that it will seldom be necessary to carry out landslide zoning at an advanced level because the costs will potentially be so much larger than the costs for intermediate level zoning and this will potentially outweigh the benefits.

It is important to link all the elements described up to now: **type of zoning, scale of zoning, zoning methods, zoning level and purpose of zoning.** The International Guidelines provide tables that explain all the connections.
With reference to the landslide susceptibility, the table below can be found.

Table VII. Landslide zoning mapping scales, methods, levels and purposes for susceptibility assessment (International Guidelines ; JTC-1, 2007)

Scale description	**Indicative range of scales**	**Zoning methods**			**Zoning levels**			**Type of zoning**	**Purpose**
		Basic	Intermediate	Sophisticated	Preliminary	Intermediate	Advanced	Susceptibility	
Small	< 1:100,000	*			*			*	Regional zoning - Information
Medium	1:100,000 to 1:25,000	*	(*)		*	(*)		*	Regional zoning - Information - Advisory
Large	1:25,000 to 1:5,000	*	*	*	*	*	*	*	Local zoning - Information - Advisory - Statutory
Detailed	> 1:5,000	[*]	(*)	*	[*]	(*)	*	[*]	Site specific zoning - Information - Advisory - Statutory - Design

Notes: * applicable; (*) may be applicable; [*] not recommended or not commonly used

At small scale, considering that only basic methods can be used (i.e., methods based on geological data and heuristic procedures), only a preliminary zoning level can be pursued and obtained. At medium scale, where statistical procedures can be used, two zoning levels may be defined. At large and detailed scale, three zoning levels are possible, respectively based on basic, intermediate and sophisticated methods.
However, the type of analysis, level and scale of zoning also depend on the complexity of the landslide features, the homogeneity of the terrain, the spatial variability of the important causal factors, geotechnical parameters and the amount of available data and expertise.

CHAPTER 4

PROPOSED APPROACH

As seen above (cfr. Chapter 2), statistical approaches are based on the analysis of the functional relationships between instability factors and the past and present distribution of landslides. Several methods exist for establishing this correlation. The most simple consists of overlapping maps of the instability factors with the landslide inventory map (Nilsen at al., 1979). More sophisticated approaches of overlapping incorporate weighting procedures of the instability factors (Yin and Yan, 1988; Bonham-Carter et al., 1990; Chung and Fabbri, 1993; Chung and Leclerc, 1994).
The main data treatment approach used in landslide susceptibility analysis is the **multivariate analysis.**

Multivariate analysis is one of the most sophisticated techniques for landslide susceptibility assessment. In the multivariate analysis, slope failure is considered as the result of several interrelated environmental factors that can vary in space and time and, as all the data treatment approaches, this technique is very sensitive to the type and quality of the factors chosen for the susceptibility analysis. Multivariate analysis allows the estimation of the relative weight of each contributing factor by means of statistical procedures such as discriminant analysis, linear and logistic regression, and neural networks. The general linear model assumes the form:

$$L = B_0 + B_1 X_1 + B_2 X_2 + B_3 X_3 + ... + B_m X_m + \varepsilon \qquad (3)$$

where L is the presence/absence (or the area percentage) of landslides in each sampling unit, the *X*'s are input predictor variables (or instability factors) measured or observed for each mapping unit, the *B*'s are coefficients estimated from the data through techniques which are dependent on the statistical tool selected (multiple regression, discriminant analysis, etc), and ε represents the model error.

When different populations are available (in the landslide susceptibility assessment they are the unfailed and failed slopes) and only one group of variables, the most suitable technique is the discriminant analysis.

4.1 THE DISCRIMINANT ANALYSIS

4.1.1 Generality

Discriminant analysis is a statistical technique for classifying individuals or objects into mutually exclusive and exhaustive groups on the basis of a set of independent variables.
The basic discriminant analysis methodology can handle either two groups or multiple groups, although in the latter case interpretation of results can be somewhat difficult.
Thus we will restrict the discussion to the two-group problem.
We consider the case of a qualitative (i.e., categorical) <u>dependent</u> variable and a set of <u>independent</u> variables. We view the categorical dependent variable as a grouping factor that places each individual or object in the sample in one and only one of two defined groups.

Having assigned individuals or objects to one of the two groups, the objective is to identify any differences between average group score profiles. In other words, we want to discriminate among the two groups on the basis of the observed scores on the set of independent variables.
This is accomplished by maximizing the between-group variance relative to the within-group variance, searching for a linear combination of the independent variables that minimalizes the probability of misclassifying individuals or objects into their respective groups.
Discriminant analysis involves deriving linear combinations of the independent variables that will discriminate between the two defined groups in such a way that the misclassification error rates are minimized.
Discriminant analysis can be thought of in terms of a rather simple "scoring system" that assigns to each individual or object in the sample a score that is essentially a weighted average of the individual's or object's values on the set of independent variables. Once a score is determined, it can be transformed into an a posteriori probability that gives the likelihood of the individual or object belonging to each of the groups.

A graphical representation of the two-group problem will help further understanding of how discriminant analysis works. Figure 18 shows a scatter diagram and the projection of the discrminant function Y in the sample space. The figure depicts two groups, A and B, and two independent variables, X_1 and X_2. The discriminant score for each individual is obtained by multiplying the discriminant weight associated with each independent variable by the individual's or object's value on the independent variable and then summing over the set of independent variables. That is, $Y=b'X$, where Y is a 1×2 vector of discriminant scores, b' is a $1\times p$ vector of discriminant weights and X is a $p\times2$ matrix containing the values for each of the n individuals on the p independent variables.
The upper part of the figure presents the scatter plot of X_1 and X_2 for each individual or object in the two groups. The ellipses drawn around the dots, which denote group-A members, and the one around the circles, which denote group-B members, would enclose some pre-specified proportion, say, 95% or more, of the sample observations in each group. The straight line trough the two points where the ellipses intersect defines a line that, when projected onto the new Y-axis, produces an overlap between the univariate distributions A' and B', that is smaller than for any other line drawn through the ellipses. Note that the new Y-axis expresses the two-variables profiles of group A and group B as single discriminate scores. Thus, by finding a linear combination of the original observed variables X_1 and X_2, the sample observations are themselves projected onto the new Y-axis as discriminant scores, and consequently we have arrived at a much more condensed version of the between-group differences.

The approach to finding linear combinations of the original observations that will discriminate between the a priori defined groups has been described as being "best". By "best" we mean that the misclassification error rates obtained with this approach are smaller than can be obtained with any other linear combination.
However, the optimality of the approach described is conditional upon certain assumptions:

- The p independent variables must have a multivariate normal distribution.

- The $p\times p$ variance-covariance matrix of the independent variables in each of the groups must be the same.

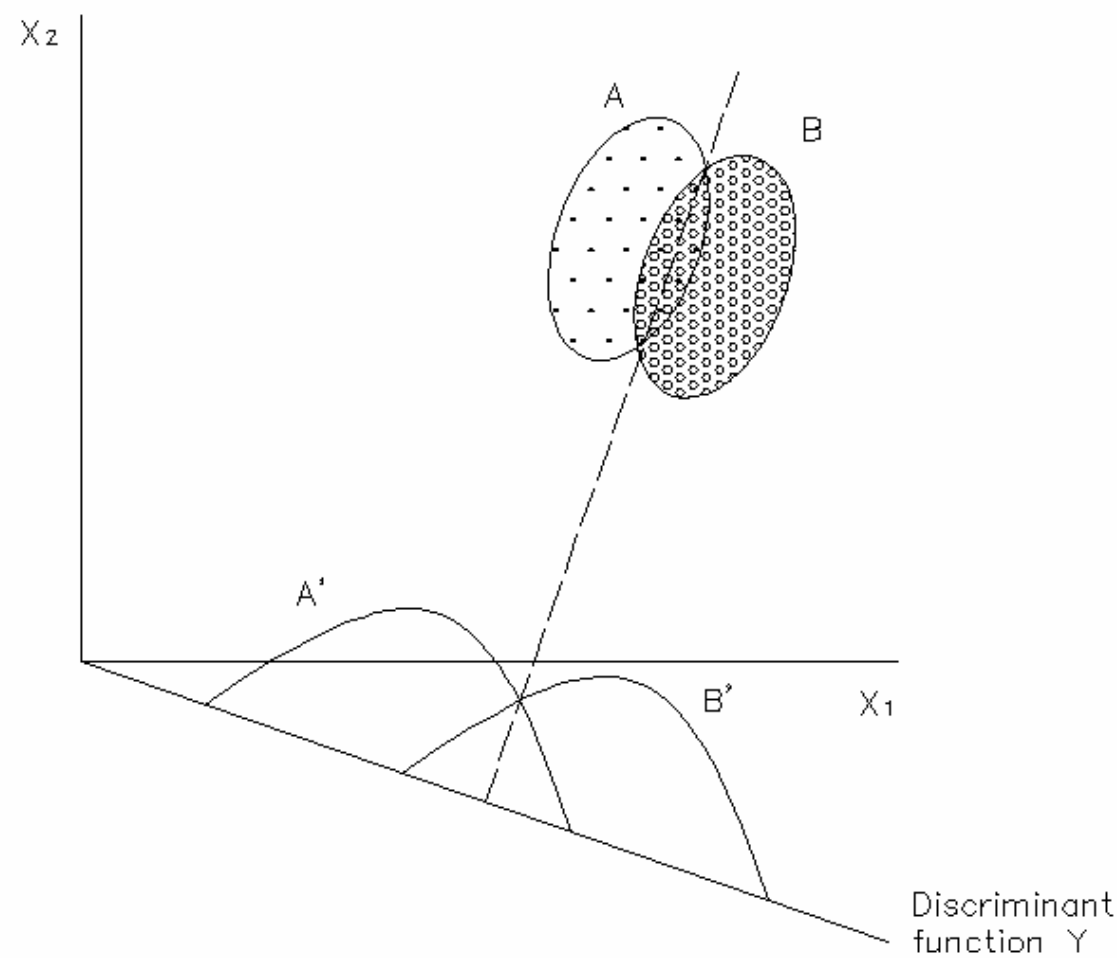


Figure 18. Discrminant function Y in the sample space (Dillon and Goldstein, 1986)

4.1.2 The two-group problem – Fisher's approach

Suppose a population G is made up of two groups G_1 and G_2. A measurement X consisting of p characteristics is observed from G. Our task is to develop an assignment rule for x that will allocate this observation to G_1 or G_2. To assist in defining a rule, the researcher has access to n observations, of which n_1 are from G_1 and n_2 are from G_2.

The earliest approach to the two-group problem was suggested by **Fisher** (1936). Under the assumption that the true mean vector for G_i is μ_i (i=1,2) and that the variance-covariance matrices Σ_1 and Σ_2 have a common value Σ, Fisher suggested finding a linear combination of X so that the ratio of the difference in the means of the linear combinations in G_1 and G_2 to its variance is maximized. In other words, denoting the linear combination by Y=b'X, we wish to find a vector of weights b so that we maximize the criterion:

$$\Delta = \frac{b'\mu_1 - b'\mu_2}{b'\Sigma b} \qquad (4)$$

It is not difficult to show that b is proportional to $\Sigma^{-1}(\mu_1 - \mu_2)$. The linear combination, therefore, is not unique; only ratios of the coefficients are. Thus, any set of coefficients can be multiplied by any constant.

In applications the parameters will usually not be known. Hence the samples of n_i observations from each G_i are used to define a sample-based rule by replacing μ_i with $\bar{x}_i$, the estimated mean vector in G_i, and Σ with S, the pooled sample variance-covariance matrix. These estimates are given by:

$$\bar{x}'_i = (\bar{x}_{i1}, \bar{x}_{i2}, ..., \bar{x}_{ip}) \qquad (5)$$

i=1,2, and

$$S = \frac{1}{n_1 + n_2 - 2}(x'_1 x_1 + x'_2 x_2) \qquad (6)$$

where $\bar{x}_{ij} = \sum_{l=1}^{ni} X_{jl} / n_i$, i=1,2, j=1,2,…,p, and where x_1 is the $p \times n_1$ matrix of observations in deviation form taken from G_1, and x_2 is the $p \times n_2$ matrix of observations in deviation form taken from G_2. Replacing parameters with their respective sample-based estimates shows that

$$\hat{b} = S^{-1}(\bar{x}_1 - \bar{x}_2) \qquad (7)$$

where S^{-1} is the inverse of the pooled sample variance-covariance matrix.

To summarize, the discriminant function is a linear composite of the original measurements on which the sum of squared differences between group means is maximal, relative to the within-groups variance.

4.1.3 Classification rules

With Fisher's linear discriminant function the assignment rule that allocates individuals or objects to the two groups becomes the following:

⇒ Assign an individual or object with realized scores x on the p independent variables to G_1 if

$$\left|\hat{b}'(x - \bar{x}_1)\right| \le \left|\hat{b}'(x - \bar{x}_2)\right| \qquad (8)$$

or to G_2 if

$$\left|\hat{b}'(x - \bar{x}_2)\right| < \left|\hat{b}'(x - \bar{x}_1)\right| \qquad (9)$$

It can be shown that if in each group the observed scores on the p independent variables are multivariate normal with mean μ_i (i=1,2) and variance-covariance matrix Σ, and if we can further assume the prior probabilities of group membership and cost of misclassifying an individual or object that actually belongs to group 1 (2) into group 2 (1) are equal, then the rules given above are sample estimates of rules that minimize the probability of misclassification.

An equivalent version of the classification rules given above uses the midpoint

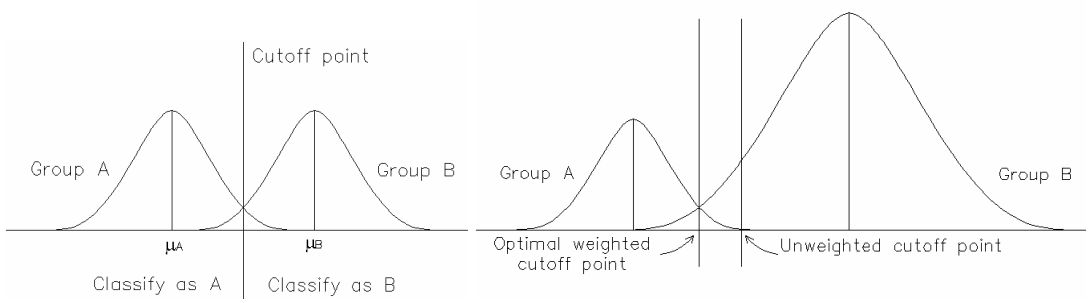
$$Y_c = \frac{\hat{b}'\bar{x}_1 + \hat{b}'\bar{x}_2}{2} = \frac{1}{2}\left(\bar{x}_1 - \bar{x}_2\right)' S^{-1}\left(\bar{x}_1 + \bar{x}_2\right) \qquad (10)$$

as the point of separation.

With unequal sample size ($n_1 \neq n_2$) the point of separation becomes:

$$Y_c^* = \frac{n_2\bar{Y}_1 + n_1\bar{Y}_2}{n_1 + n_2} \qquad (11)$$

The effects of unequal sample size on the classification rules can be better understood by examining figures 19a and 19b. The critical cut point for equal samples size in each group is shown in the figure 21a. The effect of one group being larger than the other is shown in the figure 21b.



Figures 19a and 19b. Differences between equal and different samples size (Dillon and Goldstein, 1986)

It is clear from the figures that if group A is much smaller than group B, then the optimal cut-point score will be closer to the centroid of group A than it is to the centroid of group B. Moreover, if the difference in sample size is ignored, the use of unweighted cut point results in perfect classification in group B, but substantially misclassifies members of group A.
In order for a classification rule to be optimal, it must consider both the prior probabilities of group membership and the cost of misclassifying individuals or objects into the wrong group. Let:

$P(G_i)$ = the unconditional (prior) probability of an individual or object belonging to the i group, where i=1,2.
C_{ji} = the cost of misclassifying an individual or object into group j when the actual group membership is group i.

The classification rules becomes now:

$\Rightarrow$ Assign an individual or object with scores x on the p independent variables to G_1 if

$$\frac{\left|\hat{b}'(x - \bar{x}_1)\right|}{\left|\hat{b}'(x - \bar{x}_2)\right|} \geq \ln k \qquad (12)$$

and to G_2 otherwise, where

$$k = \frac{P(G_2)C_{12}}{P(G_1)C_{21}} \qquad (13)$$

4.1.4 Variable contribution

If there is significant between-group spread in mean score profiles, then it is of interest to determine the "contribution" of each of the predictor variables to the overall discrimination.
When many predictor variables are available, the common practice is to allow some **stepwise** selection procedure to determinate which variables should enter into the discriminant function. Stepwise algorithms utilize partial F-values to select important predictors.
As known, univariate F-values associated with each predictor are a poor barometer of a variable's importance, because they ignore interrelationship among the variables, and are therefore confounded by variables that contribute redundantly to the overall discrimination.
For this reason, covariance-controlled partial F-values must be obtained.
The partial F-values approach essentially computes for each predictor variable a one-way test analysis of covariance where the covariates are the variables that are in the discriminant function at a given step.
The process works as follows. First, single predictor variable F-values are computed, treating each variable as though it were the only predictor available. The predictor with the largest F-value is then chosen to enter the discriminant function. Successive steps add (or delete) new predictors on the basis on their computed F-values conditioned on those predictors already made part of the system.

4.2 APPLICATIONS OF DISCRIMINANT ANALYSES

Discriminant analysis has been used in a wide variety of disciplines. All the applications involve the assignment of objects into mutually exclusive and exhaustive groups. The ability to classify observations correctly into their constituent groups is an important performance measure governing the success or failure of a discriminant analysis. In addition to classification accuracy, however, we will also be particularly concerned with determining the dimensions on which the groups differ. In this regard, the linear combinations of measurements used to define the assignment rule will also prove to be useful. Thus, to summarize, in discriminant analysis we are concerned with both *prediction* and *explanation*.
In this case, the discriminant analysis has been used to study **landslide susceptibility**.

4.2.1 Introduction to the problem

As seen in the Chapter 1, it is possible to use GIS technology to create landslides susceptibility map. The advantages of this technique compared to traditional ones are the quickness and the automatic capture of factors related to slope instability.
The initial work of acquiring information dealing with slope instability (factors which encourage instability) could be reduced through a GIS-based method, if we have an accurate DEM. Consequently, the first step is to create a Digital Elevation Model with high precision and resolution adequate to the dimension of movements in the study area, so to obtain some of the factors automatically with simple algorithms.

A Digital Elevation Model (DEM) is an ordered array of numbers that represents the spatial distribution of elevations above some arbitrary datum in a landscape. When discussing the use of DEMs it is important to consider the way in which the surface representation is to be used. The ideal structure for a DEM may be different if it is used a structure for a dynamic, hydrologic model than if it is used to determine the topographic attributes of the landscape. There are three principal ways of structuring a network of elevation data for its acquisition and analysis (Moore et al., 1991). **Triangulated Irregular Networks** (TINs) usually sample surface specific-points, such as peaks, ridges and breaks in slope, and form an irregular network of points stored as a set of x, y and z coordinates together with pointers to their

neighbors in the net (Peucker et al., 1978; Mark, 1975). The elemental area is the plane joining three adjacent points in the network and is known as a facet. **Grid-based methods** may use a regularly-spaced triangular, square, or rectangular mesh or a regular angular grid. The choice of a grid-based method is related primarily to the scale of the area to be examined. The data can be stored in a variety of ways, but the most efficient is as z coordinates corresponding to sequential points along a profile with the starting point and grid spacing also specified. The elemental area is the cell bounded by three or four adjacent grid-points for regular triangular and rectangular grid-networks, respectively. **Contour-based** methods consist of digitized contour lines and are stored as Digital Line Graphs (DLGs) in the form of x, y coordinate pairs along each contour line of specified elevation.

The most widely used data structures consist of square-grid network because of their ease of computer implementation and computational efficiency (Collins and Moon, 1981). However, they do have several disadvantages, including: (1) they cannot easily handle abrupt changes in elevation; (2) the size of grid mesh affects the results obtained and computational efficiency (Panuska et al.,1990); (3) precision is lacking in the definition of specific catchments areas.
Topographic attributes such as slope, specific catchments area, aspect, plan and profile curvature can be derived from all the three types of DEMs. However, the most efficient DEM structure for the estimation of these attributes is generally the grid-based method.

In this work we want to evaluate the landslide susceptibility through statistical multivariate analysis developed in GIS technique (ARC/INFO by Esri), following these successive steps:

a. Automatic or semiautomatic capture of parameters related to slope instability.
b. Statistical treatment of data.
c. Development of statistical procedures for two different cases (first time failure landslides and reactivations).
d. Definition of methods for the validation and/or evaluation of the reliability of the obtained results.

This work has been organized first considering the statistical multivariate procedure used by Baeza&Corominas (2001) to study shallow landslide susceptibility and then describing two attempts in applying the same technique in the study area to assess large landslide susceptibility. The objectives are discussed in detail below.

4.2.2 Objectives of work

Description of the methodology – Shallow landslides in Spanish Eastern Pyrenees

Different methods for studying the susceptibility of the terrain to host shallow landslides have been developed through years.
For the evaluation of the susceptibility of the terrain to develop first time landslides, we present the methodology proposed by Baeza and Corominas to study the shallow landslides in the Spanish Eastern Pyrenees (2001).
The different steps are described below:

- Collection of the necessary variables according to their suitability for susceptibility analysis.
- Statistical treatment of data through multivariate analysis and discriminant analysis, using the SPSS Inc. (1988) statistical package.
- Creation of the susceptibility map using the results obtained by the multivariate analysis.

Analysis of the activity state – Large landslides in Italian Benevento Province

The previous point focuses on predicting and mapping where new landslides will occur within a landslide-prone area, as performed in the majority of the landslide literature (Carrara et al. 1991; Soeters and van Westen, 1996; Guzzetti et al., 1999).

However, several examples, notably in Europe, show that reactivation of formerly slipped masses is a more frequent phenomenon than the occurrence of new landslides, therefore representing a higher hazard (Ardizzone et al., 2005; Catani et al., 2005). Yet, very few studies have yielded prediction maps of landslide reactivation hazard (Chung and Glade, 2004). In this case we want to apply statistical procedures to study landslides susceptibility to reactivation in the Italian Benevento Province, on the basis of the 1:25000 Inventory Map developed provided by the National Authority of Basin of rivers Liri, Garigliano and Volturno. The specific objectives are:

- Collection of the necessary variables according to their suitability for susceptibility analysis.
- Analysis of the possibility of using multivariate analysis to characterize the landslides according to their activity state, in particular to distinguish active and dormant landslides.
- Evaluation of the possible errors.

Perspectives – Large landslides in Italian Benevento Province

The same procedure has been employed to study the susceptibility to large landslides of the territory of the Italian Benevento Province. In particular, the statistical technique has been applied to distinguish stable and unstable slopes.

The scheme representing the objectives of this work is described below.

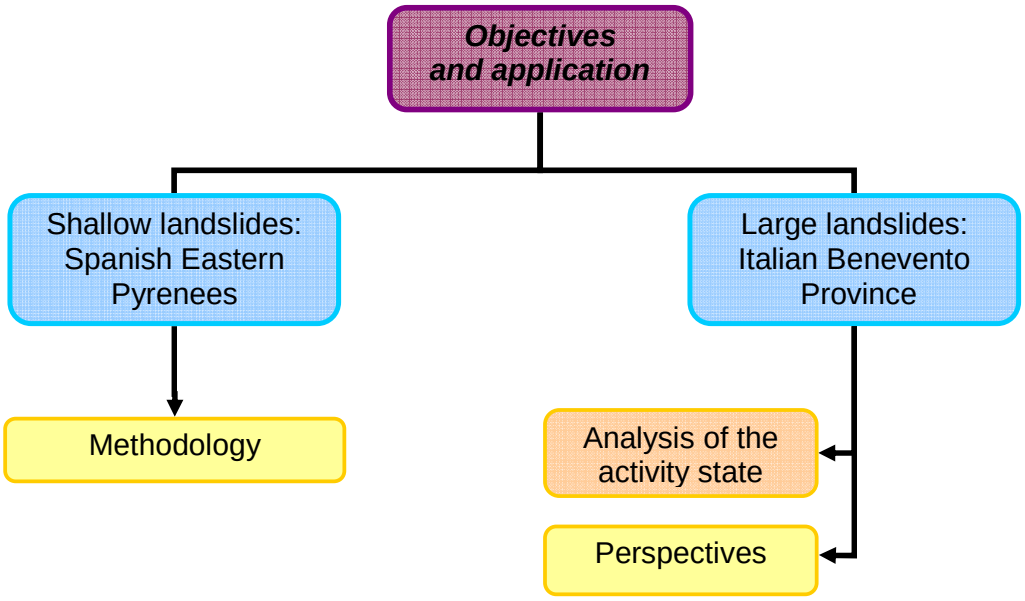


Figure 20. Objectives and applications of this thesis research

4.3 METHODOLOGY

In order to follow the predefined objectives, it is possible to delineate a common work-line, based on the application of the discriminant analysis to study landslide susceptibility.
By this analysis, each considered element of the population takes the value given by the discriminant function. Ideally, the two populations (failed and unfailed slopes or active and dormant landslides) should cluster around different values of the discriminant function.

This approach consists in the selection of a combination of independent variables that minimize the misclassification (of the slopes previously identified as failed or unfailed or the landslides defined active or dormant). In order to highlight the differences of the two populations the variables are pooled in linear combination. The main objective of this technique is to maximize the variability between the two different groups and to minimize the variability within each group.
As seen in the 3.1.1 point, the assumptions on which the discriminant analysis is based are:

- Independence of variables which have to be normal distributed.
- Equality of the variance-covariance matrix for each populations (failed and unfailed slopes or active and dormant landslides).

On the basis of these assumptions, the results of the separation of the variables may be affected by errors.
The discriminant function is expressed by a combination of weighted variables:

$$DS = C_0 + C_2 X_2 + C_3 X_3 + ... + C_n X_n \qquad (14)$$

where *X* is a variable contributing to instability, according to their statistical significance; *C* is a coefficient estimated in such a way that variability is maximal between the two groups and minimal within each group.
Thus, for a proper discrimination between the two populations, discriminant scores (*DS*) of each group should differ as much as possible. The standardized coefficients *C* are used as the relative contribution of the variable to landslide susceptibility. The discriminant scores can be used to classify new element with unknown affiliation into one of the two groups. According to the discriminant score the element is included within a landslide susceptibility class (Neuland, 1976; Carrara, 1983a).

The classification technique is based on Bayes's rule, using a conditional probability. The percentage of elements correctly classified is seen as an index of the effectiveness of the discriminant function. A detailed description of the discriminant analysis can be found in David *et al*. (1977), Lebart *et al*. (1982) and Dillon and Goldstein (1986).

The different objectives of the proposed statistical analysis are:

- Characterization of two different populations (failed and unfailed slopes or active and dormant landslides) by using different variables.
- Identification, among the chosen variables, of the ones which better explain the landslide susceptibility of the terrain.
- Definition of the relative weight for each variable.
- Definition of the discriminant function which best characterizes the distinction between the two considered groups.

In order to achieve the objectives above, the process used by Baeza&Corominas (2001) has been followed. This process consists in several steps, as described below:

1. Creation of the sample;
2. Selection of the variables and construction of the function;
3. Testing of the discriminant function;
4. Definition of the susceptibility levels and creation of the susceptibility map;
5. Validation of the susceptibility map.

This procedure has been applied by Baeza&Corominas (2001) to study shallow landslides in Spanish Eastern Pyrenees. Regarding the application in the Italian Benevento Province, the steps from 1 to 3 constitute the common work-line to be applied in two cases of failed-unfailed slopes and active-dormant landslides. However, in the first case the analysis may be improved with the steps 4 and 5, creating a susceptibility map, which represents the proneness of the terrain to develop first-time failures.
The conceptual significance of the steps is described in more detail below.

4.3.1 Creation of the sample

First it is necessary to select the samples of the two populations (failed and unfailed slopes or active and dormant landsides). The samples are selected so that the number of individuals in each one is similar. In this way, the error of classification for each individual of the two groups is similar (Dillon and Goldstein, 1986).
After creating the sample, it must be depurated from errors. For this reason, it is recommended to realize a descriptive analysis of the selected sample, in order to describe the distribution of the variables in terms of mean, standard deviation, etc., so to identify eventual errors that must be removed.

Regarding the parameters controlling the instability, it is important to select the topographic attributes that are expected to have the greatest significance. These attributes may be primary or secondary (Moore et al., 1991).
Primary attributes are directly calculated from elevation data and include variables such as elevation and slope. **Secondary** or **Compound** attributes involve combinations of the primary attributes and are indices that describe or characterize the spatial variability of specific processes occurring in the landscape such as the stream power index.

4.3.2 Selection of the variables and construction of the function

With the depurated sample begins the statistical treatment of data, in order to define the variables to be enclosed in the discriminant function. This phase consists in the identification of the variables best conditioning the slope instability and it is developed through the steps described below:

- Transformation of variables;
- Testing for normal distribution;
- Selection of independent variables;
- Obtaining the discriminant function;

Transformation of variables

Some statistical techniques (i.e. factorial or discriminant) have difficulties in dealing with qualitative data. Even though qualitative variables can be included in discriminant analysis, optimal results are not guaranteed (Dillon and Goldstein, 1986). Therefore, it is necessary to give numerical values to qualitative variables.

Testing for normal distribution

The discriminant analysis requires the variables to be normally distributed. Thus, all the selected variables must be tested for normal distribution. Among the different tests checking the adjustment to a normal distribution, the Kolmogorov–Smirnov (K–S) test at 5% confidence level has been chosen.

In general, K-S test compares the cumulative distribution function for a variable with a specific theoretical distribution, which can be the normal one, the uniform, the exponential or the Poisson distribution. For this reason, it uses mean and standard deviation parameters of the sample.

OBSERVATION – concept of Probability

Indicating with X a certain variable and with x a general value that can be assumed, the probability P that X is between a and b is equal to the area under the Probability Density Function f between a and b:

$$P[a \le X \le b] = \int_a^b f_X(x)dx \qquad (15)$$

The Probability Density Function f satisfies the following conditions:

$$f_X(x)dx \ge 0 \qquad \int_{-\infty}^{\infty} f_X(x)dx = 1 \qquad (16a.\ 16b.)$$

The distribution of values can be described also by a Cumulative Distribution Function F, which is related to the f according to:

$$F_X(x) = \int_{-\infty}^{x} f_X(x)dx \qquad (17a)$$

$$P[a \le X \le b] = F_X(b) - F_X(a) \qquad (17b)$$

Distribution of values can also be characterized by statistical descriptors:

- Mean $\quad \bar{x} = \int_{-\infty}^{x} x f_X(x)dx \qquad (18)$
- Variance $\quad \sigma_X{}^2 = \int_{-\infty}^{x} (x-\bar{x})^2 f_X(x)dx \qquad (19)$
- Standard Deviation $\quad \sigma_X = \sqrt{\sigma_X{}^2} \qquad (20)$

The most common Probability Distribution is the *Normal Distribution* described below:

$$f_X(x) = \frac{1}{\sqrt{2\pi}\cdot\sigma_X}\exp\left[-\frac{1}{2}\left(\frac{x-\bar{x}}{\sigma_X}\right)^2\right] \qquad (21)$$

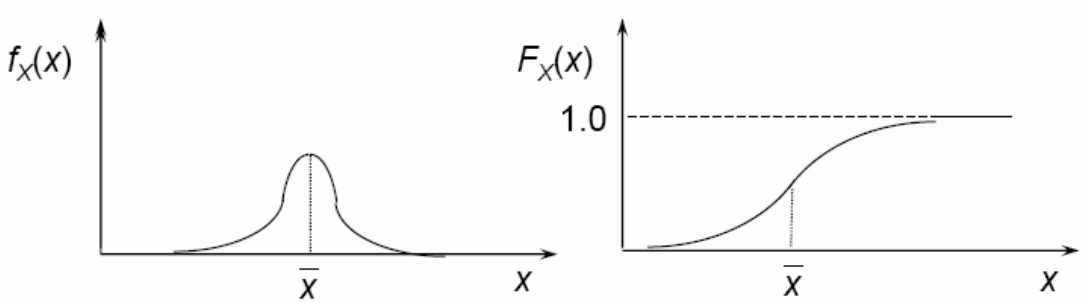


Figure 21. Normal Distribution: Probability Density Function f and Cumulative Distribution Function F

The K-S allows knowing the goodness of the adjustment of a variable's distribution to the normal one. If the adjustment is not good, a transformation of the variables is required. The transformation requires the application of a function y=r(X), so that the distribution of the new variable is more homogenous and symmetrical. After the transformation, a new KS test have to be developed, in order to verify the goodness of the results. One of the most used transformations is the lognormal. In this case the function r(X) is the log in basis 10:

$$y=\log(x) \qquad (22)$$

Hence, the normal distribution becomes a *Lognormal Distribution*, as described below.

$$f_X(x) = \frac{1}{\sqrt{2\pi}\cdot\sigma_{\ln X}} \exp\left[-\frac{1}{2}\left(\frac{\ln x - \overline{\ln x}}{\sigma_{\ln X}}\right)^2\right] \qquad (23)$$

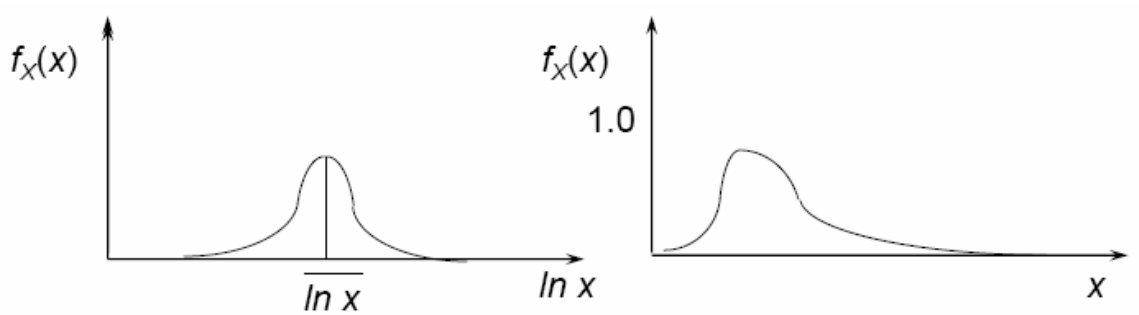


Figure 22. Lognormal Distribution: Probability Density Function *f* and Cumulative Distribution Function *F*

Selection of independent variables

Performance of a linear discriminant function is likely to be poor when dealing with populations characterized by strong correlation between variables (Dillon and Goldstein, 1986). Consequently, dependent variables must be removed. Search for possible correlation will be performed with a bivariate analysis of correlation and a principal component analysis (PCA). These techniques allow evaluating the correlation between variables. However, while the fist produces the Pearson's coefficients of correlation in a square or lower-triangular matrix, useful to characterize the dependence of each variable from all the others, the PCA provides also some insight on the structure of the overall population. Knowing how the population is structured may allow the identification of groups of elements having similar behavior, to recognize some trends within the sample and to detect correlation that is difficult to observe with a simple correlation matrix.

The PCA technique describes the dispersion of a points' cloud (variable values for each individual) in a multidimensional space, through the definition of a new system of axes (factors), so that the points' dispersion is the maximum among all the possible combinations of the original variables (David et al., 1977; Nie et al., 1981). The factors having the greater interest are those related to the greater variance.

In order to apply the PCA methodology, it is necessary to extract the minimum number of factors that explain the maximum percentage of the total variance. The weight of each variable in each factor represents the possible dependence. When a factor is defined by more than one variable, with a great saturation value (weight of the variable in the factor), a high dependence between the variables exists. On the contrary, when a factor is defined by a unique variable, with a great saturation value, a high independence of this variable respect to the others exists.

After having identified the dependent variables through the bivariate Pearson correlation and the Principal Component Analysis, before rejecting one of these, it is recommended to make a contrast analysis between the two populations (failed and unfailed slopes or active and dormant landslides), in order to identify which is the variable that best characterize each population and thus influence the separation of the two populations.
The contrast analysis is developed using two additional tests, **T-test** and **One-way test**, based respectively on the analysis of mean and variance. They provide an early understanding about the influence of each variable on stability. Only independent variables showing the highest significance in relation to the slope stability have to be selected for discriminant analysis.
With the information obtained at this step it is possible to get a first selection of variables.

Obtaining the discriminant function
From PCA and contrast analyses, a data set including the most significant independent variables was taken as input of the discriminant analysis.
Assuming that the independent variables must have a multivariate normal distribution with equal variance – covariance matrix for each of the two groups (failed and unfailed slopes or active and dormant landslides), the discriminant analysis has been applied. The objective is to create a function with the minor number of variables that separate the two populations.
The method of selecting variables for the discriminant analysis was the "stepwise" procedure, whereby variables are entered and removed one at a time from the discriminant function until the most significant model had been generated. The stepwise method selects variables by using a criterion previously defined. The used criterion, defined "Minresid", minimizes the sum of unexplained variation (residual variance) between groups previously identified by grouping variable (failed–unfailed slopes or active-dormant landslides). The statistical controls for selecting variables and for defining their entrance order in the analysis are developed through the minimum tolerance level (t=0.001) and the minimum value of F (1.0).
As a result, a linear combination of the variables having the greater significance in discriminating the two populations is obtained. The magnitudes of the standardized discriminant weights indicate the contribution of each variable to the function.

4.3.3 Testing of the discriminant function
The discriminant function has to be applied to a pilot area in order to test its performance and reliability. Several procedures exist for testing landslide prediction models:

a. selection of a random sample to build the model and use of the remaining population to verify it (Neuland, 1976);
b. derivation of models from different random sample sizes and checking whether the function coefficients change significantly (Carrara, 1984);
c. preparation of the model from landslides occurred during a specific event, and checking it with landslides triggered by a subsequent event (Luzi, 1995);
d. development of the model in a training area, and testing it in a target area with similar characteristics.

In this study, the last one will be used to validate the discriminant function.

4.3.4 Definition of the susceptibility levels and creation of the susceptibility map
Different methods for the definition of the susceptibility levels have been developed in the literature based on the scores of the discriminant function (Carrara, 1984; Baeza, 1994). The method proposed here consists in the division in ranges of the frequency distribution using its standard deviation (Baeza&Corominas, 2001). Discriminant scores are divided into several ranges, so establishing different susceptibility classes according to these ranges.

4.3.5 Validation of the susceptibility map

In order to validate the susceptibility map, an index of relative landslide density will be used (Baeza&Corominas, 2001). This index is defined by the ratio between the density of slope failures of a given susceptibility class and the overall slope failure density. The index takes the following form:

$$R = \frac{n_i / N_i}{\sum (n_i / N_i)} \cdot 100 \qquad (24)$$

where n_i is the number of slope failures within a susceptibility class and N_i is the number of cells of this class. It may be expected that slope failures will appear in cells having higher discriminant scores.

A graph showing all the phases described up to now is reported below. It represents the working scheme to be followed, with the successive steps to be developed.

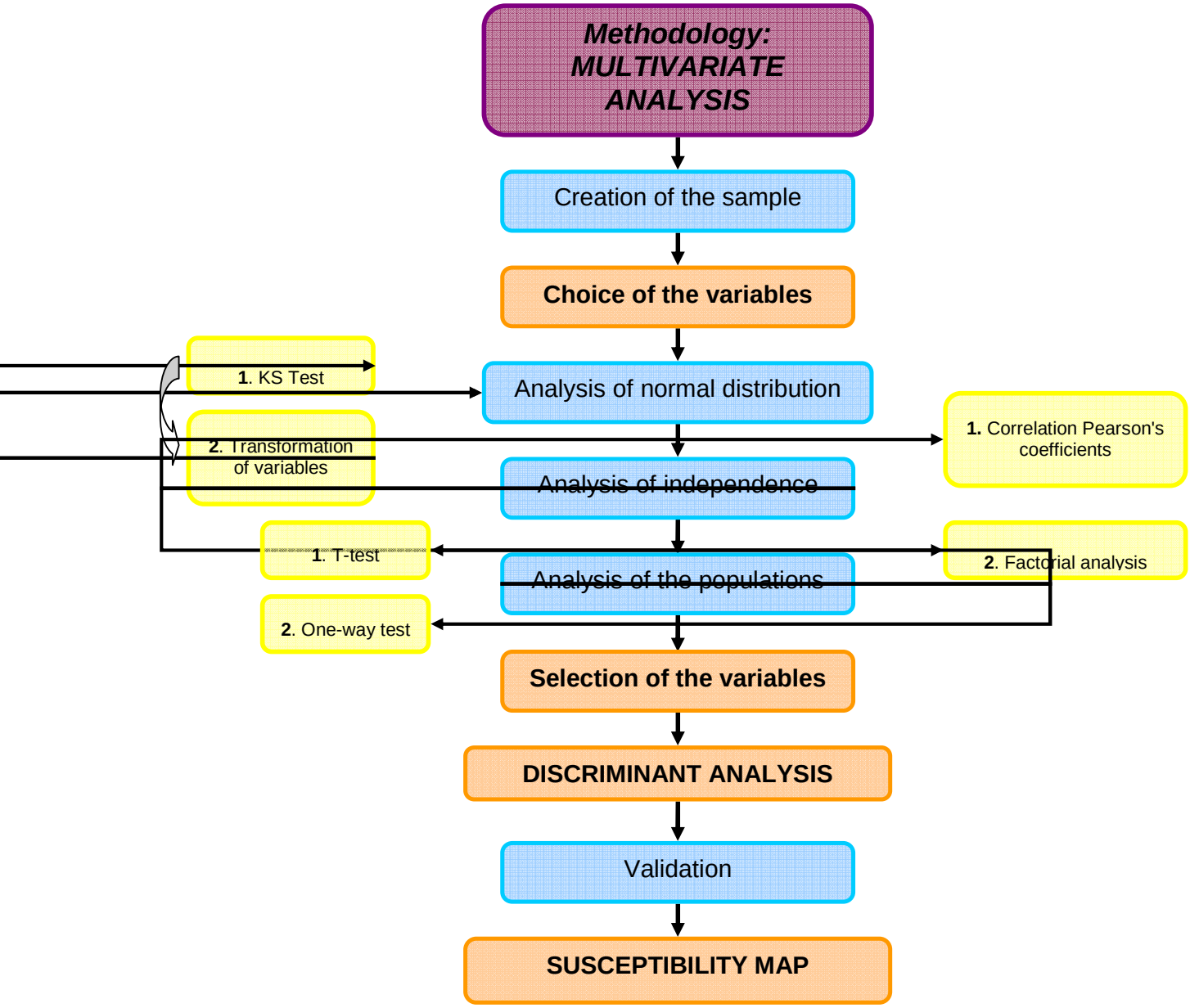


Figure 23. Working scheme to follow in applying multivariate statistical analyses

CHAPTER 5

THE CASE STUDY: LARGE LANDSLIDES

5.1 THE TERRITORY OF THE LIRI-GARIGLIANO AND VOLTURNO BASIN AUTHORITY

5.1.1 Authorities involved in landslide management

Landslides are dangerous phenomena that may cause damage and loss of lives. Regarding the aspects of forecast and prevention, many countries currently have adopted specific policies of planning in order to reduce the losses due to the occurrence of landslides.
In Italy the fundamental instrument to support any decision-making processes is the Basin Plan for the hydrogeological asset, which contains in particular the identification and perimeter of areas at hydrogeological risk.
This instrument has the aim of identifying homogeneous areas for purposes of government and territory's development through a correct planning that permits the protection of human life and the security of infrastructures, environmental and cultural heritage and the socio-economic activities, and the management of any crisis situations.
The organ responsible of the adoption of the Basin Plan for the hydrogeological asset is the Basin Authority.

The Basin Authority was established by the Italian Law 183/1989 as the competent organ on the "hydrographic basin". To find the use of the concept of basin in the L.183/89, it must be traced back to the 1960s, when after a dramatic sequence of natural disasters occurred in the 1960s, in which the disaster of Vajont (1963) and the flood of Florence (1966) are remembered, an inter-ministerial committee was set up, known as Marchi's Commission, from the name of its President, in order to identify a series of synergetic actions, programming and operational, to solve technical, economic, legislative and administrative problems connected with the defense of the soil.

At that time, the management of water resources was conducted separately from the actions of hydrogeological accommodation of soil and slopes.
The Commission's final report, completed in the early 1970s, showed instead clear and urgent need to face together all the matters relating to the defense of the soil and to the optimum use of water resources, not only with individual interventions, but especially through forms of planning able to integrate the protection and development that could only be managed by a single decision centre. The bases for the creation of an administrative structure public at scale of wide hydrographic area was posed, to which bring together all the competences for the management of the territory.

At the end of the successive decade, in a moment especially fertile from legislative point of view for the environmental issues, the L. 183/89 has reorganized the powers of central authorities of the State and local authorities in the field of defense of the soil and has set up the basin authority, giving it the task of ensuring the protection of the soil, the rehabilitation of water, the use and the management of water and the protection of the environmental aspects within the framework of the hydrographic basin.

For the first time tasks of planning and programming were attributed to an organization whose territory had been defined not by politics, but with geomorphological and environmental criteria.
In this way the attempt to overcome an administrative division that hindered, sometimes prevented, the possibility to face problems linked to water cycle and defense of the soil together and to a territorial scale adequate, became concrete.

According to L. 183/89, all the national territory was divided in hydrographic basins, with three degrees of territorial importance:

1. National basins;
2. Interregional basins;
3. Regional basins.

As regards the territory of the Campania Region, this is subdivided as follows:

Table VIII. Basin Authorities in the Italian Campania Region: importance

BASIN AUTHORITIES	IMPORTANCE
Liri-Garigliano and Volturno	National
Sele	Interregional
Destra Sele	Regional
Nord-Occidentale	Regional
Sarno	Regional
Sinistra Sele	Regional

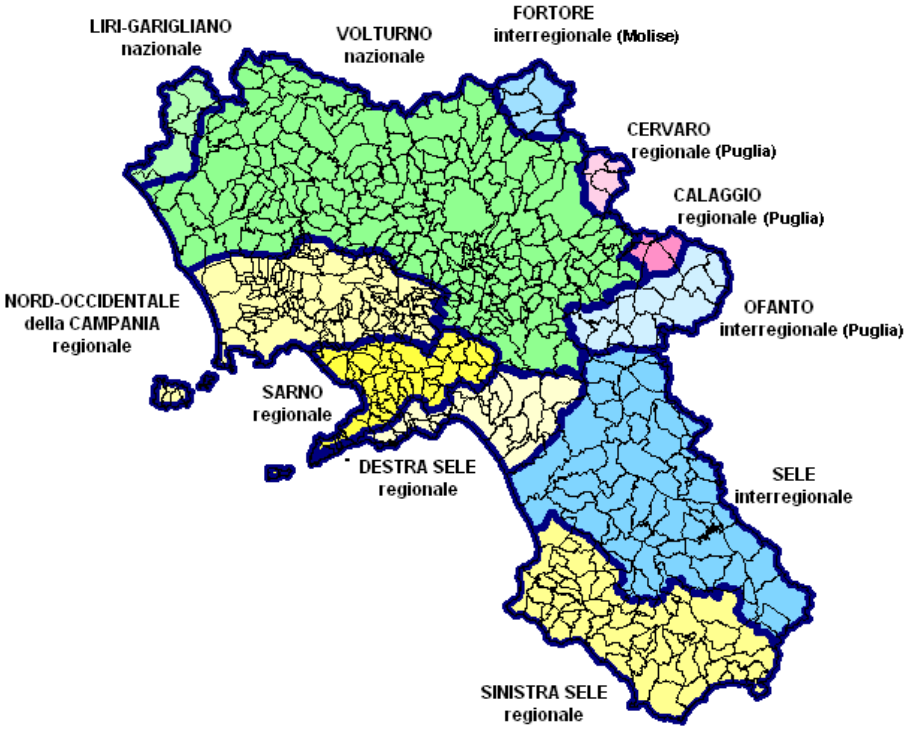


Figure 24. Basin Authorities in the Italian Campania Region: territory involved (www.sito.regione.campania.it/lavoripubblici)

Beside the setting up of the Basin Authority as the competent institution on hydrographic basin, the Italian Law 183/1989 has identified in the “Basin Plan” the cognitive, regulatory and technical-operation instrument by which the actions and the rules for the conservation, protection and exploitation of the soil and correct use of the waters are planned, on the basis of physical and environmental characteristics of the territory concerned.

The basin plan is therefore an instrument responsible to the protection of physical integrity of the territory, under its many aspects (geological, ideological, hydrogeological, urban, agricultural and environmental).
In order to reduce the timing of the editor of the entire basin plan, the Italian law n. 493 of 12-04-93 has integrated the L. 183/89 providing the possibility to draw up the basin plan for extracts concerning functional areas, which must be interrelated to the content of the basin plan.

The necessary activities for the editor of the plan may be combined substantially in three stages:

1. The informative phase. It contains the analysis of the actual situation, with the cognitive framework of the hydrographic network, the floods and the damage, the behavior of structures, the degree of vulnerability to the destabilization.

2. The programming phase. It indicates the objectives and the directives to which conform the structural and non-structural interventions, with a list of the priority interventions based on the financial resources.

3. The projective stage. It contains the technical identification of the individual priorities, with the assessment of the costs and the expected effects, with a level of detail commensurate with the importance of the work.

Subsequently, after the events that on 5th and 6th May 1998 have hit seriously some of the municipalities of the Campania Region, "urgent measures for the prevention of the hydrogeological risk" were immediately issued (DL 180/1998).

The complex legislation followed (DL 05-15-1999 n.132 and Law 365/2000) has led to the drafting of “Extraordinary Plans”, in order to cover the priority areas at hydrogeological risk for which a state of emergency has been declared.
In particular the extraordinary plans contain “the identification and the perimeter of areas where the hydrogeological risk for integrity of people and for the security of infrastructure and environmental and cultural heritage is very high”.

In this context, it is inserted the activity developed for the preparation and the editor of the Basin Plan for the hydrogeological risk (where landslides risk is distinguished from floods risk) for the territory of competence of the **National Basin Authority of rivers Liri, Garigliano and Volturno**, which relate in particular, the identification of areas at hydrogeological risk and the perimeter of the areas subject to safeguard measures.

The territory extends on an area of 11,484 km^2, including 450 municipalities. There are various regions falling within the territory of interest. In particular:

Campania Region (41.01% of territory)
Avellino Province
Benevento Province
Caserta Province
Salerno Province

Lazio Region (21.43% of territory)
Frosinone Province
Latina Province
Roma Province

Abruzzo Region (21% of territory)
L'Aquila Province

Molise Region (21.43% of territory)
Campobasso Province
Isernia Province

Puglia Region (0.1% of territory)
Foggia Province

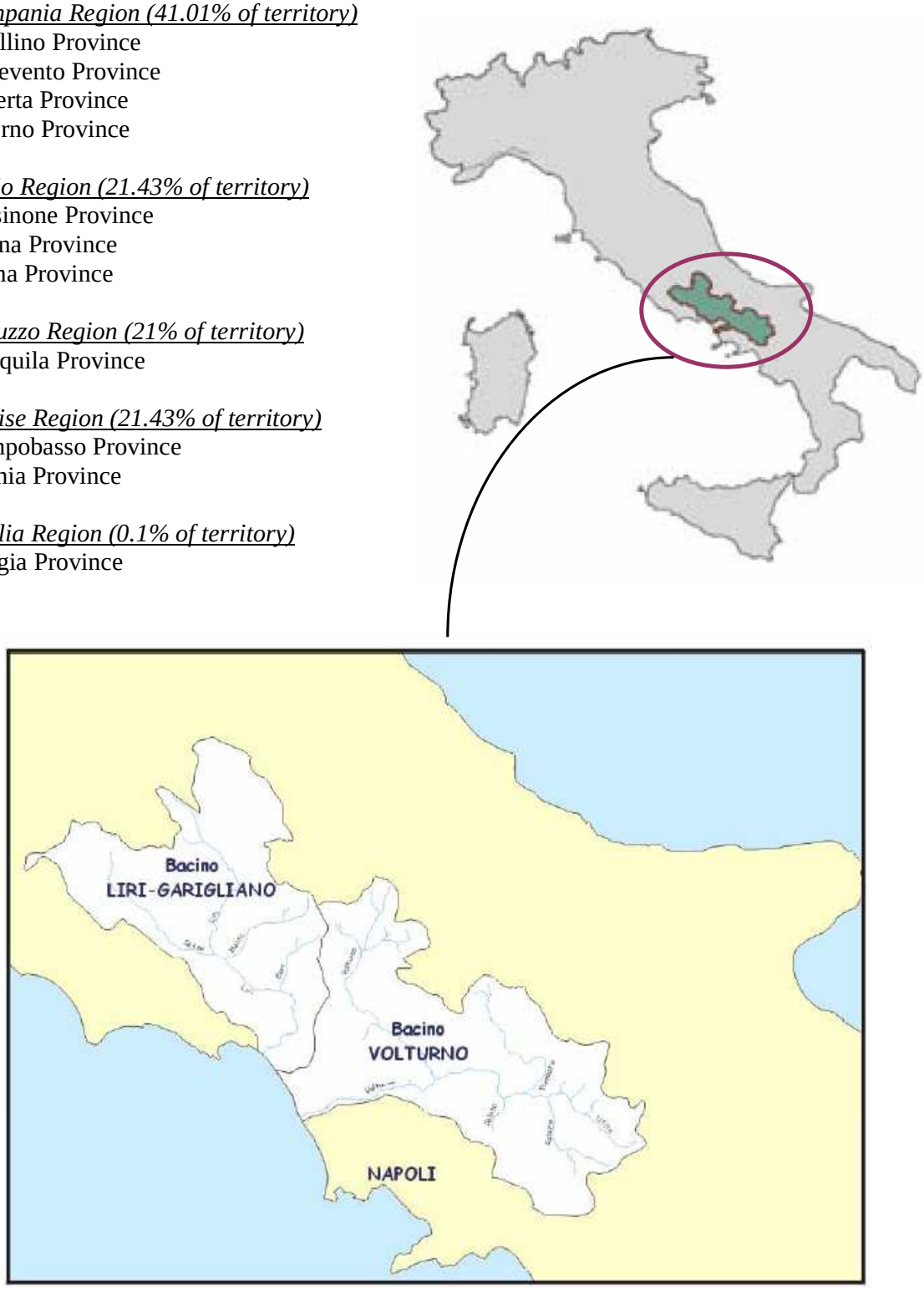


Figure 25. Location of the National Basin Authority within the Italian territory (www.autoritadibacino.it)

The Basin Plan for landslide risk in the Liri-Garigliano and Volturno Basin Authority is drawn up in continuity with several actions and specific studies developed in the years.

The process enabled to highlight the analyses of anthropogenic and physical aspects of interest (geology, geomorphological, landsliding, hydrogeology, hydraulic, urban settlements and infrastructure, damage, constraints, environmental emergencies, historic, architectural and archaeological).

For the implementation of the plan, the basin authority has executed an innovative process, from the methodological and operational point of view, in order to achieve results really significant for the purposes of planning, and the strengthening and qualifications of the property public (as defined by the Community policy and in the ongoing reform in public administration), also through the "internalization" of the technical capacities in matter of defense, use, safeguarding and physical government of the environmental system.

The assessment of landslide risk has presented difficulties for various reasons:

- the heterogeneity of the geo-environment context in which the landslides are established;
- the variety of landslides and their great diffusion in the territory;
- the articulation of the infrastructures and urban context exposed to risk of landsliding;
- the presence of numerous constraints both geological that urban.

The choice of the most appropriate approach was, therefore, subjected to the availability of basic data that help to define the various components of risk or to the concrete possibility of their adequate acquisition.
On the basis of the previous considerations, it was recommended to carry out systematic studies in many areas, preferring the homogeneous acquisition of basic data on the territory and using with caution the data provided by the competent authorities.

Therefore, the studies were aimed at defining:

- the geomorphological-structural context of the territory;
- the identification and classification of types of movement of the landslides;
- the intensity of the landslides;
- the morphological areas;
- the framework for the cognitive use of the soil for anthropogenic purposes;
- the identification and classification of areas at risk of landslide.

5.1.2 The study area

The Benevento Province territory is characterized by a great number of landslides, triggered mostly by rainfall and earthquakes. Rainfalls of high intensity trigger lots of mass movements, causing dangerous situations because may involve susceptible areas. Regarding earthquakes, results of seismic analyses on the study area show the presence of two significant conditions:

- Rockfalls may occur, mainly over steep carbonatic slopes, and may evolve in rock avalanche with particularly high disctructive power;

- Large landslides may reactivate, mainly over areas with prevalence of clayey formations.

These conditions have impacted inhabited centers and infrastructures, causing lots of damages and loss of life during the years.

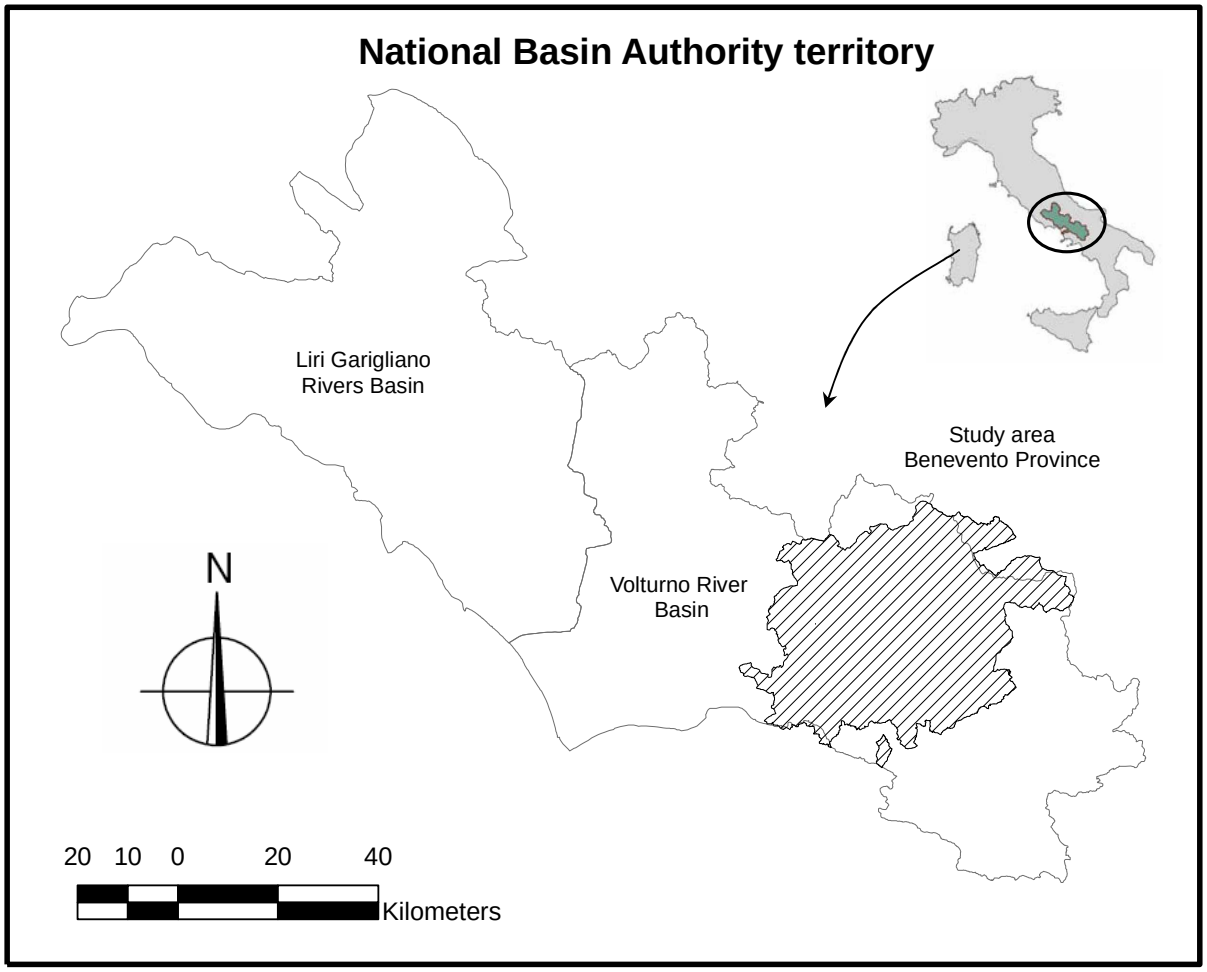


Figure 26. Location of the study area

The Benevento Province territory is collocated in one of the most geologically complex areas in Italy. The Apennine relief, considered as the result of the tectonic activity developed during Miocene Age and Pliocene Age, appears as constituted mainly of limestone and dolomitic and flysh deposits (Manfredini, 1963; D'Argenio et al., 1973; Patacca and Scandone, 1989; Sgrosso, 1998).

The Benevento Province territory is introduced in this complex geological context. The territory shows different terrains which belong to marine sedimentary successions (between the Cretaceous Age and Pliocene Age) and to continental sedimentary successions (in the Quaternary Age).

On that basis, it is possible to distinguish five main ambits, from a lithological and genetic point of view:

1. *Shelf carbonatic successions.*
   They are constituted of limestone and dolomitic limestone and are present in the western zone of the study area.
2. *Pre-orogenic basin successions.*
   Among these successions it is possible to distinguish flysh deposits, varicolored clays and arenaceous successions.
3. *Sin-orogenic and post-orogenic successions.*
   They are constituted mainly of arenaceous and arenaceous-clayey sequences and are present in the central zone and in the north-western part of the study area.
4. *Piroclastic deposits.*
   It is possible to distinguish deposits due to flowing phenomena and falling deposits; they are present in the south-western zone of the study area.
5. *Continental, fluvial and detritus deposits.*
   They are deposits of the Quaternary Age and are present mainly in the lowest part of the carbonatic relieves, as breaches and detritic deposits, and over the river valleys, as fluvial deposits.

From a morphological point of view, we can explain the influence of the described geological-structural contexts on the morphology of the area.
The western zone of the Benevento Province is characterized by the presence of high relieves, constituted of limestone successions, while the eastern part is characterized by gentle slopes. The central part of the area, instead, represents the Benevento basin, with the presence of the marine and continental clastic deposits, over which the two most important rivers flow: the Calore River and the Sabato River.

Regarding the instability conditions, it is very important to link the geology with the landslides, so to acquire information on the prevalent typologies of movements on different areas. The different geological contexts may be affected by mass movements different for typologies and cinematic characteristics, basing on the local conditions.
The analysis of this relationship all over the Benevento Province shows that the majority of the area is prone to **slow movements**, while **fast movements** can occur only in a little area.

The specific characteristics of the events recorded in the different geological contexts are below described:

1. *Rockfalls and topples.*
   In the study area these movements occur on the steep slopes of limestone relieves. The geo-mechanic conditions induce mobilized blocks, different for volume, which can be greater than various cubic meters.
   In few zones of the Province territory, where the geomorphological conditions show the presence of nearly sub horizontal relieves, rockfalls belonging to the secondary typology (Hutchinson, 1988) have been recorded.
   In other areas, first failure rockfalls have been recorded, due to particular meteoric conditions: rainfall of high intensity during the coldest winter periods. That highlights how the slope instability conditions are related not only to the pluviometric conditions, but also to the crioclastism phenomena, with the sequence of different temperature

periods, favored by the presence of intense fractured lithotypes. This condition facilitates the rock alteration, due to the continuous action of the frost and thaw mechanisms, to the tree roots action and to the pore water pressure effect, that induces slope instability conditions.

2. *Debris-flows, rock-avalanche and debris-avalanche.*

   These movements belong to the category of slope instabilities defined as "flow-like landslides" (Hutchinson, 1988), due to their characteristics similar to the flow category. These movements are characterized by rapid cinematic conditions and generally occur over the upper part of steep slope, where portions of rocks, loose or lithoidal, may move, generally after heavy rainfalls or earthquakes. These flows, from rapid to extremely rapid, are considered very dangerous over mountainous areas, because can involve also the piedmont.

   Debris-flows are common on carbonatic relieves, characterized by steep and denudated slopes. In few areas, the detached rock detritic materials create a rock-avalanche evolving in a debris-flow, which puts down in the lower part of the slope (valley area).

   Generally the mechanisms that trigger rock-avalanches may be similar to those described for rockfalls; the difference consists in the volume involved in the movement.

   However, if the material involved is prevalently detritic, the triggering mechanism is different and it is linked to the transformation of superficial detritic formations in flow, due to the occurrence of a shallow landslide.

   Debris-flows and debris-avalanches occur in the southern zone of the Province territory, involving the piroclastic deposits present on the carbonatic relieves.

3. *Earth-flows.*

   The slow-cinematic flows are defined in literature as earth-flows (Carrara et al., 1985) and represent phenomena in which the mass movement presents similar conditions to those of a viscous fluid. Actually, they consists in viscous-plastic movements, generally from slow to rapid, characterized by relative displacements in the moving mass along one or more defined surfaces.

   In the majority of the cases, the material flows into the gullies, where it is rapidly eroded. However, significant is also the presence of the lobed typology, in which the materials spread over the lower part of the slope, creating the typical fan shape.

   The greater part of the earth-flows recorded in the study area presents a translational movement and is therefore classified as "mudslides", due to the presence of a well defined shear surface. However, local conditions may induce viscous flows, characterized by a greater mobilization of material (Picarelli, 2001).

   From a morphological point of view, it is possible to distinguish three parts in this typology of movements: the alimentation part, where local rotational and translational movements occur, the flow channel, with the trajectory among which the moving masses run, and the depositional area, where materials accumulate.

   It has to be noted that the mechanisms occurring in the alimentation part are considered as first-time movements, because involve undisturbed material, while the mechanisms occurring over the channel and in the depositional area are considered as reactivations.

We present below three notable examples of earth-flows: the one occurred in Sant'Agata dei Goti in the January 1997, the one occurred in san Giorgio La Molara (1980) and the other occurred in San Bartolomeo in Galdo (1958).
In the first case, the phenomenon has involved prevalently the piroclastic formation present above a clayey flysch substrate over a gentle slope (5°-10°) and it is characterized by the presence of a unique alimentation area.
In the second case, the phenomenon is characterized by different alimentation areas involving the whole slope (multi-source event): several alimentation zones merge in a unique flow channel. This configuration is typical in areas with clayey lithotypes involved in tectonic activity, in particular under complex lithological and structural conditions.
In the last case, the lithological and structural conditions were homogenous and the landslide, even though characterized by several movements, can't be defined as a multi-source event. Actually widespread phenomena may occur, all over the area, like several and independent movements. This typology of landslides is defined as "coalescent landslides", in order to underline the spatial and temporal independence of different mass movements, due to the uniform basin conditions.

The described typologies of slope instability occur mainly on clayey and clayey-marnous lithotypes and on the upper part of slopes, characterized by high values of slope angle. In these areas, water infiltration has a great effect. Even though clayey materials are characterized by low values of permeability, the presence of discontinuity favor water infiltration within the clayey deposits. These altered formations, which may reach about ten meters, are characterized by greater impermeability rather than the deep formations: that causes the possibility to have water tables in the superficial formations, with pore water pressure reducing the mass strength over the terrain surface. This condition belongs to the triggering factors defined by Terzaghi (1936) as "internal causes".
Among the "external causes" which may trigger these types of movements there is the rivers action at the toe of the slope, that produces an increment of shear stresses along the slope, and the earthquakes action.

4. *Slides.*
In the Benevento Province territory rotational slides are not very common. It can be explained considering the particular lithological and structural conditions. Translational slides, instead, are widespread all over the study area. These movements occur prevalently on mechanic discontinuity in flysch formations or involve detritic masses on lithological discontinuity. Moreover, local rotational and translational movements are recorded in the upper part of earth-flows. In this case, the phenomena are classified as "complex" (Cruden and Varnes, 1996), because represent a movement constituted by rotational slide and earth flows. The example of movements occurred in Campolattaro is notable. The structural condition was characterized by the presence of limestone sequence, causing huge amount of water, demonstrated by several sources near the crown of the movement.

The first step for any susceptibility assessment study is the evaluation of the current situation. In this research, the basis of the study is the Landslide Inventory Map, provided by the National Basin Authority of the rivers Liri, Garigliano and Volturno. It is a detailed map, developed at 1:25000 scale all over the territory of the Basin Authority competence. The

spatial location of the landslides were extracted from aerial photographs and then transformed in this map.
In the Inventory Map, the terrain units and the existing and potential landslides are mapped.
On the Benevento Province study area, there are classified different typologies of movements, on different lithotypes. In particular, we focus on the landslides that can be considered belonging to the slow and moderate intensity classes.

Landslides are usually classified by their intensity. Considering the definition given by Hungr (1997) and reported in the International Guidelines currently being developed by JTC-1 (Joint Technical Committee on Landslides and Engineered Slopes), the landslide intensity is "a set of spatially distributed parameters related to the destructive power of a landslide."
The parameters may be different, being the maximum movement velocity the most accepted one although total displacement, differential displacement, depth of the moving mass, depth of deposited mass and depth of erosion are alternative parameters. By keeping in mind the design of protective structures, other derived parameters like peak discharge per unit width, kinetic energy per unit area, maximum thrust or impact pressure may be also considered.
Regarding the maximum movement velocity, the classification taken in account is the one provided by Cruden and Varnes (1996). This classification considers seven classes of landslides, based on their velocity.

Table IXa. Velocities and landslides consequences (Cruden and Varnes, 1996)

Class	Description	Probable Destructive Significance	Typical velocity	Velocity (m/s)
7	Extremely rapid	Catastrophe of major violence; buildings destroyed by impact of displaced material; many deaths; escape unlikely.	5 m/s	5
6	Very rapid	Some lives lost; velocity too much to permit to the persons to escape.	3 m/min	$5 \cdot 10^{-2}$
5	Rapid	Escape evacuation possible; structures, possessions, and equipment destroyed.	1.8 m/hr	$5 \cdot 10^{-4}$
4	Moderate	Some temporary or little damageable structures can be temporarily maintained.	13 m/mounth	$5 \cdot 10^{-6}$
3	Slow	Remedial construction can be undertaken during movement; intensive structures can be maintained with frequent maintenance work if total movement is not large during a particular acceleration phase.	1.6 m/year	$5 \cdot 10^{-8}$
2	Very slow	Some permanent structures can not be damaged from the movement	16 mm/year	$5 \cdot 10^{-10}$
1	Extremely slow	Imperceptible without monitoring instruments; construction possible with precaution.	<16 mm/year	$<5 \cdot 10^{-10}$

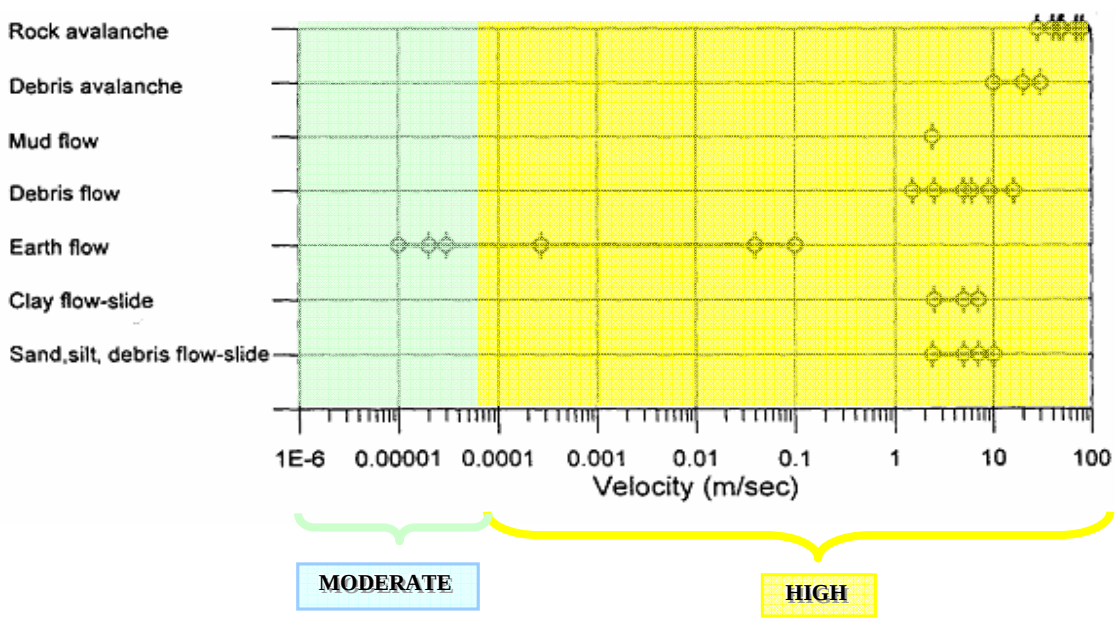
Referring to the International Guidelines, the landslide intensity can be grouped from the previous seven classes into three, individuating the categories of *high*, *moderate* and *low* intensity, as described in the table IXb.

Table IXb. Velocities and landslides consequences (Cruden and varnes, 1996; mod.)

Intensity	Class	Description	Probable Destructive Significance	Typical velocity	Velocity (m/s)
HIGH	7	Extremely rapid	Catastrophe of major violence; buildings destroyed by impact of displaced material; many deaths; escape unlikely.	5 m/s	5
	6	Very rapid	Some lives lost; velocity too much to permit to the persons to escape.	3 m/min	$5 \cdot 10^{-2}$
	5	Rapid	Escape evacuation possible; structures, possessions, and equipment destroyed.	1.8 m/hr	$5 \cdot 10^{-4}$
MODERATE	4	Moderate	Some temporary or little damageable structures can be temporarily maintained.	13 m/mounth	$5 \cdot 10^{-6}$
	3	Slow	Remedial construction can be undertaken during movement; intensive structures can be maintained with frequent maintenance work if total movement is not large during a particular acceleration phase.	1.6 m/year	$5 \cdot 10^{-8}$
LOW	2	Very slow	Some permanent structures can not be damaged from the movement	16 mm/year	$5 \cdot 10^{-10}$
	1	Extremely slow	Imperceptible without monitoring instruments; construction possible with precaution.	<16 mm/year	$<5 \cdot 10^{-10}$

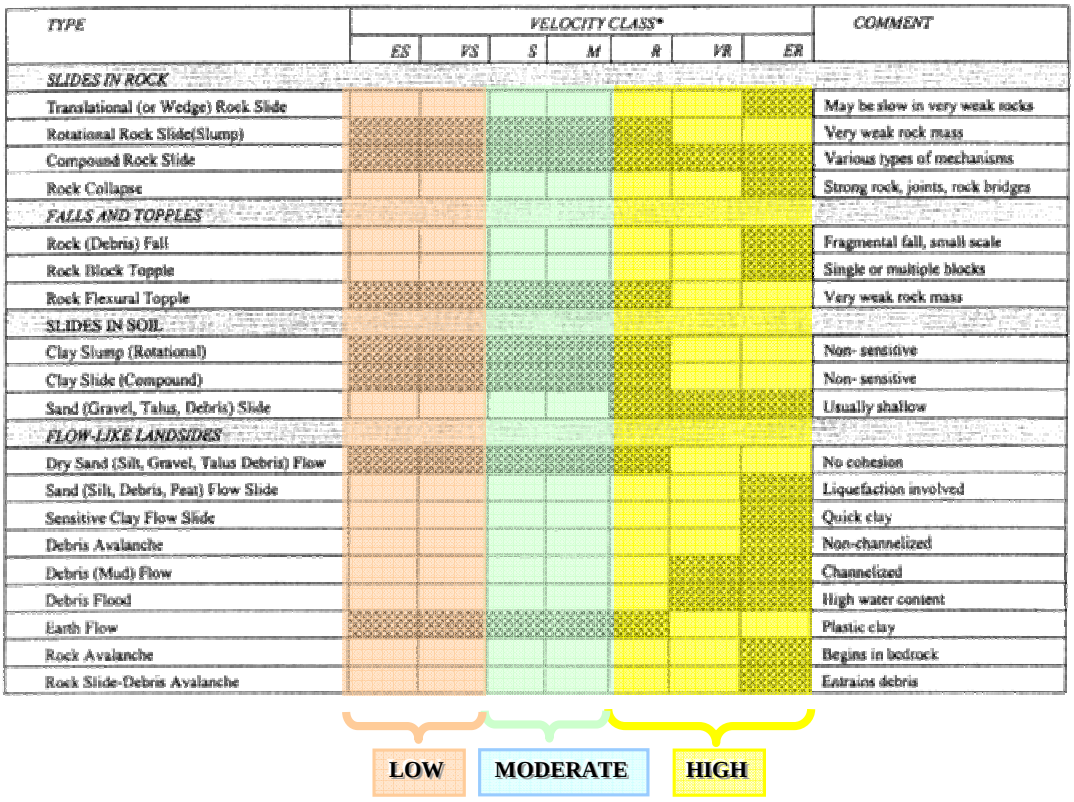
In order to relate the typology of movement to the classification based on landslide intensity, different authors has made various hypothesis during the years.
Referring to the experience of Hungr et al. (2001), the flows can be grouped in different categories, with a specific range of the velocity values.
Considering the velocity value of 5×10^{-5} as discriminating the two classes of high and moderate intensity, the flows can be characterized as in the figure below.

Figure 27. Typical velocity of landslide of the flow type (Hungr et al., 2001)



Referring to the observations made by Hungr, Corominas and Eberhardt (2005), a general assessment of typical failure behavior for various types of landslides is given in the figure below.

Figure 28. A simple classification of landslides, showing typical ranges of velocities (Hungr et al., 2005)



According to these hypotheses, the seven classes given by Cruden and Varnes can be related not only to different intensities, but also to different typologies of movement, as shown in the table X.

Table X. Intensity classes and phenomenologies

Intensity	Class	Description	Phenomenology
HIGH	7	Extremely rapid	Falls Topples Fast moving debris flows Fast moving earth flows
	6	Very rapid	
	5	Rapid	
MODERATE	4	Moderate	Slow moving earth flows Rotational and translational slides
	3	Slow	
LOW	2	Very slow	Lateral spreads Shallow creep Deep creep Deep-seated landslides
	1	Extremely slow	

Complex landslides	Slide-flow Flow-creep Slide-creep

Considering the study area of the Benevento Province, in the Inventory Map a total of nearly 5500 landslides is distinguished. Among these movements, nearly the 62% in number and the 77% in area are classified as belonging to moderate and low intensity classes. Thus, we concentrate our work to these categories of movements.
A table with the main characteristics of the selected typologies of landslides is given below:

Table XI. Moderate and low intensity landslides characteristics

		Number	Area (m^2)	% number	% area
MODERATE INTENSITY	**Earthflows**	2152	88448219	39,16	34,81
	Slides	1779	69094387	32,37	27,19
	Slide-earthflows	919	70047564	16,72	27,57
	Earthflow-creep	131	4966544	2,38	1,95
	Slide-creep	3	131752	0,05	0,05
LOW INTENSITY	**Shallow creep**	371	10624524	6,75	4,18
	Deep creep	83	5024299	1,51	1,98
	Deep seated landslides	57	5755530	1,04	2,27

Regarding the state activity of the recorded movements, the Inventory Map contains further information.
To each movement belonging to the classes of earthflow, slide and slide-earthflow, the state of activity has been assigned.
Three possible states for activity have been considered by the National Basin Authority: inactive, dormant and active, the latter including both active and reactivated phenomena.

As **active** there were indicated the landslides whose last movement has been recorded in the last 5-10 years. To this category belong the landslides defined by Cruden and Varnes (1996) as *active* and *suspended* (cfr. Chapter 2).
As the probability of reactivation of these landslides is nearly 0.1 and referring to the International Guidelines, to this category belong movements with a Very High Hazard Descriptor.

Table XII: Examples of descriptors for hazard zoning (International Guidelines ; JTC-1, 2007)

Hazard Descriptor	**Rock Falls from Natural Cliffs or Rock Cut Slope**	**Slides of Cuts and Fills on Roads or Railways**	**Small Landslides on Natural Slopes**	**Individual Landslides on Natural Slopes**
	Number/annum/km of cliff or rock cut slope	Number/annum/km of cut or fill	Number/square km/annum	Annual probability of active sliding
Very High	>10	>10	>10	10^{-1}
High	1 to 10	1 to 10	1 to 10	10^{-2}
Moderate	0.1 to 1	0.1 to 1	0.1 to 1	10^{-3} to 10^{-4}
Low	0.01 to 0.1	0.01 to 0.1	0.01 to 0.1	10^{-5}
Very Low	< 0.01	<0.01	< 0.01	$< 10^{-6}$

This class includes also the *active* phenomena, as defined by Leroueil. Referring to the slope movement stage as introduced by Leroueil et al. (1996), landslides can be separated into two main categories: first-time failures and reactivated landslides. First-time landslides commonly are characterized by high velocity and can produce fatal consequences. Reactivated landslides commonly cause great economic damage and, sometimes, temporary or permanent evacuation of large zones.

Therefore, while the terminology proposed in the WP/WLI (cfr. Chapter 2) assigned the term *reactivated* to a landslide which is again active after being inactive, the classification given by Leroueil indicates as *reactivated* a landslide after the post-failure, which may be an occasional reactivation or an active landslide.
The landslides defined as *active* by the National Basin Authority involve also the reactivated landslides defined by Leroueil as *active*.

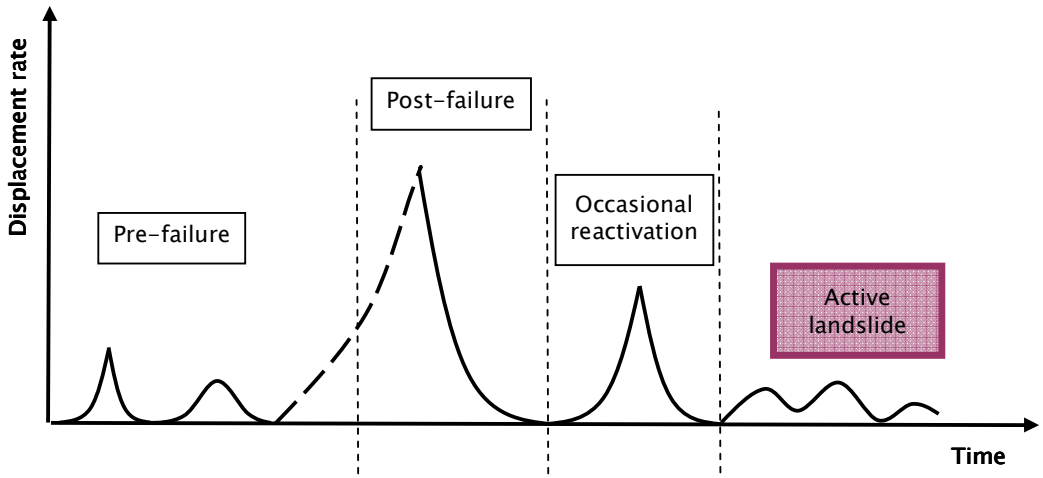


Figure 29. Graph showing the different stages of activity during a landside's life (Leroueil, 1996)

As **dormant** there were indicated the landslides so defined by Cruden and Varnes (1996) (cfr. Chapter 2).
As that the probability of reactivation of these movements can be supposed 0.01 and referring to the International Guidelines, to this category belong movements with a High Hazard Descriptor .

Table XII: Examples of descriptors for hazard zoning (International Guidelines ; JTC-1, 2007)

Hazard Descriptor	**Rock Falls from Natural Cliffs or Rock Cut Slope**	**Slides of Cuts and Fills on Roads or Railways**	**Small Landslides on Natural Slopes**	**Individual Landslides on Natural Slopes**
	Number/annum/km of cliff or rock cut slope	Number/annum/km of cut or fill	Number/square km/annum	Annual probability of active sliding
Very High	>10	>10	>10	10^{-1}
High	1 to 10	1 to 10	1 to 10	10^{-2}
Moderate	0.1 to 1	0.1 to 1	0.1 to 1	10^{-3} to 10^{-4}
Low	0.01 to 0.1	0.01 to 0.1	0.01 to 0.1	10^{-5}
Very Low	< 0.01	<0.01	< 0.01	$< 10^{-6}$

As **inactive** there were indicated the landslides defined by Cruden and Varnes (1996) as *relict* (cfr. Chapter 1). These movements represent an insignificant percentage of the total and thus can be ignored.

Among all the movements presenting information about the state activity, a percentage in number between 73% and 87% (74% - 85% in area) is classified as dormant, while a percentage in number between 11% and 25% (11% - 23% in area) is classified as active.
The inactive landslides constitute the lower part (<2% in number).
A table with the main characteristics of the selected typologies of landslides with the relative state activity is given below:

Table XIIIa and XIIIb. Classification of landslides in number and in area

% in number						
	active	**inactive**	**dormant**	**%active**	**%inactive**	**%dormant**
Earthflows	556	16	1580	25,84	0,74	73,42
Slides	365	26	1388	20,52	1,46	78,02
Slide-earthflows	108	6	805	11,75	0,65	87,60

% in area						
	active	**inactive**	**dormant**	**%active**	**%inactive**	**%dormant**
Earthflows	20704242	1642922	66101056	23,41	1,86	74,73
Slides	9077632	4214342	55802414	13,14	6,10	80,76
Slide-earthflows	8266573	2070980	59710011	11,80	2,96	85,24

All the information obtained from the tables XIIIa and XIIIb are better described in graphs (Figures 30 and 31).
The graphs highlight that the greater part of landslides belongs to the class of moderate intensity. In particular earth flows, slides and complex movements of slide and earth flow are in the majority. Moreover, in each typology of movement the 70-80% of landslides is classified as dormant. This predominance of dormant landslides is probably due to a very cautious method of classification used.

As the inactive category represents an insignificant percentage, the analyses described in the next paragraph have been developed considering only the two categories of **dormant landslides and active landslides** (the latter including both active and reactivated phenomena) and ignoring the inactive landslides.

In various cases, landslides have a great extents and their morphology implies that they are produced by associations of many smaller landslides. Among the total of landslides considered in the analysis, the areal extent of the smallest observable is approximately 370 m^2 and the largest is 2 km^2 approximately (Cfr. Appendix 1).

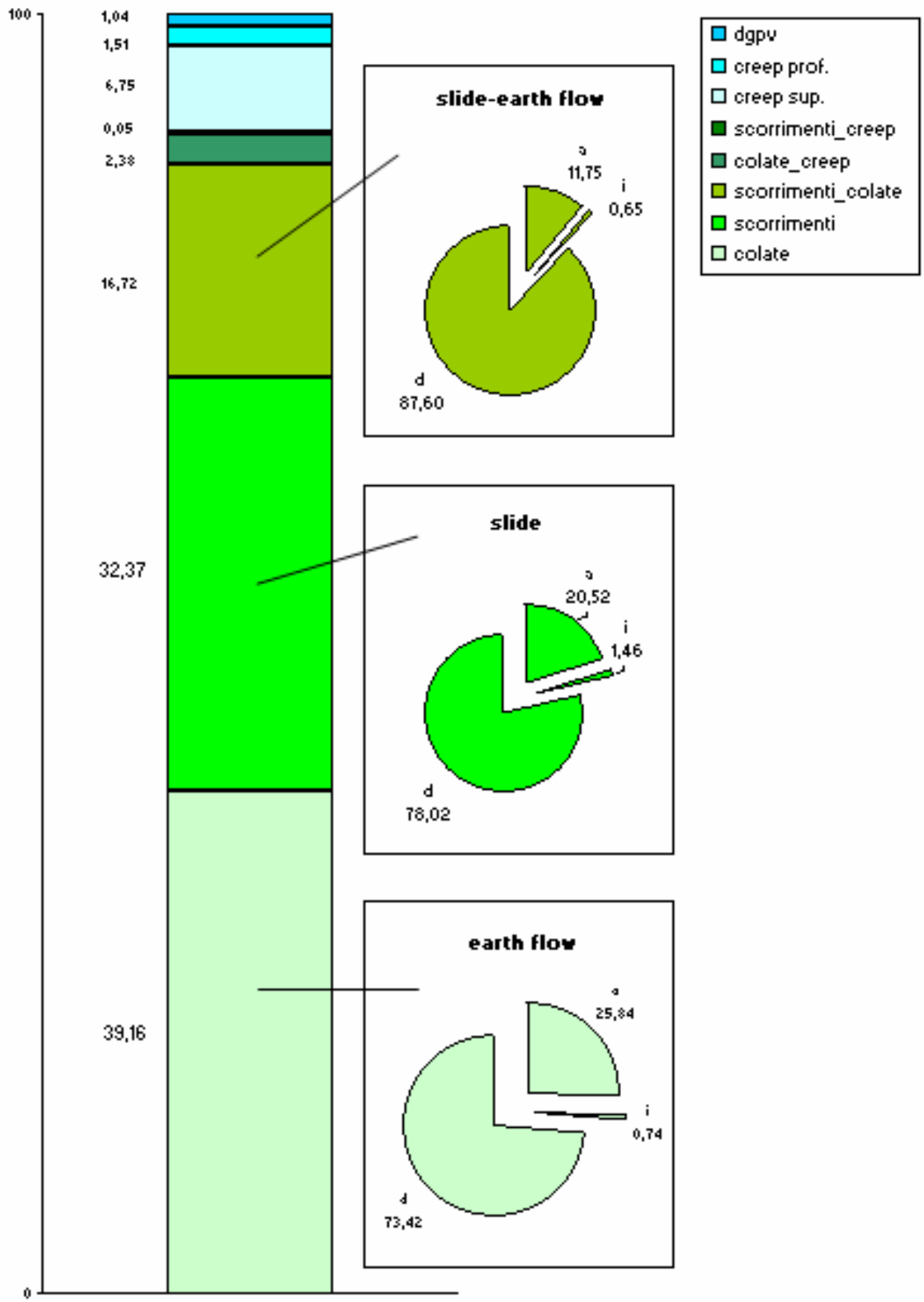


Figure 30. Classification of landslides in number (*a*=active; *d*=dormant; *i*=inactive)

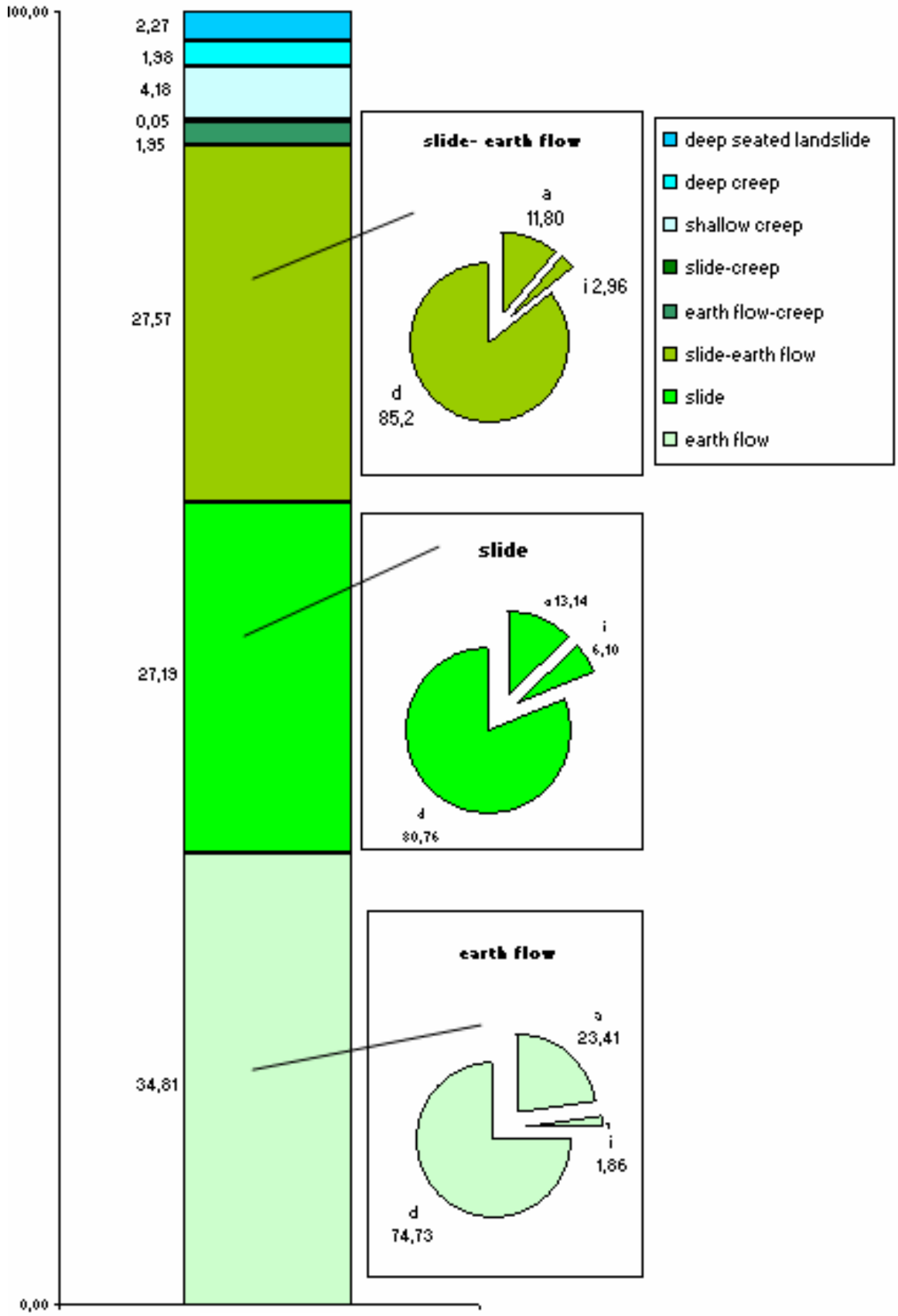


Figure 31. Classification of landslides in area (*a*=active; *d*=dormant; *i*=inactive)

5.2 ANALYSIS OF THE ACTIVITY STATE

As described, the recorded movements have been distinguished for typology of movement and for state activity. Our aim is the analysis of the comparison between the two categories of **active earthflows** and **dormant earthflows**, which constitute the greater percentage of the total of movements, in order to find values of the selected variables that discriminate the belonging to each of the two groups. The comparison will be done considering also the interrelationship between the movements and the hydrographic network.
The inactive movements were ignored, due to their scarce presence.

5.2.1 Data collection

The first step in a GIS-based analysis is the creation of a Digital Elevation Model. The DEM used here is a grid of 10×10m created from the 5m altitude contours map, provided by the National Basin Authority of Rivers Liri, Garigliano and Volturno.
In order to create the grid, a TIN has been derived from the 5m altitude contours. Then the TIN has been converted to a raster of elevation, in the form of a 10m×10m grid.
An effective DEM must be depurated from error. In particular, a digital elevation model free of sinks – a depressionless DEM – is the required input to the flow direction process. In some cases, there may be legitimate sinks in the data. In other cases, the presence of sinks may result in an erroneous calculation of the variables related to the flow process. Hence, it is important to understand the morphology of the area well enough to know what features may truly be sinks on the surface of the earth and which are merely errors in the data.
In this study, after having identified all the sinks on the surface of the earth in the study area, an algorithm has been used to fill some of the sinks. It is important to determine an appropriate depth (z-limit) for fill the sinks, to understand the type of errors present in the data and to determine if the sinks are legitimate morphological features.
The second step in our data collection was to identify the movements.
Each moving mass is simply identified by each polygon with its code and each variable is extracted by the polygons and assigned to them.

5.2.2 Variables used in the analysis

The variables used in the analysis in order to carry on the statistical treatment are representative of the behavior of the moving mass and are not related to the scarps. These variables were all derived from automatic capture processes or from the Landslide Inventory Map provided by the National Basin Authority.
Several parameters have been derived from the DEM or from the Inventory Map, in order to be included into the multivariate statistical analysis. A total of 8 variables have been produced.
These variables have been classified as belonging to different groups, according to the type of information provided:

1. Digital Elevation Model
2. Geometry
3. Geology
4. Activity state
5. Hydrographic network

Among these 9, 6 are quantitative variables, derived directly from the DEM with commands available in the GIS used (Arc/Info version 9.2) or from information in the Inventory Map. The other 3 variables are qualitative and are derived from the Inventory Map.
Table XIV shows the variables used in the analysis and the method of capture of them.

Table XIV. Variables used in the analysis

	Quantitative variables			Qualitative variables derived from Inventory Map
Type of information	Variable	Derived from DEM	Derived from Inventory Map	Variable
DEM	**ΔZ** – Maximum height difference (m)	×		
Geometry	**AREA** – Polygon area(m^2)		×	
	P – Polygon perimeter (m)		×	
	P2A – Shape factor		×	
	DIST – Polygon length(m)	×		
	PEND – Mean slope angle (°)	×		
Geology				**ST** – Soil type
Activity state				**AS** - Grouping variable
Hydrographic net				**CONF –** Typology of confinement

Here we propose to describe each variable and the expected relationship with the occurrence of slope failures.

1. Digital Elevation Model

Variable ΔZ

Definition

This variable is the difference in height between the point situated at the highest elevation and the point situated at the lowest elevation. It is measured in meters above mean sea level and it is calculated for each selected polygon. It is a quantitative variable with values depending on the study area (areas near the sea, mountainous areas, etc.).

Function

It is expected that the two categories of active and dormant landslides have different values of the maximum elevation and thus different values of this variable.

2. Geometry

Variable AREA

Definition

This variable represents the area occupied by the moving masses. It is a quantitative variable, measured in square meters, that can be evaluated directly from the Inventory Map.

Function

As all the other parameters related to the geometry of the movements, the area is important to describe the main characteristics of the moving mass.

Variable P

Definition
This variable represents the perimeter of the moving masses. It is a quantitative variable, measured in meters, that can be evaluated directly from the Inventory Map.

Function
As all the other parameters related to the geometry of the movements, the perimeter is important to describe the main characteristics of the moving mass.

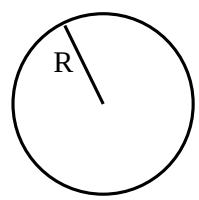
Variable P2A

Definition
It is the ratio between the variables P and AREA. It is a quantitative variable derived from the ones previously described.

Function
This variable is useful to characterize the shape of the landslide. It has been calculated for the most common figures, and the results are shown below:

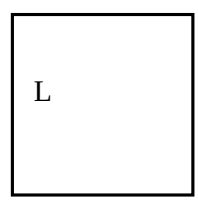
- *CIRCLE*
  The parameter assumes a unique value, independently from the dimension of the figure.

$$\left.\begin{array}{l} P = 2\pi R \\ \\ A = \pi R^2 \end{array}\right\} \Rightarrow P^2 / A = 4\pi = 12.56 \qquad (25)$$



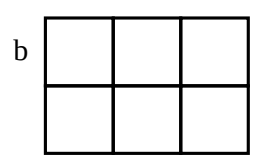
- *SQUARE*
  The parameter assumes a unique value, independently from the dimension of the figure.

$$\left.\begin{array}{l} P = 4L \\ \\ A = L^2 \end{array}\right\} \Rightarrow P^2 / A = 16 \qquad (26)$$



- *RECTANGLE*
  The parameter does not assume a unique value, but depends on the shape of the figure. Thus, considering two rectangles, having different values in the basis and height but having the same area, it is possible to operate a comparison, as shown below.

$$\left.\begin{array}{l} P = 10b \\ \\ A = 6b^2 \end{array}\right\} \Rightarrow P^2 / A \approx 16.67 \qquad (27)$$



$$\left.\begin{array}{l} P = 14b \\ \\ A = 6b^2 \end{array}\right\} \Rightarrow P^2 / A \approx 32.67 \qquad (28)$$

The more the rectangle is slim, the greater is the value of the parameter. Thus, the variable may be thought as a shape factor, useful to characterize the geometry of the moving mass.

Moreover, it is possible to define some ranges of the values assumed by the parameter. My hypothesis is described below.

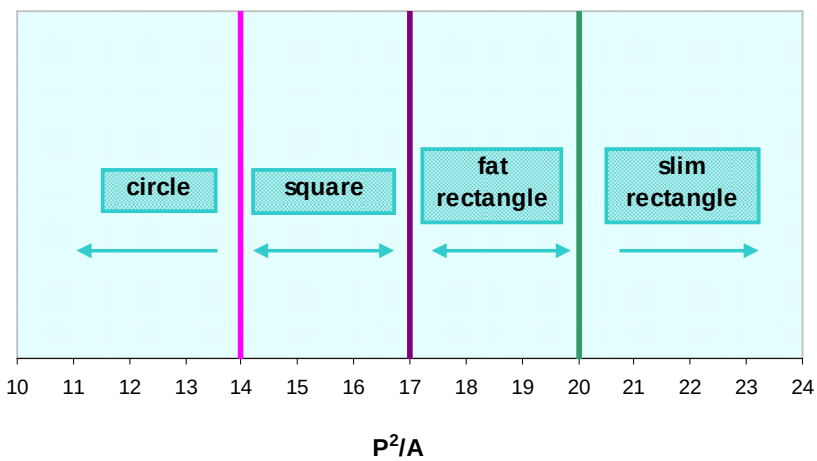


Figure 32. Shape factor ranges

Variable DIST

Definition

This variable represents the length of the moving mass and is calculated as the distance between the points situated at the highest and lowest elevation. The research of these points for each polygon, necessary to define the variable ΔZ, brings to the evaluation of a couple of coordinates. From these coordinates, it is possible to evaluate the distance, using the expression reported below:

$$DIST = \sqrt{(X_1 - X_2)^2 + (Y_1 - Y_2)^2} \qquad (29)$$

It is a quantitative variable, measured in meters.

Function

The variable DIST is representative of the travel distance of a movement. It is expected that active and dormant landslides, having different evolution, have different values of this parameter.

Variable PEND

Definition
The variable represents the mean slope angle of the moving mass and is calculated from the variables ΔZ and DIST, through the expression below:

$$PEND = arcsen\left(\frac{\Delta Z}{DIST}\right) \tag{30}$$

Function
The slope angle is probably the main factor of stability as it affects the magnitude of both normal and shear stresses on the potential surface of failure.
The higher the angle the greater is the shearing component of the forces acting at the potential surface of failure (Jones et al., 1961).

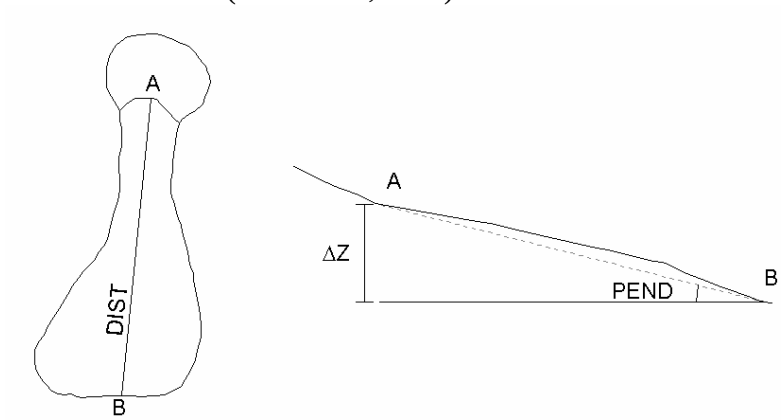


Figure 33. Adopted variables' scheme

3. Geology

Variable ST

Definition
The variable represents the typology of soil.
The geo-lithological map provided by the National Basin Authority distinguishes five homogeneous geolithological complexes in the study area:

- M - intercalate basin limestone and clays;
- Q - quaternary clastic deposits;
- C - shelf limestone;
- P - sequence of sands, clays and conglomerates;
- Mar - flysh;

However the majority of the selected movements (active and dormant earthflows) occurs only in three types of lithologies:

- M - intercalate basin limestone and clays;
- P - sequence of sands, clays and conglomerates;
- Mar - flysh;

Thus, these three typologies have been selected as variables in the statistical analysis.
It is a qualitative variable, which has to be transformed in quantitative in order to be used in the statistical analysis.

Function
Lithology is one of the most influential parameters on slope instability, because each material has different shear strength and hydraulic conductivity.

4. Activity state

Variable AS

Definition
This variable is the state of activity of the landslide, as identified and inventoried in the study area. It will be used as the grouping variable in the statistical analysis. It is a qualitative variable.

Function
The variable AS shows the presence of active or dormant landslides and it is used as grouping variable in the statistical analysis, in order to establish the relationship with the conditioning factors.

5. Hydrographic net

Variable CONF

Definition
This variable represents the type of confinement of the landslide and its relationship with the hydrographic network.
Observing the conformations of the moving masses, three possible types of confinement have been distinguished:

- A: landslides running almost perpendicular towards the river;
- B: landslides without any influence from the river;
- C: landslides following the river course.

Examples of the three typologies are described below.

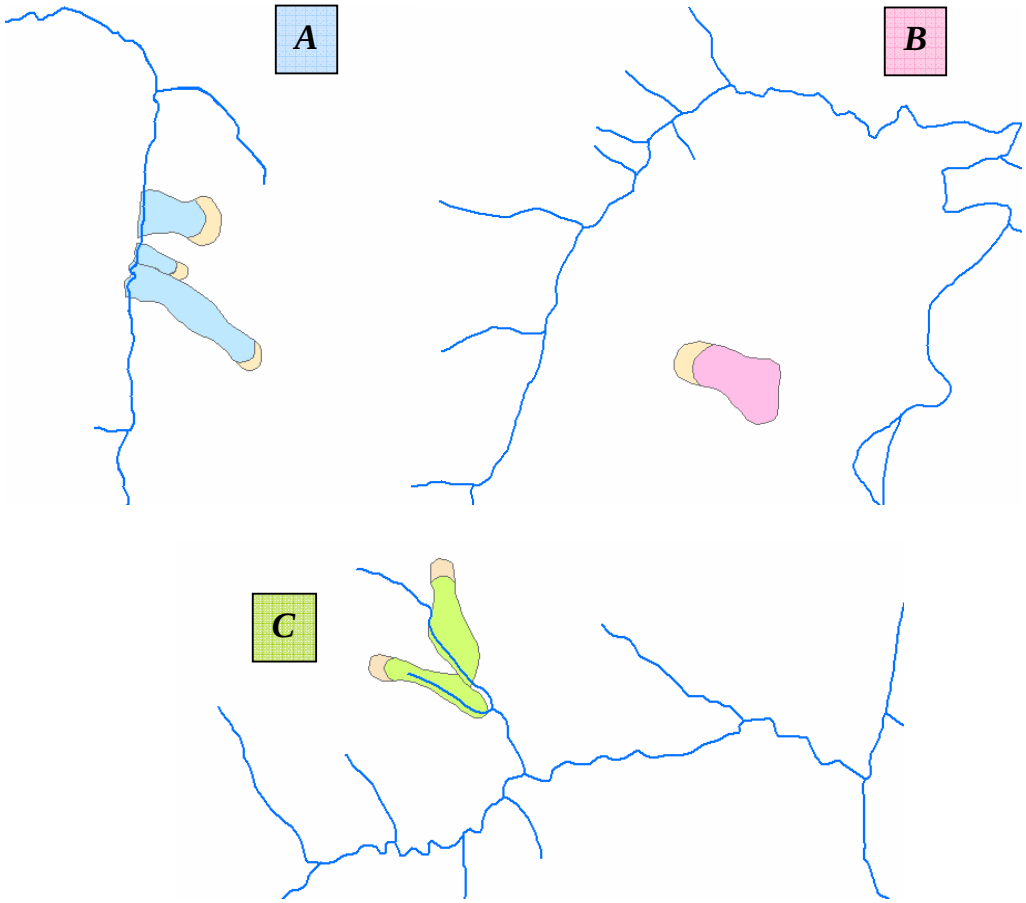


Figure 34. Typologies of confinement

This variable is qualitative and it must be transforming in quantitative to be used in the statistical analysis. However, before using this parameter, it is necessary to verify that it is an adequate variable for the analysis. It means that only if it is proved that differences in the variables values exist among the three classes, CONF can be used in the statistical analysis.

Function

The variable is representative of the interrelationship between existing landslides and the drainage network. As described in the point 2.1.5, toe erosion is considered one of the factors that contribute to increase shear stress, thus reducing the safety factor of a slope. Hence, it is important to identify the phenomena that interact at the toe with streams.

According to this observation, the three selected category of *CONF* identify three different relation between the recorded movement and the drainage network. In particular:

- **A: Strong erosion**
  Landslides that run almost perpendicular towards the river may cross the stream line and the deposit material at the toe suffer erosion by the river.
- **B: Without obstacles**
  Landslides without any influence from the river may expand and propagate without obstacle.
- **C: Obstacle to propagation**
  Landslides following the river course find in the river a great obstacle.

5.2.3 Statistical treatment

The statistical treatment was carried out using the SPSS Inc.(1988) statistical package, following the steps described in the chapter 4.3.

Creation of the sample

The two samples of active and dormant earthflows have been obtained simply selecting by hand the movements in the Inventory Map showing the evidence of membership to each typology of confinement.

Due to the prevalence of dormant earthflows on active eartflows and to the difficulties in recognizing the typology of confinement for many movements, the two samples do not have similar number of individuals, as it is recommended (Dillon and Goldstein, 1986).

General information are reported below:

Table XV. Statistics

Type of confinement	Typology of movement	Number
CONF=A	Active earthflows	44
	Dormant earthflows	141
CONF=B	Active earthflows	40
	Dormant earthflows	121
CONF=C	Active earthflows	42
	Dormant earthflows	130

Regarding the variable *ST*, information about the membership of the different lithologies are reported below.

Table XVI. Statistics

Type of confinement	Typology of soil	Typology of movement	Number
CONF=A	**Mar**	Active earthflows	19
		Dormant earthflows	82
	P	Active earthflows	12
		Dormant earthflows	48
	M	Active earthflows	13
		Dormant earthflows	11
CONF=B	**Mar**	Active earthflows	22
		Dormant earthflows	98
	P	Active earthflows	9
		Dormant earthflows	18
	M	Active earthflows	9
		Dormant earthflows	5
CONF=C	**Mar**	Active earthflows	26
		Dormant earthflows	92
	P	Active earthflows	7
		Dormant earthflows	19
	M	Active earthflows	9
		Dormant earthflows	19

Transformation of variables

As already said, the discriminant analysis has difficulties in dealing with qualitative data. Therefore, it is necessary to transform qualitative variables in quantitative variables, as shown in the following table XVII.

Table XVII. Transformation of the variables

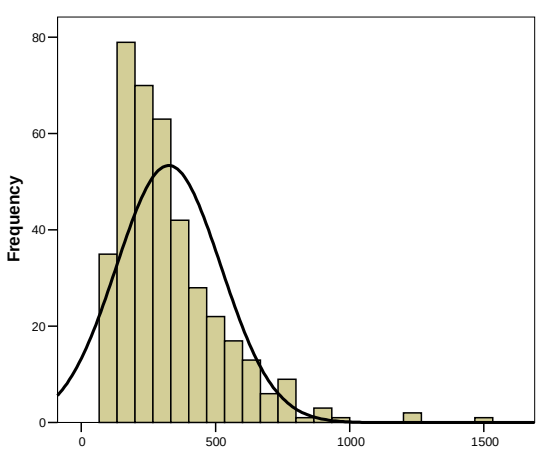
Variable	Value	Observations
ST – Soil type		
P - Sequence of sands, clay and conglomerate	1	Higher values express higher proneness to fail.
M - Intercalate basin limestone and clays	2	
Mar - Flysch	3	
AS – Grouping variable		
Active landslides	1	Variable used only in discriminant analysis to classify the sample.
Dormant landslides	0	
CONF– Typology of confinement		
C - Obstacle to propagation	1	Higher values express lower influences suffered by the river.
A - Strong erosion	2	
B - Without obstacles	3	

Regarding the variable CONF, a hypothesis on the greater and lower influence exercised by the river has been done, based on general considerations.

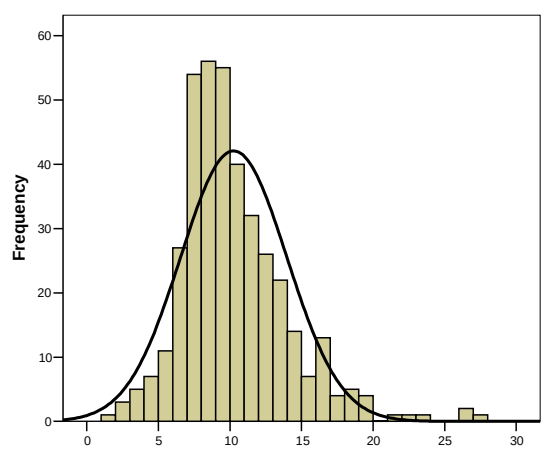
Testing for normal distribution

In order to work with normal distributions of the variables, the normalization test has been developed. To observe if each variable has a normal distribution, the histogram of each one has been created. In particular, two graphs have been developed for each variable: the first refers to the population of dormant landslides, the second to the group of active landslides. From them, it is possible to observe that some variables have a normal distribution and others no. As examples, two graphs have been reported below (for the others cfr. Appendix 2).

DIST – POLYGON LENGTH



PEND – MEAN SLOPE ANGLE



The log-transformation of the variable is necessary for the variable length (*DIST*), area (*A*), perimeter (*P*), difference in height (ΔZ) and shape factor (*P2A*). The new variables will be renamed, using the term *log* before the name of the variable.

Regarding the parameter *ST*, the transformation is not necessary because it is a qualitative parameter, previously transformed in quantitative, that assumes only three values.

Selection of independent variables

Before defining the independent variables, it is necessary to verify that the parameter *CONF* is an adequate variable for the analysis. It means that only if it is proved that differences in the variables values exist among the three classes, CONF can be used in the statistical analysis.
In order to verify that, three different T-test has been done, each considering two typologies of confinement at time.

CONF 1 and 3

The analysis has demonstrated that substantial differences among the two groups exist. All the variables chosen give significant result, except for the parameter *ST*.

Table XVIII. T-test and One-way test results – CONF 1 and CONF 3

				One-way test		T-test	
Variable	**CONF**	**Mean**	**Standard Deviation**	**F**	**Sig.**	**t**	**Sig.**
PEND	1	9.7394	3.22843	8.439	0.004	-3.968	0.000
	3	11.3924	4.32540				
logDIST	1	2.6330	0.19439	1.950	0.164	13.053	0.000
	3	2.3403	0.21477				
logAREA	1	4.5488	0.30731	0.121	0.728	10.975	0.000
	3	4.1831	0.30020				
logP	1	3.0062	0.18217	0.005	0.945	12.437	0.000
	3	2.7604	0.17828				
log∆Z	1	1.8386	0.20240	5.103	0.025	9.698	0.000
	3	1.6087	0.22828				
ST	1	2.53	0.745	0.037	0.847	-0.517	0.605
	3	2.58	0.764				
logP2A	1	1.4636	0.11311	0.186	0.667	10.276	0.000
	3	1.3378	0.10994				

CONF 1 and 2

The analysis has demonstrated that substantial differences among the two groups exist. All the variables chosen give significant result, except for the parameter *PEND*.

Table XIX. T-test and One-way test results – CONF 1 and CONF 2

				One-way test		T-test	
Variable	**CONF**	**Mean**	**Standard Deviation**	**F**	**Sig.**	**t**	**Sig.**
PEND	1	9.7394	3.22843	10.903	0.001	-1.490	0.137
	2	10.3282	4.20631				
logDIST	1	2.6330	0.19439	0.065	0.798	14.206	0.000
	2	2.3342	0.20245				
logAREA	1	4.5488	0.30731	1.764	0.185	8.971	0.000
	2	4.2671	0.28596				
logP	1	3.0062	0.18217	2.208	0.138	12.248	0.000
	2	2.7815	0.16443				
log∆Z	1	1.8386	0.20240	17.677	0.000	11.406	0.000
	2	1.5504	0.27212				
ST	1	2.53	0.745	32.970	0.000	3.573	0.000
	2	2.22	0.909				
logP2A	1	1.4636	0.11311	4.501	0.035	15.332	0.000
	2	1.2958	0.09154				

CONF 2 and 3

The analysis has demonstrated that substantial differences among the two groups exist. All the variables chosen give significant result, except for the parameter *logDIST* and *logP*.

Table XX. T-test and One-way test results– CONF 2 and CONF 3

				One-way test		T-test	
Variable	**CONF**	**Mean**	**Standard Deviation**	**F**	**Sig.**	**t**	**Sig.**
PEND	2	10.3282	4.20631	0.021	0.886	-2.317	0.021
	3	11.3924	4.32540				
logDIST	2	2.3342	0.20245	1.265	0.262	-0.274	0.784
	3	2.3403	0.21477				
logAREA	2	4.2671	0.28596	0.921	0.338	2.665	0.008
	3	4.1831	0.30020				
logP	2	2.7815	0.16443	2.039	0.154	1.141	0.255
	3	2.7604	0.17828				
log∆Z	2	1.5504	0.27212	4.237	0.040	-2.167	0.031
	3	1.6087	0.22828				
ST	2	2.22	0.909	30.104	0.000	-3.960	0.000
	3	2.58	0.764				
logP2A	2	1.2958	0.09154	2.605	0.107	-3.875	0.000
	3	1.3378	0.10994				

The results demonstrate that substantial differences among the three groups caught two each time exist. The next step consists in verifying the independence of the variables. As already said, dependent variables must be removed, searching for possible correlation.

A bivariate analysis of correlation will be first done, in order to search for possible relationship between the variables. Then the principal component analysis (PCA) will be applied, in order to view the membership of each variable to the different factors.

- **Bivariate correlation**

This procedure produces Pearson product-moment correlations with significance level. One or more matrices of correlation coefficients are produced. By default a simple square matrix of correlation coefficients is printed. If a simple list of variable is provided, the procedure prints the correlations of each variable with every other variable in the list in a square or lower-triangular matrix.

The table with the coefficient of Pearson is reported below. The variable showing high correlation have the coefficient colored in yellow.

- *AREA, P* and *P2A*

It can be observed that the area, the perimeter and their ratio are strongly correlated. It means that moving masses with an homogenous shape prevail. Long and slim movements, with a great value of perimeter and little area, seem to be rare.

Moreover, these variables are correlated also to the roughness of the terrain. It can be explained considering the way this parameter is obtained, that is the ratio between the surface area (depending on the slope angle) and the planimetric area.

- *AREA, P* and *DIST*

As expected the variable area and perimeter of the moving mass are related to its maximum distance. As already said, we expect not to find very slim moving mass, with high values of distance and little values of area, nor very squat watersheds, with little values of distance and great values of area.

- *DIST* and *ΔZ*

The elevation difference between the two points at highest and lowest altitude is strongly correlated to the length of the moving mass. It means that the greater is the difference in height, the greater is the distance covered by the moving mass.

Table XXI. Pearson coefficients of correlation matrix

	logDIST	logAREA	logP	logP2A	logΔZ	ST	PEND	CONF
logDIST	1							
logAREA	0.906	1						
logP	0.964	0.964	1					
logP2A	0.748	0.500	0.712	1				
logΔZ	0.806	0.712	0.771	0.629	1			
ST	0.116	0.121	0.110	0.039	-0.023	1		
PEND	-0.259	-0.252	-0.251	-0.152	0.333	-0.239	1	
CONF	-0.482	-0.441	-0.482	-0.405	-0.350	0.017	0.165	1

- **Factorial analysis**

The principal component analysis has been developed with the 8 variables considered. The goodness of the results depends on total variance explained. In this case the first two factors explain the 72.8 per cent of the variance (cfr. Table XXII).

Table XXII. Total variance explained in the Principal Component Analysis
(Results of the sum of the square saturation of the rotation)

Component	Total	% of the Variance	% Cumulative
1	4.395	54.934	54.934
2	1.429	17.865	72.799

From the results obtained in the table XXIII it is clear that the total variance of each variable is greater then the 30%, being the variables *logDIST*, *logP* and *logΔZ* the best representative, with more than 90% of the variance, followed by *logAREA* and *PEND*, with more than 82%. The results are interpreted in the way that the greater is the extraction, the more independent is the variable and the lower is the correlation with the others.

Table XXIII. Communalities or total variance of each variable
(Extraction method: Principal component analysis)

Variable	Extraction
logDIST	0.964
logAREA	0.839
logP	0.962
logP2A	0.611
logΔZ	0.913
ST	0.376
PEND	0.821
CONF	0.338

The rotated matrix with two factors is presented in the table XXIV, where the greater loads are in yellow.

Table XXIV. Varimax rotated component matrix
(Extraction method: Principal component analysis)

	Factor	
	1	2
logDIST	**0.971**	-0.141
logAREA	**0.899**	-0.174
logP	**0.969**	-0.150
logP2A	**0.781**	-0.030
log∆Z	**0.868**	0.399
ST	0.048	**-0.611**
PEND	-0.111	**0.900**
CONF	**-0.569**	0.117

FACTOR 1: This factor represent nearly the 55% of the total variance and is defined by the variables *logDIST, logAREA, logP, logP2A, log∆Z* and *CONF*.

FACTOR 2: This factor represent nearly the 18% of the total variance and is defined by the variables *ST* and *PEND*.

After having identified the dependent variables, before rejecting one of these, **One-way test** and **T-test** were developed in order to understand the influence of each variable on stability.
In the SPSS package, the ONEWAY procedure calculates a one-way analysis of variance, as well as a variety of multiple comparison procedures, and the T-TEST procedure calculates a test for the equality of two means for independent or paired samples.
ONEWAY can analyze several dependent variables by one independent variable with one specification of the procedure. By default, the procedure produces a standard analysis of variance table for each dependent variable.

T-TEST produces Student's t, degrees of freedom, and two-tailed probability for a comparison of two means. In addition, the mean, standard deviation, and standard error are displayed for each variable. An independent-samples test divides the case into two groups and compares the group means on a single variable.

The One-way test analyze the variance through the parameter F. Based on the value assumed by the significance level of F (Sig.), different hypothesis about the variance are done. In particular:

- if Sig > 0.05, a hypothesis of equal variances is done;
- if Sig < 0.05, a hypothesis of different variances is done.

According to this, different function for the calculation of the mean are considered, so that the value t and its significance level (Sig.) are evaluated. The results are interpreted in the way that the lower is the significance level the greater is the difference in mean between the two populations. In particular,

- if Sig > 0.05, significant differences in mean between the two groups are not present;
- if Sig < 0.05, significant differences in mean between the two groups are present.

The Tables of the results of One-way test and T-test are reported below.
It can be observed that all the variables have mean and standard deviation values very near among the two populations. Thus, no one of the selected variables has high discriminant power. The calculation has been developed first considering all the active earthflows and all the dormant earthflows; then distinguishing the three typologies of confinement; finally distinguishing also the type of soil. No one of the considered case has shown consistent results.

Table XXV. T-test and One-way test results

ALL				One-way test		T-test	
Variable	**Activity state**	**Mean**	**Standard Deviation**	**F**	**Sig.**	**t**	**Sig.**
PEND	active	11.0982	4.72869	8.005	0.005	1.819	0.071
	dormant	10.2594	3.71684				
logDIST	active	2.4048	0.26086	1.105	0.294	-1.598	0.111
	dormant	2.4451	0.24143				
logAREA	active	4.2819	0.34741	0.245	0.621	-2.030	0.043
	dormant	4.3514	0.32968				
logP	active	2.8275	0.21802	0.765	0.382	-1.379	0.168
	dormant	2.8566	0.20272				
log∆Z	active	1.6541	0.25728	1.201	0.274	-0.488	0.626
	dormant	1.6675	0.27140				
ST	active	2.31	0.815	0.047	0.828	-1.984	0.048
	dormant	2.48	0.828				
logP2A	active	1.3730	0.13574	0.476	0.491	0.855	0.393
	dormant	1.3619	0.12412				
CONF	active	1.98	0.810	0.028	0.867	0.086	0.931
	dormant	1.98	0.801				

The tables demonstrate that the variables selected for the analysis have a very low disciminating power. Analogous calculations have been developed distinguishing the typology of soil (cfr. Appendix 2).

5.3 PERSPECTIVES FOR THE APPLICABILITY OF THE USED APPROACH

5.3.1 Data collection

An attempt in applying the statistical procedure described to study large landslide susceptibility has been developed. In this case, we aim at comparing failed and unfailed slopes. As *failed* we indicate the movement mapped in the Inventory Map, while as *unfailed* we indicate the areas without any recorded movement. To create the category of failed slopes, only the polygons corresponding to those movements that show the evidence of a scarp have been selected. In particular, three typologies have been considered:

- Earthflows
- Slides
- Complex phenomena of slide – earthflows

All the other typologies of landslide mapped in the Inventory Map have been ignored, as they do not have a scarp surface.
To these typologies of movements, information about the state activity are also associated.
However, in this first analysis the state of activity may be ignored, considering as one the two

categories of active and dormant landslides (the inactive ones have been ignored).
The aim of this analysis is the assessment of landslide susceptibility, by comparing cells with the presence of movements (failed cells) and cells without any recorded movement (unfailed cells). Thus, the activity state of the movement can be ignored.
The first step in a GIS-based analysis is the creation of a Digital Elevation Model. The DEM used here is the grid of 10×10m already created for the previous work.

The second step in our data collection was to identify all the scarps. In the previous application in the Spanish Eastern Pyrenees, each landslide had a dimension so that each movement falls in a unique cell. In this case, instead, dealing with large landslides, the dimension is bigger than one cell area. As a consequence, each movement is represented by a polygon occupying a certain area. Having selected the polygons corresponding to the scarps, it is necessary to find the best way to deal with these polygons.
In order to do identify the scarp, the cell corresponding at the highest altitude has been selected. Among the grid cells having their centre point in a scarp, the point corresponding to the biggest value of the elevation has been identified. In this way, each scarp is localized by its point.
In few cases, different points having the same maximum elevation have been identified within the same scarp. In such cases, due to the impossibility to work with more than one point for each scarp, the adopted procedure has been improved: among the points having the same maximum elevation, the one having the greatest slope angle has been selected.

Having identified each polygon with its point, the third step is to individuate a methodology to capture the values of the variables that are expected to influence the slope instability.
Below two different methodologies are illustrated, while the variables used in the analysis and the comparison between the two procedures are described in detail in the following paragraphs.
The first methodology uses the variables associated to the selected maximum elevation points.
The second methodology is developed considering a 20m buffer for each selected scarp.
In this way we have two different ways to work, so to compare the values of the variables in the two procedures.

Points. From each point corresponding to the highest elevation in each scarp, the value of the adopted variables has been extracted.
In this case, as we work with points, a unique value of each parameter is associated to each movement. The disadvantage of such procedure is the presence of zero-value of few variables in different elements. Most of those unacceptable values regard parameters related to the water flowing toward the point.

Buffer polygons. For each scarp polygon, a 20 m buffer has been created. The aim of using such procedure is to extract not a unique value of the selected variables for each movement (as done in the previous points procedure), but a series of data. As it is considered that the best morphological conditions contributing in the first-time landslide susceptibility assessment would be extracted from the vicinity of the scarp polygon, a buffer zone to it has been created.

The tool BUFFER creates a new feature class of buffer polygons around the specified input features, as illustrated in the figure 35.

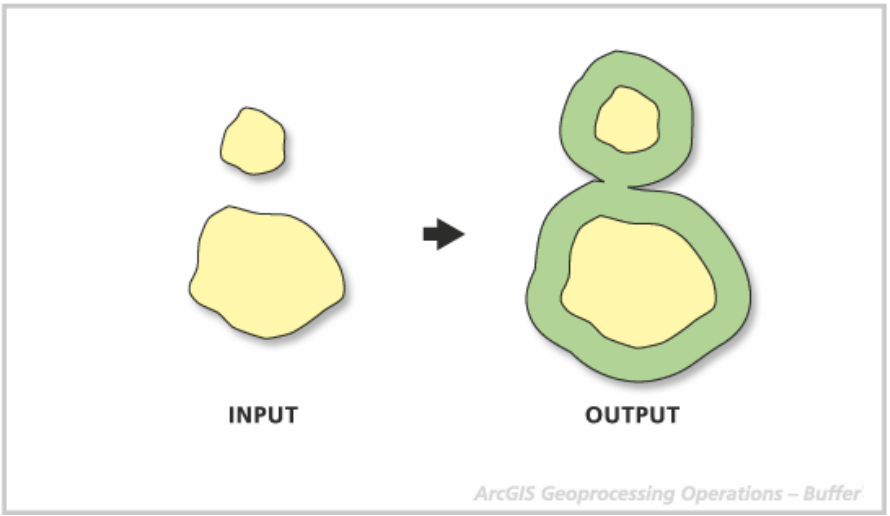


Figure 35. The *buffer* tool

The tool works in Euclidean space and uses a two dimensional algorithm. A buffer will be the same width no matter what the coordinate system is. The width of the buffer has been specified as a fixed distance. In particular, a distance of 20 meters has been selected. This value has been chosen to have at least two cells of width for each buffer.

Observation

When a map is realized, and also while using it, a metric error is done, because the objects are represented with a graphical feature that creates a degree of indetermination.

When the topographer puts the points on a map and joins them with a line, it uses a pencil; the drawn points and the graphical line which joins them has a certain dimension, that is minimum of 0,2 millimeters. This value is called **graphics error**. A point that has a diameter of 0.2 millimeters on a 1:2.000 scale map represents a circle of 40 cm of diameter. This value is obtained multiplying the graphics error for the denominator of the scale (0.2 millimeters × 2.000). Equally, the cross-sectional dimension of the segment that joins the two points on the map represents on the land a feature of 40 cm. This means that, dealing with a 1:2.000 scale, the points on the map have an indetermination degree of approximately 40 cm.

It is important to notice that all the types of maps demand that the deformations are inferior to the graphics error (correspondent to 0.2 millimeters on the sheet of the design). The table XXVI shows the maximum resolution correspondent to the scale of the map.

Table XXVI. Maximum resolution at different scales

Map Scale	*Max Resolution*
1: 500	10 cm
1: 1000	20 cm
1: 2000	40 cm
1: 5000	1 m
1: 10000	2 m
1: 25000	5 m
1: 50000	10 m
1: 100000	20 m
1: 1000000	200 m

However, in order to estimate how much the effect of the graphics error influences the qualitative and metric content of a map, we do not have to consider the value of 0.2 millimeter, which is the minimum thickness of a line used while drawing a map. Actually it is assumed that the graphics error can carry to a indetermination degree that can reaches the value of 0.5 millimeters. This value is called **graphical tolerance**, and practically it determines the maximum value of error that the cartographer can commit in the construction of a map and also the degree of metric reliability that the customer can expect when carrying out measures on the paper.
If we consider a map in 1:2.000 scale, the graphical tolerance is 1 meter. It means that, in carrying out measures on the map, we can commit error of wrong distances of 1 m respect to their real value. This value corresponds to the mean value of the resolution degree.
If the graphics error is defined as the minimum thickness of a line used while drawing a map, generally a line cannot be drawn much narrower than about 0.5 millimeters. Therefore, on a 1:20,000 scale paper map, the minimum distance which can be represented (mean resolution) is about 10 m. The table XXVII shows the mean resolution correspondent to different scales.

Table XXVII. Mean resolution at different scales

Map Scale	***Mean Resolution***
1: 500	25 cm
1: 1000	50 cm
1: 2000	100 cm
1: 5000	2.5 m
1: 10000	5 m
1: 25000	12.5 m
1: 50000	25 m
1: 100000	50 m
1: 1000000	500 m

Finally, the minimum resolution is evaluated considering that a line cannot be drawn much larger than 1 millimeter. In this case, we talk about the minimum degree of precision. As a consequence, called D the denominator of scale of the map, the minimum resolution in the acquaintance of coordinated of a point is given by: $e = 0{,}001 \times D$ [m].
For example, if the topographical map has a 1:2000 scale, e is 2 m, corresponding to 200 cm.

The table XXVIII shows the minimum resolution corresponding to the scale of the map.

Table XXVIII. Minimum resolution at different scales

Map Scale	***Min Resolution***
1: 500	500 cm
1: 1000	1 m
1: 2000	2 m
1: 5000	5 m
1: 10000	10 m
1: 25000	25 m
1: 50000	50 m
1: 100000	100 m
1: 1000000	1000 m

In our case, we work with a map developed in 1:25000 scale. Thus, the resolution values are:

Table XXIX. Maximum, mean and minimum resolution at different scales

Map Scale	*Max Resolution*	*Mean Resolution*	*Min Resolution*
1: 500	10 cm	25 cm	500 cm
1: 1000	20 cm	50 cm	1 m
1: 2000	40 cm	100 cm	2 m
1: 5000	1 m	2.5 m	5 m
1: 10000	2 m	5 m	10 m
1: 25000	5 m	12.5 m	25 m
1: 50000	10 m	25 m	50 m
1: 100000	20 m	50 m	100 m
1: 1000000	200 m	500 m	1000 m

The choice of a buffer zone of 20 meters can be considered correct, because the value brings to a resolution degree between the mean and the minimum values.

After creating the buffers, these polygons must be cut with respect to the boundary of the corresponding moving mass and scarp, through the use of the tool ERASE, that works as illustrated in the figure 36.

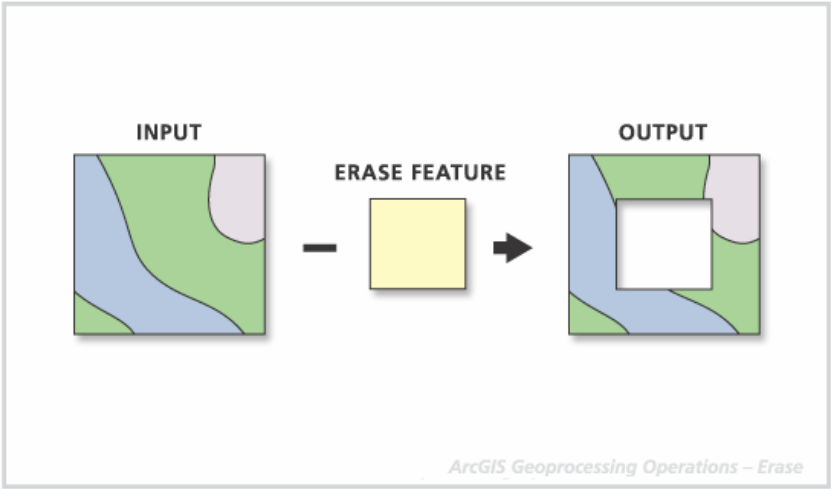


Figure 36. The *erase* tool

The tool ERASE creates a new feature classes by overlaying two sets of features. The Erase Features polygons define the erasing region. Input Features that are within the erasing region are removed. The output feature classes contain only those Input Features that are outside the erasing region.

An example showing the described methodology is illustrated in the figure 37.

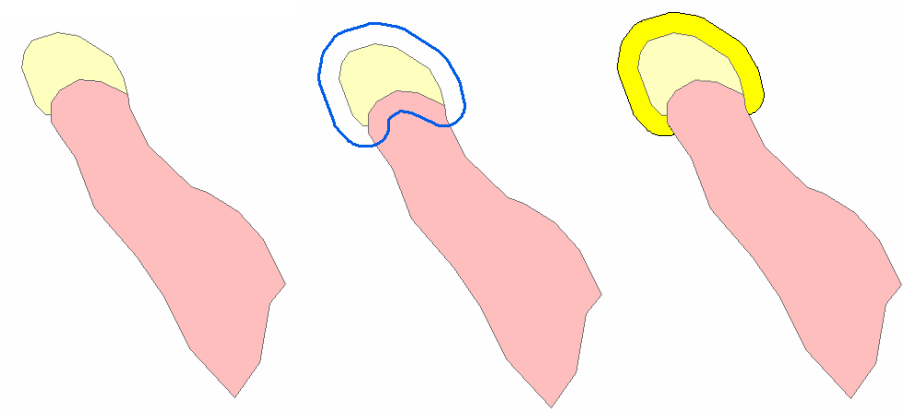


Figure 37. Creation and cut of the buffer zone

After cutting the buffer polygons, the value of the adopted variables has been extracted.
In this case, as we work with polygons, the value associated to each point within a buffer polygon is extracted. In order to catch the parameters, a specified tool extracts the cell values of a grid in the given points (in this case the points within the buffer zones).
As a consequence, in this case a table of values of each parameter is associated to each buffer.
In order to understand how the tool that extracts values from a grid to given points works, an example of the output INFO table is reported below:

Table XXX. Info-table extracted from the *extract values to points* tool

VALUE	COUNT	AREA	MEAN	MIN	MAX	...
0	5	125.0000	0.6	0.0	.0	...
1	5	125.0000	1.0	0.0	3.0	...
2	3	75.0000	1.667	1.0	2.0	...
4	2	50.0000	3.0	3.0	3.0	...

The tool summarizes values of a raster within the zones of another dataset (the buffer zones dataset) and reports the results to a table, recording in the output INFO table the mean, minimum, maximum, range, sum, standard deviation, variety, majority, and median of the values of all cells in the value raster that belong to the same zone.

Having described two different approaches, it is necessary to identify which is the more correct and significant. In order to do that, the values assumed by the variables chosen for the analysis have to be compared. Thus, before comparing the two procedures, the variables used have to be described.

5.3.2 Variables used in the analysis

In this case the variables used in the analysis in order to carry on the statistical treatment are all derived from automatic capture processes or from the Landslide Inventory Map provided by the National Basin Authority.
Several terrain parameters related to the occurrence of the slope failures have been derived from the DEM or from the Inventory Map, in order to be included into the multivariate statistical analysis. A total of 13 variables have been produced.

These variables have been classified as belonging to different groups, according to the type of information provided:

1. Digital Elevation Model
2. Geometry
3. Watershed dimension
4. Geology
5. Landslide

Among these 13, 11 are derived from the DEM, while 2 from the Inventory Map.
Among the 11 derived from the DEM, 6 DTMs (cell height, slope angle, slope aspect, curvature, transverse curvature and longitudinal curvature) were obtained directly with commands available in the GIS used (Arc/Info version 9.2), 4 parameters (watershed area, watershed length, mean watershed angle and stream power index) were created from algorithms especially defined, and the last one (slope roughness) was created through a GIS extension working with the Arc/Info.
Table XXXI shows variables used in the analysis and the method of capture of them.

Table XXXI. Variables selected for the statistical analysis

Quantitative variables derived from DEM					**Qualitative variables derived from the Inventory Map**
	Variable	GIS function	Algorithm	GIS extension	
DEM	**H**-Height above mean sea level (m)	×			
Geometry	**β** - Slope angle (°)	×			
	α - Slope aspect (°)	×			
	RUGOS - Roughness			×	
	CURVAR - Curvature	×			
	PF - Longitudinal curvature	×			
	PL - Transverse curvature	×			
Watershed	**A** - Watershed area		×		
dimension	**LONG**- Watershed length		×		
	PENDM - Mean watershed angle		×		
	SPI – Stream power index		×		
Geology					**ST –** Soil type
Landslide					**SL**- Grouping variable

Here we propose to describe the selected variables and their expected relationship with the occurrence of slope failures. All the maps are presented in the Appendix 4.

1. Digital Elevation Model

Variable H

Definition
It is the variable height above mean sea level of the points which constitute the regular grid of the Digital Elevation Model. It is a quantitative variable with values depending on the study area (areas near the sea, mountainous areas, etc.). It is the variable from which it is possible derive the geometrical variables.

Function
In many cases this variable is used as an indirect means of accounting for spatial variations in temperature and/or precipitation.
Several researchers have found a close relationship between the amount of rainfall (the main triggering factor of shallow landslides in the study area) and the elevation (Carrara, 1983; Gallart and Clotet, 1988; Baeza, 1994). It has been demonstrate an increase of rainfall with altitude and a consequent increase in the number of slope failures.

2. Geometry

Variable β

Definition
It is the slope angle in the failure area, defined as the angle between the terrain surface and the horizontal. It is expressed in degree between in a range from 0 to 90. It is a quantitative variable which derives from DEM and therefore it is represented in ArcGis as a regular grid of floating points (without an associated table of attributes).
The Slope function calculates the rate of change between each cell and its neighbors, for example, the steepest downhill descent for the cell (the change in elevation over the distance between the cell and its eight neighbors). Every cell in the output raster has a slope value. The lower are the slope value, the flatter is the terrain; the higher are the slope value, the steeper is the terrain. The output slope raster can be calculated as percent of slope or degree of slope.
When the slope angle equals 45 degrees, the rise is equal to the run. Expressed as a percentage, the slope of this angle is 100 percent. Note that as the slope approaches vertical (90°), the percentage slope approaches infinity.

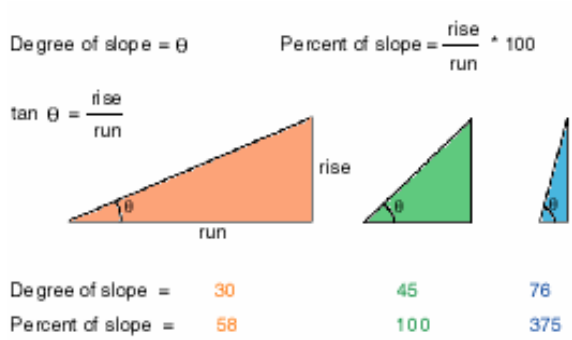


Figure 38. Examples of different slope angle values

Function

The slope angle is probably the main factor of stability as it affects the magnitude of both normal and shear stresses on the potential surface of failure.

The higher the angle the greater is the shearing component of the forces acting at the potential surface of failure (Jones et al., 1961).

Variable α

Definition

This variable defines the aspect of the scarp in a certain point, and it represents the direction that a slope faces. It identifies the steepest downslope direction from each cell of the grid to its neighbors. It can be thought of as the slope direction or the compass direction a hill faces. It is calculated from the angle between the geographic North and the horizontal projection of the vector normal to the surface in a certain point. It is a quantitative and continuous variable derived from the DEM, measured clockwise in degrees from 0 (due north) to 360 (again due north, coming full circle). The value of each cell in an aspect dataset indicates the direction the cell's slope faces. Flat areas having no downslope direction are given a value of -1.

It is represented in ArcGis as a regular grid of floating points.

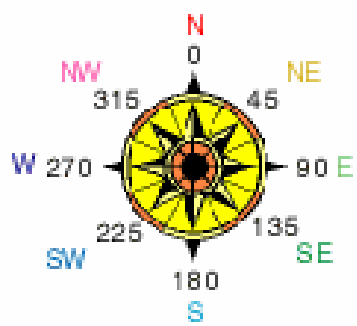


Figure 39. Representation of sloe aspect values

To avoid working with a circular statistical analysis, the variable's range [0-360°] has to be transformed in [0-180°]. The transformation is done as described below.

$$IF(\alpha < 180) \Rightarrow \alpha' = \alpha \qquad (25a)$$

$$IF(\alpha > 180) \Rightarrow \alpha' = 360 - \alpha \qquad (25b)$$

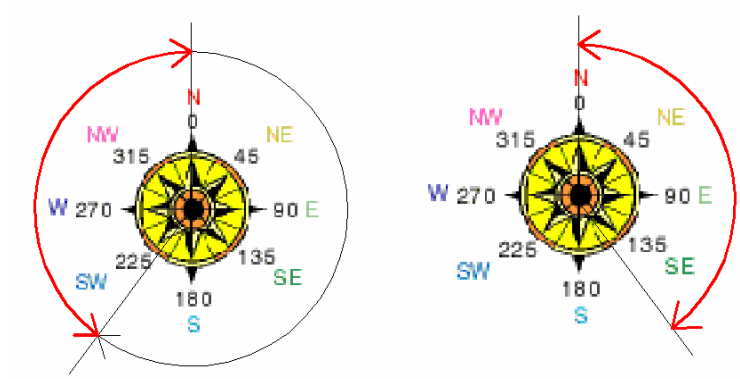


Figure 40. Transformation of slope aspect range

Function

Although, the relation between slope aspect and mass movement has long been investigated, no general agreement exists on the effect of aspect on landslide occurrence (Carrara et al. 1991). However, slope aspect is related to the general physiographic trend of the area and/or the main precipitation direction, and direction of the landslides is roughly perpendicular to general physiographic trend (Duman et al., 2006)**.** Beside the rainfall storm path, this variable is related also to available soil moisture and amount of vegetation. In particular, the aspect of the slope has an indirect influence on moisture content of the soil, which is related to the reduction of the effective stresses at the potential failure surface (Neuland, 1976; Carrara, 1983a).

Variable RUGOS

Definition

The RUGOS is the terrain roughness, defined here as the slope variation in a certain area (in this case in an area of 3×3 cells). This variable is here considered as the *Surface Ratio* for the land area contained within that cell's boundaries. This ratio can be calculated by dividing the surface area of the region by the planimetric area.

The *Surface Ratio Grid* will be a new, floating point grid whose cell values reflect the cell's surface area divided by the planimetric area of that cell. Surface areas are always greater than or equal to the planimetric area, so surface ratios will always be greater than or equal to 1. The grid has been generated using an extension available in Arc/Info.

In summary, this extension calculates surface ratios using the following strategy:

- For each cell in the grid, surface areas are based on triangle areas derived from eight triangles.
- Each triangle connects the center point of the central cell with the center points of two adjacent cells. These triangles are located in three-dimensional space, so that the area of the triangle represents the true surface area of the space bounded by the three points.
- The triangle area is adjusted so that it only represents the portion of the triangle that overlays the central cell.
- The areas of the eight triangles are summed to produce the total surface area of that cell.
- The surface ratio of the cell is calculated by dividing the surface area of the cell with the planimetric area of the cell.

The following figures explain as this methodology works. They represent an example of a Surface Ratio calculation, for a cell with an elevation value of 165 surrounded by cells at 190, 170, 155, 183, 145, 175, 160 and 122.

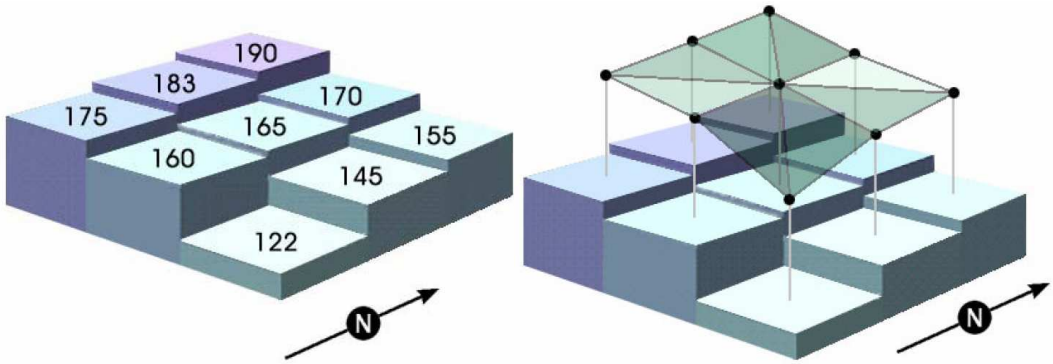


Figure 41. The *roughness* tool

This grid of elevation values can be pictured in 3-dimensional space as a set of adjacent columns, each rising as high as it's specified elevation value.
Then the 3-dimensional centerpoints of each of these 9 cells are evaluated, in order to calculate the lengths of the 8 lines that connect the central cell's centerpoint with the center points of the 8 surrounding cells. Then the lengths of the lines that connect each of the 8 surrounding cells with the one right next to it are calculated, to end up with the evaluation of the lengths of the sides of the 8 triangles that all meet at the center point of the central cell. Using these lengths, it is possible to calculate the areas of each of the triangles. By dividing the surface area with the planimetric area, the surface ratio is so evaluated.

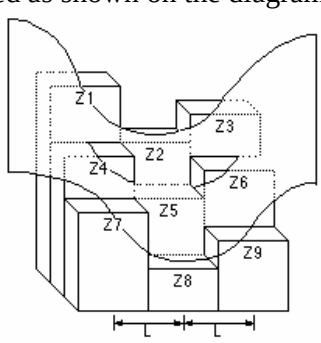
Function
High roughness slopes are more prone to landsliding because gradient changes favor rainfall infiltration into the soil and thus its instability.

Variable CURVAR

Definition
The variable CURVATURE represents the convexity/concave degree of terrain. It is obtained through the curvature ray of each cell by a vertical plane. It is defined as the change rate of slope and depends from the second derivative of the surface (i.e., the slope of the slope).
The curvature of a surface is calculated on a cell-by-cell basis. For each cell, a fourth-order polynomial is fit to a surface composed of a 3×3 window, in the form:

$$Z = Ax^2y^2 + Bx^2y + Cxy^2 + Dx^2 + Ey^2 + Fxy + Gx + Hy + I \qquad (31)$$

The coefficients a, b, c, and so on, are calculated from this surface.
The relationships between the coefficients and the nine values of elevation for every cell numbered as shown on the diagram are as follows:




```
A = [(Z1 + Z3 + Z7 + Z9) /4  - (Z2 + Z4 + Z6 + Z8) /2 + Z5] /L4
B = [(Z1 + Z3  - Z7 - Z9) /4  -  (Z2 - Z8) /2] /L3
C = [(-Z1 + Z3 - Z7 + Z9) /4 + (Z4 - Z6)] /2] /L3
D = [(Z4 + Z6) /2 - Z5] /L2
E = [(Z2 + Z8) /2 - Z5] /L2
F = (-Z1 + Z3 + Z7 - Z9) /4L2
G = (-Z4 +Z6) /2L
H = (Z2 - Z8) /2L
I = Z5
```

Figure 42. The *curvature* tool

This quantitative and continuous variable assumes different values, based on the study area and thus on the used DEM, and it is expressed in one over 100 z-units, or 1/100 (z-units). A positive curvature indicates that the surface is upwardly convex at that cell. A negative curvature indicates that the surface is upwardly concave at that cell. A value of zero indicates that the surface is flat. It is a variable derived from the DEM as a regular grid of floating points.

Function

Curvature can be used to describe the physical characteristics of a drainage basin in an effort to understand erosion and runoff processes.

The slope curvatures indicate the capability of water run-off concentration or dispersion. The subsurface flows concentrate in concave areas while disperse in convex areas. In the first case, it is expected to have an increase of pore water, element that favors a landslide.

Variable PF

Definition

This variable represents the convexity/concave degree of terrain surface in the direction of slope (longitudinal). This quantitative and continuous variable assumes different values, based on the study area and thus on the used DEM, and it is expressed in one over 100 z-units, or 1/100 (z-units).

A negative profile indicates that the surface is upwardly convex at that cell. A positive profile indicates that the surface is upwardly concave at that cell. A value of zero indicates that the surface is flat. It is a variable derived from the DEM as a regular grid of floating points.

Function

As the variable CURVAR, this one indicates the capability of the topography to concentrate or disperse the water run-off. As it represents the curvature in the slope direction, this variable affects water infiltration and the acceleration/deceleration of water flow, and therefore influences erosion and deposition.

Variable PL

Definition

This variable represents the convexity/concave degree of terrain surface perpendicular to the slope direction (transversal). It is described as the curvature of a contour line formed by the intersection of a horizontal plane with the surface. This quantitative and continuous variable assumes different values, based on the study area and thus on the used DEM, and it is expressed in one over 100 z-units, or 1/100 (z-units).

A positive plan indicates that the surface is upwardly convex at that cell. A negative plan indicates that the surface is upwardly concave at that cell. A value of zero indicates that the surface is flat. It is a variable derived from the DEM as a regular grid of floating points.

Function
This variable assumes the same function as the variables CURVAR and PF. In particular, as it represents the curvature direction perpendicular to the slope, the influence of plane curvature on the erosion processes is the convergence or divergence of water during downhill flow. In addition, this parameter constitutes one of the main factors controlling the geometry of the terrain surface where landslide is occurred.

3. Watershed dimension

Variable A

Definition
This variable is the watershed area, which is defined as an area that drains water to a common outlet as concentrated drainage. It represents the total area flowing to a given outlet, or pour point and it is calculated as the sum of the areas of those cells contributing to the flows towards a certain point. This variable derives from the DEM and assumes values from 0 to infinite. Even though it is a quantitative variable, the values of watershed areas are multiplier of the single cell area; so, it is not a continuous variable. The area is expressed in square meters.

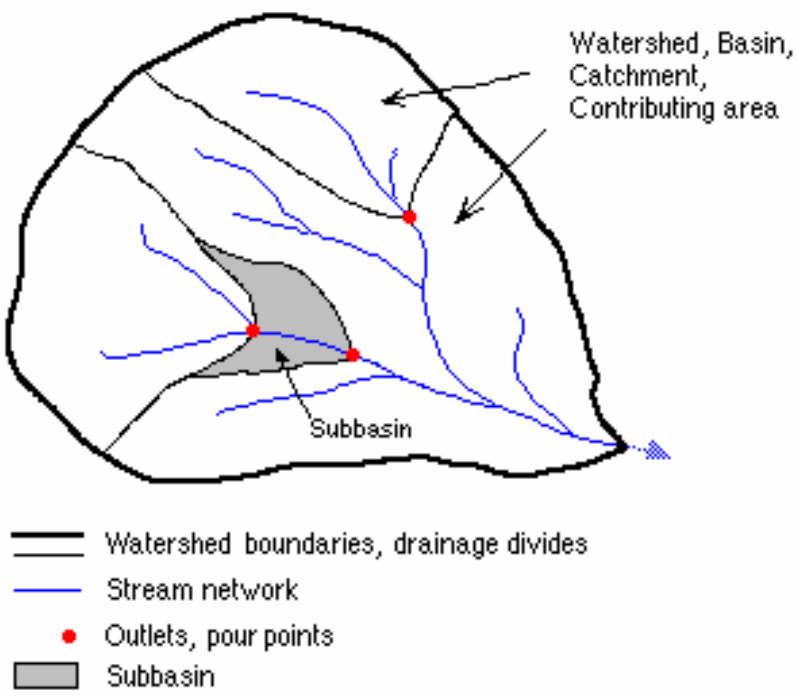


Figure 43. Watershed representation

Function
The watershed area is related to the amount of water that the soil can collect and infiltrate. A greater area is associated with more water infiltrated and higher chances of landsliding (Oyagi, 1984). It is proved that the watershed area influences the occurrence of slope instability.
Furthermore, the presence of colluvial deposits is more frequent in large watersheds.

Variable L

Definition
This variable indicates the maximum distance of the watershed area contributing to a certain cell. It represents the length of the longest flow path within a given basin, calculated on the slope direction and not on the horizontal projection. Therefore, the flow length of each cell is related to its slope angle. This quantitative variable derives from the DEM and assumes values from 0 to infinite. The flow length is expressed in meters.

Function
The length of watershed indicates the size of watershed area, the capability to concentrate groundwater and to accumulate sediments (superficial deposits). A close relationship has been found between the distance to water divide and location of the slope failure (Oyagi, 1984). The cells in proximity of water divide show less failure than the others.

Variable γ

Definition
The variable is the mean slope angle of the watershed contributing to a certain point. It is defined as the mean value of the slope angle of the watershed following the maximum slope direction. It is obtained from the DEM by algorithm and it is a quantitative and continuous variable.

Function
The mean slope angle of the watershed area indicates its capacity to help water infiltration into the soil and thus its instability.

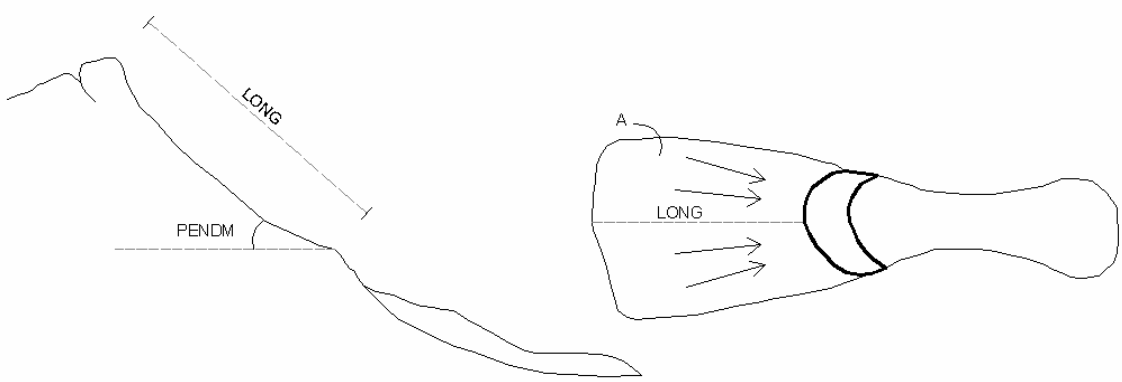


Figure 44. Variables related to watershed dimension

Variable SPI

Definition
The STREAM POWER INDEX represents the measure of erosive power of water flow. It is defined as:

$$SPI = As \cdot \tan \beta \qquad (32)$$

where *As* is the specific catchment area (previously defined as *A*) and β is the slope gradient in degree (previously defined as *PENDM*).
This definition is based on the ideal case of temporally and spatially uniform rainfall excess. Basing on this assumption, the steady-state discharge per unit width is directly proportional to the specific catchment area (As) (Moore et al. 1991), which is a measure of surface or shallow subsurface runoff at a given point on the landscape, integrating the effects of upslope contributing area and catchment convergence and divergence on runoff. Even though the ideal condition above described rarely exists in the natural environment, the assumed relationship is used extensively in hydrology.
This quantitative variable is a mathematic transformation of two variables previously defined (*A* and *PENDM*).

Function
This index is one of the main factors controlling the erosion processes. Besides, the erosion processes can be considered as one of the main conditioning factors of landslide occurrence.

4. Geology

Variable ST

Definition
The variable represents the typology of soil.
The geo-lithological map provided by the National Basin Authority distinguishes five homogeneous geolithological complexes in the study area:

- intercalate basin limestone and clays;
- quaternary clastic deposits;
- shelf limestone;
- sequence of sands, clays and conglomerates;
- flysh;

It is a qualitative variable, which has to be transformed in quantitative in order to be used in the statistical analysis.

Function
Lithology is one of the most influential parameters on slope instability, because each material has different shear strength and hydraulic conductivity. According to the typologies of soil present in the study area, it is expected that the high susceptibility is in fine grained arenaceous complex, with silty clays levels and in arenaceous-clayey complex, with varicolored clays.

5. Landslides

Variable SL

Definition
This variable represents the landslides identified in the study area. It is a variable that utilize a point or a polygon to localize the failure zone of each movements. It will be used as the grouping variable in the statistical analysis. It is a qualitative variable.

Function
The variable SL shows the presence or absence of landslides and it is used as grouping variable in the statistical analysis, in order to establish the relationship with the conditioning factors. It shows the evidence of past movements, on which is based the forecast.

Comparing procedures
The two adopted methodologies have both advantages and drawbacks. The use of a point rather than a polygon makes the work computationally simpler. On the other hand, in few cases it creates the problem of zero-values parameters, mostly for the variables related to the water flow. The use of a buffer, although it is more complex computationally, avoids the zero-values problem. Moreover the procedure is conceptually more correct. As different cells belong to the buffer zone, the extraction of many values of the variables related to the slope instability may give better contributions rather than the use of a unique value related to a single point. Thus, the use of buffer zones, used to extract the values of the parameters that are expected to be responsible for the landslide mechanism, is conceptually more accurate. Notwithstanding that, in order to create a susceptibility map it is necessary to associate the variables values to points and not to polygons. If we get the values from the buffer zones and assign the mean value extracted to the whole polygon representing the scarp buffer, the resulting susceptibility map will be constituted by polygons. A unique value of the susceptibility degree will be assigned to these polygons, which will have the buffer shape. In order to overcome this problem, the use of points rather than buffer zone, even though physically less exact, is favorable.

According to the previous observations about advantages and drawbacks in the described methodologies, a new approach is followed in the definition of the parameter values to the selected movements. The mean values extracted from the buffer zones have been assigned to the points localized at the highest altitude. In this way, the disadvantages of the two procedures are overcame, so to work with points having more significant values.

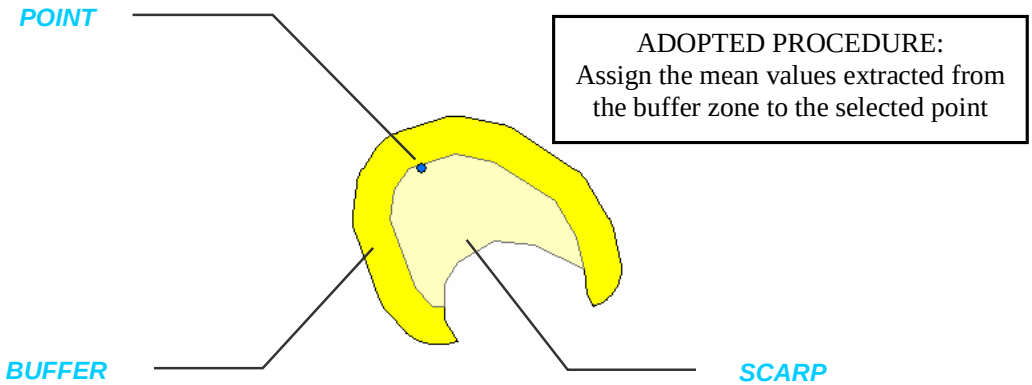


Figure 45. Adopted procedure

In order to achieve this goal, it is necessary to verify that the two procedures give almost the same results. Two tables with the comparison between the values extracted from points and the mean values extracted from buffer zones are reported below. The first distinguishes the three typologies of activity states considered, while the second distinguishes also the typology of movement.

Table XXXIIa. Comparison between points and buffers approaches

Activity state	Approach	β	A	α	CURVAR	PF	PL	LONG	PENDM	RUGOS	SPI
ACTIVE	buffer	12,40	1700,53	107,58	0,05	-0,02	0,03	76,65	12,23	1,04	334,79
	point	13,57	609,88	97,51	-0,18	0,06	-0,12	56,19	12,89	1,04	123,06
DORMANT	buffer	11,73	1634,695	106,43	0,04	-0,02	0,02	75,64	11,58	1,03	300,65
	point	12,65	620,78	93,74	-0,17	0,07	-0,10	56,90	12,01	1,03	114,85

Table XXXIIIb. Comparison between points and buffers approaches

Activity state	Type of movement	Approach	β	A	α	CURVAR	PF	PL	LONG	PENDM	RUGOS	SPI
ACTIVE	earthflows	buffer	11,90	1560,27	107,67	0,04	-0,02	0,03	75,43	11,77	1,03	298,16
		point	13,20	605,78	97,85	-0,23	0,09	-0,14	55,67	12,57	1,04	119,92
	slides	buffer	13,10	1860,91	106,47	0,08	-0,04	0,03	77,48	12,74	1,04	378,4
		point	14,15	616,40	97,77	-0,07	-0,02	-0,08	56,57	13,25	1,04	126,28
	earthflows-slides	buffer	12,62	1937,18	110,73	0,00	0,02	0,02	80,04	12,85	1,04	397,51
		point	13,56	609,55	94,46	-0,31	0,16	-0,16	57,87	13,48	1,04	130,20
DORMANT	earthflows	buffer	11,31	1494,94	106,03	0,05	-0,02	0,03	73,98	11,10	1,03	265,79
		point	12,24	636,26	93,04	-0,18	0,06	-0,12	57,75	11,50	1,03	113,47
	slides	buffer	12,43	1804,51	107,27	0,04	-0,02	0,02	77,13	12,26	1,04	352,17
		point	13,46	577,72	95,59	-0,13	0,05	-0,07	54,09	12,90	1,04	114,51
	earthflows-slides	buffer	11,39	1659,87	105,81	0,03	-0,01	0,02	76,59	11,42	1,03	295,54
		point	12,18	664,66	92,14	-0,24	0,12	-0,13	60,04	11,61	1,03	118,46

As expected, the mean values extracted from the buffer zones are sufficiently comparable to the ones extracted from the points, except for the variables related to the amount of water flowing towards the point/buffer.
The parameter *A* takes in account the numbers of cells which contribute to the water flow towards the selected point. Using the points procedure, the contribute area is calculated once, for the selected point at the highest elevation. Using the buffers procedure, instead, the calculation is repeated different times, for each point within the buffer zone.
Thus, the mean values of the parameters extracted from the buffer areas are always greater than the ones calculated for the points. The same happen for the variables *LONG* and *SPI.*
We can conclude that the proposed approach, based on the extraction of the mean values from the buffer areas and the assignment of these values to the points, is sufficiently correct.

5.3.3 Statistical treatment

The statistical treatment was carried out using the SPSS Inc.(1988) statistical package, following the steps described in the 3.3 point.

Creation of the sample

In order to develop the statistical analysis, a sample from the populations of failed slopes (cells with rupture) and a sample from the populations of unfailed slopes (cells without rupture) must be selected.
The number of cells showing slope failures is much smaller that the number of cells of the study area (about 15000000 cells). In order to avoid the bias of the function obtained, the discriminant analysis requires population sets having a similar number of individuals (Dillon

and Goldstein, 1986). Therefore, because the number of failed cells is small, the sample set of unfailed slopes must be also small. Consequently, to derive the discriminant function, we have selected a random sample of nearly 9000 unfailed cells.
The sample was obtained by the GIS with an algorithm created for sampling.
From this sample, a sub-sample of unfailed cells has to be selected, in order to work with the same number of individuals in the failed cells' samples, distinguished for the typology of movements of landslides.

As the charateristics of the typologies of movements are different, different disciminant analysis have to be used. In each failed/unfailed sample a certain number of cells (nearly the 50%) are chosen to create the discriminant function, while the remaining failed and unfailed cells will be used in the step of validation of the discriminant function.
This is done through the creation of a variable in the SPSS package. This variable, called "validate", takes two values: the cells with 1 will be used to create the discriminant function, while the cells with 0 will be used to validate it.

Transformation of variables
As already said, the discriminant analysis has difficulties in dealing with qualitative data. Therefore, it is necessary to transform qualitative variables in quantitative variables.
Regarding the grouping variable SL, it has been transformed assigning the value 1 to the class of failed cells (with the presence of movements) and the value 0 to the class of unfailed cells.

As regard the variable *SOIL TYPE*, it has been transformed from qualitative to quantitative parameter through an index of relative landslide density. This index is defined by the ratio between the density of slope failures in a given lithology and the overall slope failure density.

The index is calculated for each typology of movements (earthflows, slides and complex phenomena of slide-earthflows) and takes the following form:

$$I_i = \frac{a_i / A_i}{\sum_i (a_i / A_i)} \cdot 100 \qquad (33)$$

where a_i is the area occupied from the scarps of a given typology within a certain lihology and A_i is the area occupied by that typology.

It may be expected that the lithology showing the greatest value of this index is the most prone to landsliding and thus the most contributing to the slope instability.
The table XXXIV shows the transformed variables.

Table XXXIV. Transformation of the variables

Variable	Value			Observations
ST – Soil type	*Earthflows*	*Slides*	*Complex*	
Flysch	39.67%	40.84%	40.22%	Higher values express higher proneness to fail.
Sequence of sands, clay and conglomerate	34.31%	24.68%	23.97%	
Intercalate basin limestone and clays	16.09%	18.44%	21.23%	
Shelf limestone	6.03%	10.13%	9.14%	
Quaternary clastic deposits	3.90%	5.91%	5.44%	
SL – Grouping variable				
Unfailed slope		1		Variable used in discriminant analysis to classify the sample.
Inunfailedslope		0		

The next steps of the statistical analysis will be done separately for the three typologies of movements.

EARTHFLOWS

Parameters description

The first step is creating a general description table, in order to give information about the mean value assumed by all the selected parameters in the two groups of failed and unfailed cells.

As regards the values assumed by the variable β (slope angle) in the group of failed cells, considering that the majority of movements occur in clayey materials, it is demonstrate in literature that earthflows present on slope with a slope angle lower than 6° can not exist. Their presence in the database can be considered as errors due to the program or other inaccuracies. Therefore, we have decided to eliminate the earthflows with slope angle lower than 6°. As a consequence, as it is recommended to work with the same range of values in the two groups of failed and unfailed cells, we have decided to eliminate from the unfailed cells caught randomly the ones having β lower than 6° and the ones having β. The same for the maximum value of slope angle (45°).

Below the processes employed are described.

Failed cells → In this case, from the population of earthflows the ones having β<6° were eliminated. The population results reduced from N=2136 to N=1962.

Unfailed cells → In this case, in order to avoid a great reduction in number of the unfailed cells, a new sample has been created. After having previously selected the unfailed cells having 6°<β<45°, a sample of 2580 cells has been obtained.

The table XXXV shows the main statistical information about the two populations.

Table XXXV. Statistics

FAILED	H	β	α'	CURVAR	PF	PL	PENDM	A	LONG	SPI	RUGOS	ST
N	1962	1962	1962	1962	1962	1962	1962	1962	1962	1962	1962	1944
Mean	470,01	12,06	106,31	0,04	-0,01	0,02	11,86	4975,45	92,60	1037,74	1,03	34,56
Std. Error of Mean	4,59	0,10	0,96	0,006	0,0043	0,0042	0,105	704,36	1,88	149,06	0,00069	0,22
Median	428,25	11,03	106,79	0,03	-0,016	0,023	10,73	1315,39	74,82	257,86	1,02	39,67
Mode	275,00(a)	9,58(a)	90,00	0,00	0,00	0,00	8,46(a)	700,00	128,89	21,02(a)	1,02	39,67
Std. Deviation	203,33	4,56	42,58	0,30	0,19	0,18	4,6	31199,56	83,63	6602,78	0,03	9,99
Variance	41343,76	20,81	1813,14	0,095	0,037	0,035	21,73	973413080,90	6995,18	43596824,14	0,001	99,95
Skewness	0,706	1,624	-0,113	-0,603	1,327	-0,287	1,565	23,508	7,740	23,362	4,245	-1,842
Std. Error of Skewness	0,055	0,055	0,055	0,055	0,055	0,055	0,055	0,055	0,055	0,055	0,055	0,056
Kurtosis	0,049	4,143	-0,965	5,716	12,626	8,593	3,514	711,093	110,914	708,387	29,781	1,875
Std. Error of Kurtosis	0,110	0,110	0,110	0,110	0,110	0,110	0,110	0,110	0,110	0,110	0,110	0,111
Minimum	74,557	6,00	0,00	-2,355	-1,109	-1,548	4,19	100,00	14,67	21,02	1,01	3,90
Maximum	1394,46	42,66	180,00	1,47	1,79	1,36	39,72	1060522,50	1725,24	224890,99	1,41	39,67

UNFAILED	H	β	α'	CURVAR	PF	PL	PENDM	A	LONG	SPI	RUGOS	ST
N	2580	2580	2580	2580	2580	2580	2580	2580	2580	2580	2580	2580
Mean	520,54	15,07	96,58	-0,48	0,19	-0,29	14,55	15765,03	130,05	3419,39	1,05	27,29
Std. Error of Mean	5,60	0,163	1,025	0,036	0,022	0,019	0,162	6436,17	5,831	1204,75	0,001	0,298
Median	469,00	12,017	96,34	0,00	0,00	0,00	11,70	600,00	63,80	143,04	1,02	39,67
Mode	287,00	8,04	135,00	0,00	0,00	0,00	7,29	100,00	14,26	12,81	1,01	39,67
Std. Deviation	284,628	8,28	52,08	1,85	1,14	0,99	8,24	326917,02	296,22	61193,96	0,06	15,17
Variance	81013,21	68,69	2712,52	3,44	1,30	0,98	68,04	106874739320,08	87748,42	3744701331,05	0,004	230,409
Skewness	0,735	1,234	-0,183	-1,146	0,695	-1,761	1,148	43,739	13,986	40,626	2,499	-0,596
Std. Error of Skewness	0,048	0,048	0,048	0,048	0,048	0,048	0,048	0,048	0,048	0,048	0,048	0,048
Kurtosis	0,070	0,864	-1,100	10,820	13,582	13,426	0,689	2076,284	307,852	1853,707	7,036	-1,543
Std. Error of Kurtosis	0,096	0,096	0,096	0,096	0,096	0,096	0,096	0,096	0,096	0,096	0,096	0,096
Minimum	44,00	6,05	0,00	-18,00	-10,00	-11,61	1,56	100,00	10,05	7,90	1,01	3,90
Maximum	1714,00	44,75	180,00	11,00	10,78	4,81	45,76	15729900,00	8573,72	2860758,72	1,44	39,67

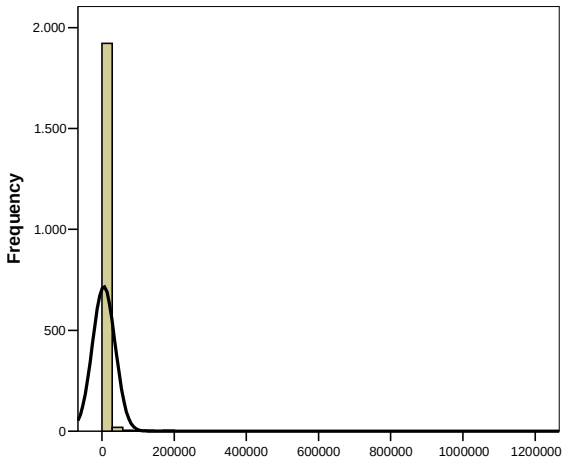
a Multiple modes exist. The smallest value is shown

Testing for normal distribution

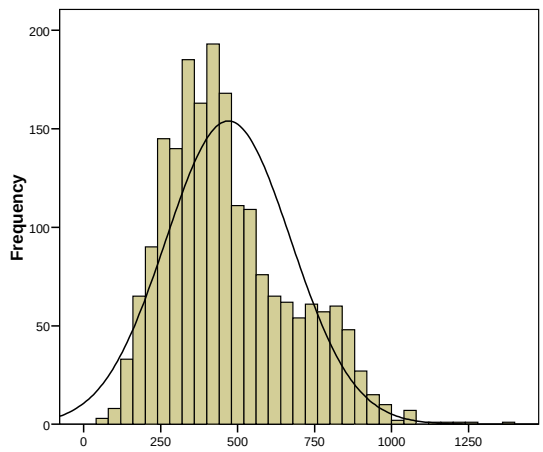
In order to work with normal distributions of the variables, the normalization test has been developed. To observe if each variable has a normal distribution, the histogram of each one has been previously created, and then the test of Kolmogorov-Smirnov has been applied.

As regards the histograms, two graphs have been developed for each variable: the first refers to the population of unfailed cells, the second to the group of failed cells (cells with the presence of earthflows). From them, it is possible to observe that some variables have a normal distribution and others no. As examples, two graphs have been reported below (for the others cfr. Appendix 2).

A – WATERSHED AREA



H – ELEVATION



In order to confirm all that has been observed from the histograms, a KS test has been developed. In the table XXXVI the values of the KS test and the significance level have been reported. The results are interpreted in the way that the lower is the KS value, the better is the adjustment of the variable distribution to the normal one. On the contrary, the greater is the KS value, the worse is the adjustment. The greatest values of KS are highlighted in orange.

Table XXXVI. K-S test results – without transformation

Variable	Z de Kolmogorov-Smirnov
H	5,589
β [°]	10,421
A	32,501
α' [°]	5,074
CURVAR	13,630
PF	13,577
PL	14,629
LONG	21,991
PENDM [°]	9,878
RUGOS	16,939
SPI	32,315
ST	22,986

The log-transformation is necessary for the variable *A*, *LONG*, *RUGOS* and *SPI*. The new variables will be renamed, using the term "log" before the name of the variable.
Regarding the parameter *ST*, the transformation is not necessary because it is a qualitative parameter, previously transformed in quantitative, that assumes only five values.
The results in term of KS values are reported in the table XXXVII. A lower value of the KS test is observed in all the transformed variables.

Table XXXVII. K-S test results – with transformation

	Before transformation		After transformation
Variable	Z de Kolmogorov-Smirnov	Variable	Z de Kolmogorov-Smirnov
H	5,589		5,589
β [°]	10,421		10,421
A	32,501	log A	3,713
α' [°]	5,074		5,074
CURVAR	13,630		13,630
PF	13,577		13,577
PL	14,629		14,629
LONG	21,991	log LONG	2,083
PENDM [°]	9,878		9,878
RUGOS	16,939	log RUGOS	16,143
SPI	32,315	log SPI	3,643
ST	22,986		22,986

For each variable two graphs have been developed: the first refers to the population of unfailed cells, the second to the group of failed cells (cfr Appendix 2).

It can be observed that the frequency distribution of all the variables in the population of failed cells (cells representing earthflows) is almost normal, while it doesn't adjust well to a normal distribution in the group of unfailed cells. The assumption of normal distribution that is at the basis of discriminant analysis (cfr. 3.1.1 point) seems to be violated. However, the large number of studies that have utilized linear discriminant analysis without any concern for whether this assumption was satisfied supports the contention that robustness is not a problem (Dillon and Goldstein, 1984).

Selection of independent variables
Having verified the normal distribution of the variables, the next step consists in verifying the independence of the variables. A bivariate analysis of correlation will be first done, in order to search for possible relationship between the variables.
Then the principal component analysis (PCA) will be applied, in order to view the membership of each variable to the different factors.

- **Bivariate Pearson correlation**

The table with the coefficient of Pearson is reported below. The variable showing high correlation have the coefficient colored in yellow.

- β, *PENDM* and *RUGOS*

It can be observed that the slope angle and the mean watershed angle are strongly correlated. It means that the change in slope angle between the watershed area and the scarp area is little.

Moreover, these variables are correlated also to the roughness of the terrain. It can be explained considering the way this parameter is obtained, that is the ratio between the surface area (depending on the slope angle) and the planimetric area.

- *logA*, *logLONG* and *logSPI*

Regarding the parameters that define the watershed dimension, as expected the three variables representing the watershed area, the watershed length and the stream power index are strongly correlated. It means that the watersheds have a dimension so that the greater is the length, the greater is the area. As a consequence, we expect not to find very slim watersheds, with high values of watershed length and little values of watershed area, nor very squat watersheds, with little values of watershed length and great values of watershed area.

- *CURVAR*, *PF* and *PL*

Regarding the parameters defining the curvature, the global, profile and plan curvature are strongly correlated. It means that the characteristics explained from these variables are the same. The behavior of the scarp in term of concavity/convexity is the same if we consider the global one, the profile one or the plan one.

Table XXXVIII. Pearson coefficients of correlation matrix

	H	β [º]	log A	α' [º]	CURVAR	PF	PL	log LONG	PENDM [º]	log RUGOS	log SPI	ST
H	1											
β [º]	0,243(**)	1										
log A	-0,026	-0,129(**)	1									
α [º]	0,062(**)	-0,094(**)	0,040(**)	1								
CURVAR	-0,051(**)	-0,071(**)	-0,202(**)	0,019	1							
PF	0,018	0,068(**)	0,080(**)	-0,019	-0,886(**)	1						
PL	-0,074(**)	-0,056(**)	-0,285(**)	0,014	0,851(**)	-0,512(**)	1					
log LONG	-0,014	-0,036(*)	0,926(**)	0,008	-0,208(**)	0,085(**)	-0,290(**)	1				
PENDM [º]	0,233(**)	0,889(**)	-0,074(**)	-0,084(**)	-0,126(**)	0,155(**)	-0,059(**)	0,020	1			
log RUGOS	0,245(**)	0,960(**)	-0,101(**)	-0,091(**)	-0,111(**)	0,098(**)	-0,094(**)	-0,021	0,845(**)	1		
log SPI	0,044(**)	0,160(**)	0,943(**)	0,014	-0,231(**)	0,123(**)	-0,289(**)	0,900(**)	0,251(**)	0,159(**)	1	
ST	-0,122(**)	-0,423(**)	0,080(**)	0,122(**)	0,081(**)	-0,048(**)	0,097(**)	0,022	-0,441(**)	-0,408(**)	-0,059(**)	1

(**) The correlation is significative at 0.01 level (bilateral).
(*)The correlation is significative at 0.05 level (bilateral).

- **Factorial analysis**

The principal component analysis has been developed with the 12 variables considered. This method, usually used to reduce the number of variables to work with, is applied here only to find possible correlation among the variables.
The goodness of the results depends on total variance explained. In this case the first 4 factors explain nearly the 80 per cent of the variance (cfr. Table XXXIX).

Table XXXIX. Total variance explained in the Principal Component Analysis
(Results of the sum of the square saturation of the rotation)

Component	Total	% of the Variance	% Cumulative
1	3,194	26,617	26,617
2	2,875	23,959	50,576
3	2,513	20,944	71,520
4	1,082	9,018	80,538

From the results obtained in the table XL it is clear that the total variance of each variable is greater then the 44% (with the variable *H*), being the variables *CURVAR*, *logA* and *logSPI* the best representative, with more than 97% of the variance, followed by β and *logLONG* with more than 92%. The results are interpreted in the way that the greater is the extraction, the more independent is the variable and the lower is the correlation with the others.

Table XL. Communalities or total variance of each variable
(Extraction method: Principal component analysis)

Variable	Extraction	Variable	Extraction
H	0,445	PL	0,748
β [°]	0,928	log LONG	0,935
log A	0,979	PENDM [°]	0,870
α' [°]	0,718	log RUGOS	0,894
CURVAR	1,000	log SPI	0,973
PF	0,794	ST	0,379

The rotated matrix with the factors is presented in the table XLI, where the greater loads are in yellow.

Table XLI. Varimax rotated component matrix (Extraction method: Principal component analysis)

	Factor			
	1	2	3	4
H	0,292	-0,015	-0,039	0,599
β [°]	0,959	-0,007	-0,008	0,093
log A	-0,126	0,976	-0,102	0,002
α' [°]	-0,208	0,021	0,020	0,821
CURVAR	-0,061	-0,112	0,992	-0,008
PF	0,064	-0,018	-0,889	-0,008
PL	-0,040	-0,227	0,833	-0,025
log LONG	-0,032	0,961	-0,104	-0,018
PENDM [°]	0,925	0,054	-0,061	0,084
log RUGOS	0,940	0,001	-0,048	0,096
log SPI	0,171	0,965	-0,113	0,028
ST	-0,593	0,028	0,061	0,153

FACTOR 1: This factor represent nearly the 27% of the total variance and is defined by the variables β, *PENDM* and *RUGOS*. In this case the factor is defined by more than one variable, with a great saturation value (weight of the variable in the factor). It implies that a high dependence between the variables exists. Actually, as observed in the bivariate correlation matrix, these variables are strongly correlated. The meaning can be found considering that these variables are representative of the slope direction of the moving mass.

FACTOR 2: This factor represent nearly the 24% of the total variance and is defined by the variables *logA*, *logLONG* and *logSPI*. Also in this case the factor is defined by more than one variable, with a great weight of the variable in the factor. It implies that the variables are strongly correlated, as observed in the bivariate correlation matrix. The meaning can be found considering that these variables are representative of the geometric characteristics of the watersheds.

FACTOR 3: This factor represent nearly the 21% of the total variance and is defined by the variables *CURVAR*, *PF* and *PL*. The three variables are strongly correlated and it seems to explain the same condition of concavity/convexity of the scarp zones.

FACTOR 4: This factor represent nearly the 9% of the total variance and is defined by the variables α and *H*, representing respectively the aspect and the elevation of the scarp.

The graph showing the first three factors in a rotated space is presented below.

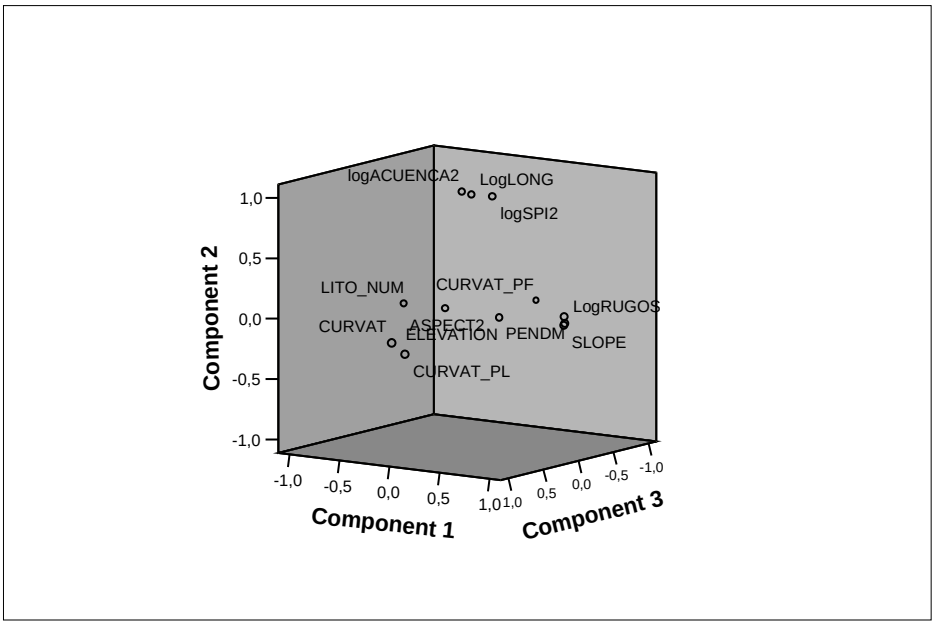


Figure 46. Component plot in a rotated space

After having identified the dependent variables, before rejecting one of these, it is recommended to make a contrast analysis between the two populations, in order to identify which is the variable that best characterize each population and thus influence the separation of the two populations.
The contrast analysis is developed using two additional tests, **One-way test** and **T-test**, developed in order to understand the influence of each variable on stability.
The Tables of the results of One-way test and T-test are reported below.

Table XLII. T-test and One-way test results

Variable	Activity state	Mean	Standard Deviation	One-way test		T-test	
				F	Sig.	t	Sig.
H	0	520,54	284,63	242,566	0.000	6,975	0.000
	1	470,02	203,33				
β [°]	0	15,08	8,29	645,780	0.000	15,616	0.000
	1	12,07	4,56				
log A	0	2,87	0,67	200,638	0.000	-17,912	0.000
	1	3,18	0,47				
α' [°]	0	96,58	52,08	136,772	0.000	-6,921	0.000
	1	106,31	42,58				
CURVAR	0	-0,49	1,86	1286,871	0.000	-14,199	0.000
	1	0,042	0,31				
PF	0	0,19	1,14	1069,570	0.000	9,002	0.000
	1	-0,01	0,19				
PL	0	-0,29	0,99	1115,965	0.000	-16,087	0.000
	1	0,03	0,19				
log LONG	0	1,81	0,48	724,034	0.000	-5,810	0.000
	1	1,87	0,26				
PENDM [°]	0	14,55	8,25	615,534	0.000	13,910	0.000
	1	11,86	4,66				
log RUGOS	0	0,02	0,03	545,412	0.000	14,696	0.000
	1	0,01	0,01				
log SPI	0	2,22	0,72	201,427	0.000	-13,909	0.000
	1	2,48	0,51				
ST	0	27,30	15,18	1595,008	0.000	-19,362	0.000
	1	34,56	10,00				

According to all the analysis done in order to select independent variables, we have rejected the correlated variable showing the lowest significance on the stability of the slope by the T-test and One-way test.

Results displayed by T-test (Table XLII) show that all the variable have high statistical significance, as the means of the failed slope and unfailed slope groups are different (in all cases Sig=0.00, that means that significant differences in mean between the two populations are present). Therefore, all the variables can be used in the successive step of defining the discriminant function.
One-way test indicates which of the correlated variables best characterizes the stability of the slope, according to the value assumed by the parameter F.
As regard the variables defining the watershed dimension, according to the F value we kept variable *logLONG* (F=724.034) in front of *logA* (F=200.638). Actually, the greater is the F value, the greater is the discriminant power of that variable among the two populations.
Moreover, we decide to use the variable *logSPI*.
As regard the variables defining the curvature, the variables *CURVAR*, *PF* and *PL* have F values so close (F=1286,871 , F=1069,570 and F=1115,965 respectively) that we decided to follow the procedure of discriminant analysis using the three variables separately.

Finally, regarding the variables defining the slope direction, we decide to use both β and PENDM in the discriminant analysis, in order to evaluate the variable that shows the greater importance (with F=645,780 and F=615,534 respectively).

Obtaining the discriminant function

From PCA, a data set including the most significant independent variables was taken as input of the discriminant analysis. In order to keep the independent variables assumption, the analysis was performed different times.

The first discriminant analysis was performed with the following input variables: *H*, β, α', *PL*, *logLONG*, *PENDM*, *logSPI*, *logRUGOS* and *ST*.

The discriminant function (*DF1*) obtained and the main statistical parameters are shown in Table L. Elevation (*H*), slope angle (β), slope aspect (α') and roughness (*logRUGOS*) are the most ininfluential variables (*H* has a weight of -0.085 and β, α' and *logRUGOS* have been rejected), while the most influential variables are the stream power index (*logSPI*) with 2.552 and the watershed length (*logLONG*) with -2.024. Mean watershed angle (*PENDM*), plan curvature (*PL*) and soil type (*ST*), with lower weights, were also chosen.

A new discriminant function (*DF2*) was obtained using the same variables except elevation (*H*) and the ones rejected in *DF1* (β, α' and *logRUGOS*). The stepwise method gives nearly equal results, as the Wilks λ increases insignificantly, while the percentage of correctly classification remains almost the same (Table XLIII) The most influential variables are the stream power index (*logSPI*) with 2.557, the watershed length (*logLONG*) with -2.022 and the mean watershed angle (*PENDM*) with -0.810. Plan curvature (*PL*) and soil type (*ST*), with lower weights, were also chosen.

A final analysis was performed without the variable plan curvature (*PL*), that is the last caught in the *DF2*, in order to evaluate its importance. The stepwise method gives worse results, as the Wilks λ increases from 0.662 to 0.716 and the percentage of correctly classification decreases.

The table below presents the main statistical characteristics of the analyses developed. The best discriminant analysis up to now is shown in yellow.

Discriminant	DF1	DF2	DF3

Table XLIII. DF1, DF2 and DF3 discriminant function results

	input	used*	rejected	Standardized coefficient	input	used*	rejected	Standardized coefficient	input	used*	rejected	Standardized coefficient
H	*	6		-0.085								
β	*		*									
logA												
α'	*		*									
CURVAR												
PF												
PL	*	5		0.494	*	5		0.502				
logLONG	*	3		-2.024	*	3		-2.022	*	3		2.226
PENDM	*	4		-0.790	*	4		-0.810	*	4		0.839
logRUGOS	*		*									
logSPI	*	2		2.552	*	2		2.557	*	2		-2.637
ST	*	1		0.321	*	1		0.321	*	1		-0.390
Wilks lambda λ	0.660				0.662				0.716			
% correctly classified general	81.1				81.4				77.1			
% correctly classified Unfailed	81.1				81.4				78.5			
% correctly classified Failed	81.0				81.3				75.4			
Function at group centroid - Unfailed	-0.642				-0.640				0.564			
Function at group centroid - Failed	0.801				0.799				-0.704			

used*: the number indicates the order of capture

Regarding the variables defining the curvature's characteristics of the two populations, on the basis of the variables used in the discriminant analysis *DF2*, other two analyses were made: one incorporating the variable *PF* (*DF4*), the other one with *CURVAR* (*DF5*).
The results (tables XLIV and XLV) demonstrate that the variable showing the better discriminant power is the plan curvature. Therefore, the variable *CURVAR* and *PF* were rejected in front of the variable *PL*.

The best discriminant analysis up to now is shown in yellow.

Table XLIV. DF2 and DF4 discriminant function results

Discriminant Function / Variables	DF2			DF4		
	input	used*	Standardized coefficient	input	used*	Standardized coefficient
H						
β						
logA						
α'						
CURVAR						
PF				*	5	0.236
PL	*	5	0.502			
logLONG	*	3	-2.022	*	3	2.203
PENDM	*	4	-0.810	*	4	0.795
logRUGOS						
logSPI	*	2	2.557	*	2	-2.628
ST	*	1	0.321	*	1	-0.390
Wilks lambda λ	0.662			0.704		
% correctly classified general	81.4			76.7		
% correctly classified Unfailed	81.4			77.4		
% correctly classified Failed	81.3			75.9		
Function at group centroid - Unfailed	-0.640			0.580		
Function at group centroid - Failed	0.799			-0.723		

used*: no one of the variables has been rejected and the number indicates the order of capture.

Table XLV. DF2 and DF45discriminant function results

Discriminant Function / Variables	DF2			DF5		
	input	used*	Standardized coefficient	input	used*	Standardized coefficient
H						
β						
logA						
α'						
CURVAR				*	5	
PF						
PL	*	5	0.502			
logLONG	*	3	-2.022	*	3	
PENDM	*	4	-0.810	*	4	
logRUGOS						
logSPI	*	2	2.557	*	2	
ST	*	1	0.321	*	1	
Wilks lambda λ	0.662			0.680		
% correctly classified general	81.4			79.0		
% correctly classified Unfailed	81.4			79.2		
% correctly classified Failed	81.3			78.6		
Function at group centroid - Unfailed	-0.640			-0.613		
Function at group centroid - Failed	0.799			0.766		

used*: no one of the variables has been rejected and the number indicates the order of capture.

Finally, regarding the two correlated variables defining the watershed dimension (*logLONG* and *logA*), on the basis of the variables used in the discriminant analysis *DF2*, a comparation was made using *logA* in front of *logLONG*, defining a new discriminant function (*DF6*). The results (table XLVI) demonstrate that the variable showing the better discriminant power is the *logLONG*.

Table XLVI. DF2 and DF6 discriminant function results

Variables \ Discriminant Function	DF2			DF6		
	input	used*	Standardized coefficient	input	used*	Standardized coefficient
H						
β						
logA				*	2	-3.396
α'						
CURVAR						
PF						
PL	*	5	0.502	*	3	0.618
logLONG	*	3	-2.022			
PENDM	*	4	-0.810	*	4	-1.586
logRUGOS						
logSPI	*	2	2.557	*	5	4.300
ST	*	1	0.321	*	1	0.367
Wilks lambda λ	0.662			0.780		
% correctly classified general	81.4			75.0		
% correctly classified Unfailed	81.4			68.4		
% correctly classified Failed	81.3			83.2		
Function at group centroid - Unfailed	-0.640			-0.475		
Function at group centroid - Failed	0.799			0.593		

The table XLVI presents in yellow the discriminant function considered as the best discriminating. It has been chosen because contains the smaller number of variables which explain the behavior of the selected cells regarding the slope stability/instability.
The main statistical parameters of the selected function (*DF2*) are reported in the table XLVII.

Table XLVII. DF2 discriminant function results

Variables	Function coefficients Standard	Unstandard	Correctly classified (%)
PL	0.502	0.696	General, 81.4
logLONG	-2.022	-5.074	Partial (unfailed-failed), 81.4 – 81.3
PENDM	-0.810	-0.116	
logSPI	2.557	4.043	Group centroids
ST	0.321	0.024	Unfailed slopes, -0.640
Constant		0.814	Failed slopes, 0.799

Positive discriminant coefficients are associated with failed slopes and negative ones with unfailed slopes.

Therefore, high values of *PL*, *logSPI* and *ST* increase discriminant scores and, consequently, slope instability. High values of *logLONG* and *PENDM*, with a negative coefficient, increase stability.
This result is in accord with the mean value assumed by the selected variables in the two populations of failed and unfailed cells, as reported in the descriptive table below.

Table XLVIII. Mean values of the selected variables

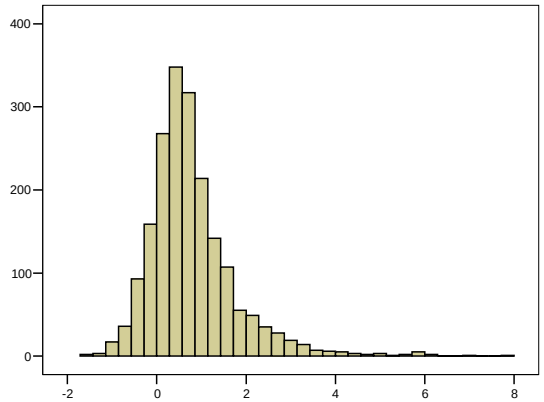
	PL	PENDM	ST	logLONG	logSPI
Activity=0	-0,29	14,55	27,30	1,81	2,22
Activity=1	0,03	11,86	34,56	1,87	2,48

Regarding the parameters *PL*, *ST* and *logSPI*, the mean value assumed in the unfailed cells is lower than the mean value in the category of failed cells. It means that the coefficients related to these variables are expected to be positive like the centroid of failed slopes.
In this case the coefficients obtained in the discriminant analysis are in accord with the described situation.
Regarding the parameter *PENDM*, instead, the mean value assumed in the unfailed cells is greater than the mean value in the category of failed cells. It means that the coefficients related to this variables is expected to be negative like the centroid of unfailed slopes. Also in this case the discriminant analysis gives correct results.
Finally, regarding the parameter *logLONG*, the mean values assumed in the two categories of failed and unfailed cells are very similar. In this case it is not expected to have a positive o negative coefficient that can be in accord with the centroids of failed or unfailed cells.

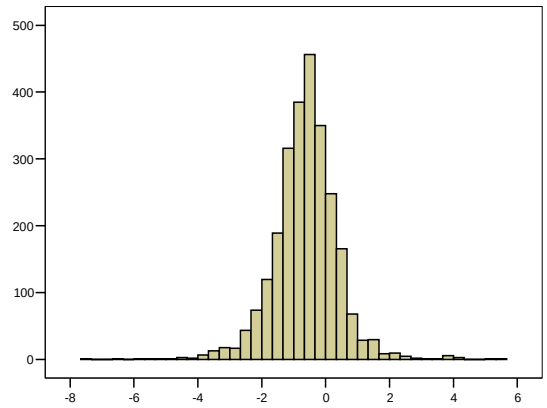
According to this, failures are expected on slopes with gentle watershed, where the ability of groundwater flow to reach the potential landslide site is relevant.
The importance of the *SPI* is particularly high, demonstrating the relevance of the erosive power of water flow on slope instability.
Regarding the soil type (*ST*), the lithology showing the greatest index of landslide density is the most prone to landsliding and thus the most contributing to the slope instability. Thus it is expected that first-time failure earthflows occur mostly in flysh deposits.

Failed and unfailed slope populations are characterized respectively by centroid values of C_f = 0.799 and C_{un} = -0.640. The distance between centroids shows that separation between the two groups by the discriminant function has been successfully achieved.

Failed



Unfailed



- **ROC curves**

In order to assess the performance of the deterministic analysis developed, the ROC curve (Receiver Operating Characteristic) was used in this study.
The ROC curve is a useful method of representing the quality of deterministic and probabilistic detection and forecast systems (Swets, 1988). The curve is a plot of the probability of having true positive identified landslides versus the probability of having false positive identified landslides.
The ROC curve is a measure of performance of a predictive rule and it is used when decision making is done with uncertainty. The quality of a forecast system is characterized by the area under the ROC curve (AUC), which describes the system's ability to anticipate correctly the occurrence or non-occurrence of predefined events.
The area under the curve is a useful statistic summary of the accuracy of the analysis and it varies from 0.5 to 1 and the most ideal model shows a curve that has the largest AUC.
The information contained in the curve is described below:

- If the model does not predict the occurrence of the landslide any better than a random approach, the AUC would equal 0.5. In this case the probability of having false positive identified landslides and the probability of having true positive identified landslides are equal and the curve is diagonal (0,0) (1,1).
- A ROC curve of 1 shows a perfect prediction. In this case, for any value assumed by the false positive %, the true positive % is always 1. Therefore, the curve has the point (0,1) and the point (1,1).
- Usually the curves are contained between the two ideal curves described above.

In our study, the ROC curve has been obtained either for the training area (cells used in the disciminant analysis) and for the target area (cells not used in the discriminant analysis), considering only the failed cells.
As already said (cfr. 3.3.3), the strategy adopted in order to obtain the landslide control set consists in carrying out the analysis in a part of the study area (training area) and testing it in another part (target area).
The two curves obtained are reported in figure 47. It can be observed that in both the cases the area under the curve (AUC) is nearly equal to 0.9, that represents a good result. It means that the analysis developed has good capacities of predict the occurrence of landslides.

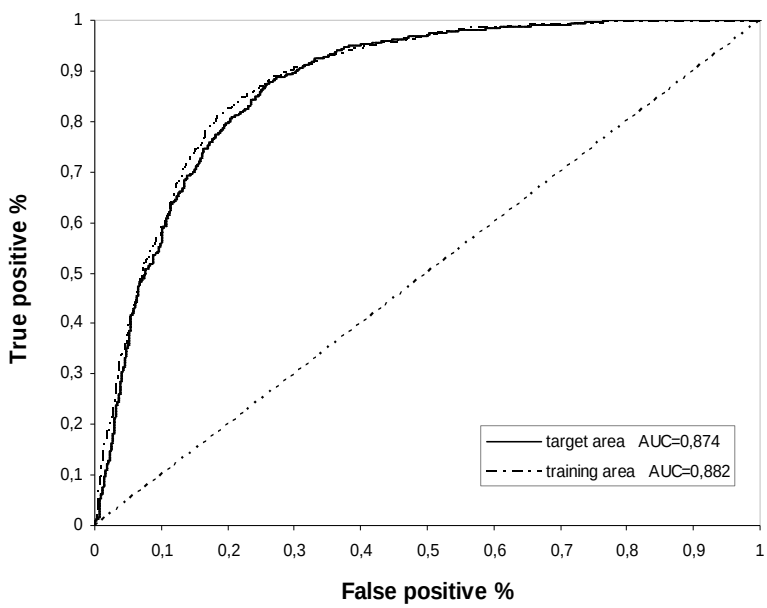


Figure 47. ROC-curves

Testing of the discriminant function

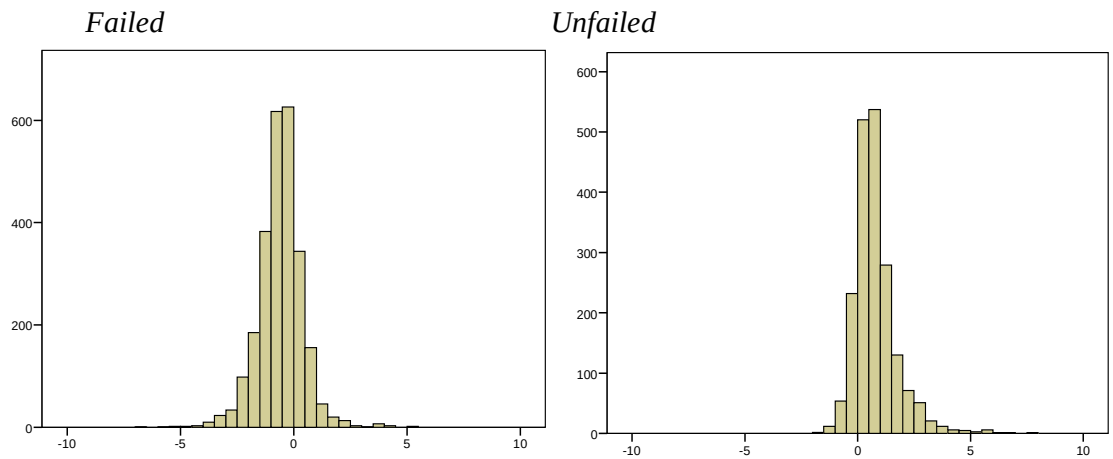
In order to confirm the obtained results, a new sample of unfailed cells has been compared with the populations of failed cells (cells representative of earthflows).

From these two populations, similar steps as the previous ones have been developed.

The results of the discriminant function obtained, with a Wilks λ of 0.665, are summarized below.

Table XLIX. Discriminant function result

Variables	Function coefficients		Correctly classified (%)
	Standard	Unstandard	
PL	0.475	0.611	General, 80.0
logLONG	-1.978	-5.062	Partial (unfailed-failed), 79.5 – 80.6
PENDM	-0.847	-0.119	
logSPI	2.487	3.998	Group centroids
ST	0.319	0.024	Unfailed slopes, -0.620
Constant		-1.451	Failed slopes, 0.812



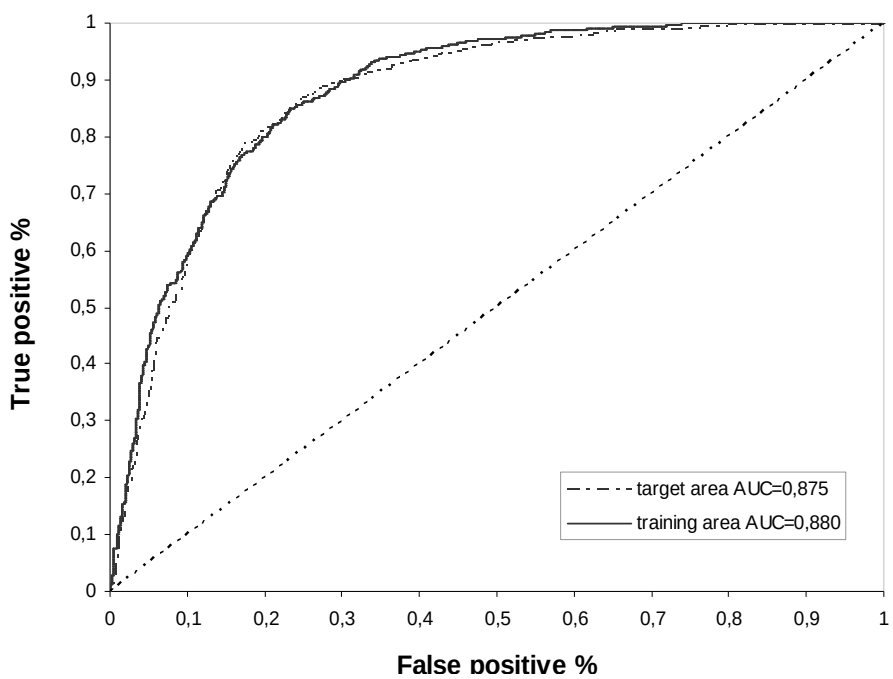


Figure 48. ROC-curves

Definition of the susceptibility levels and creation of the susceptibility map

Finally a landslide-susceptibility map was prepared using the discriminant function *DF2* (cfr. Appendix 4). The discriminant score is described as the combination defined in the (14).
Applying the coefficients obtained in *DF2*, the disciminant scores are described by the sequent function:

$$DS = 0.814 + 0.696 \cdot PL - 5.704 \cdot \log LONG - 0.116 \cdot PENDM + 4.043 \cdot \log SPI + 0.024 \cdot ST \tag{34}$$

The susceptibility map has been obtained using this function in ArcMap. The resulting discriminant scores range from -11 to 7. Basing on the methodology adopted to divide in classes the discriminant scores, different map can be created. In order to best interpret the susceptibility map, discriminant scores were divided into several ranges, based on the standard deviation of the distribution obtained, as described in figure 49.

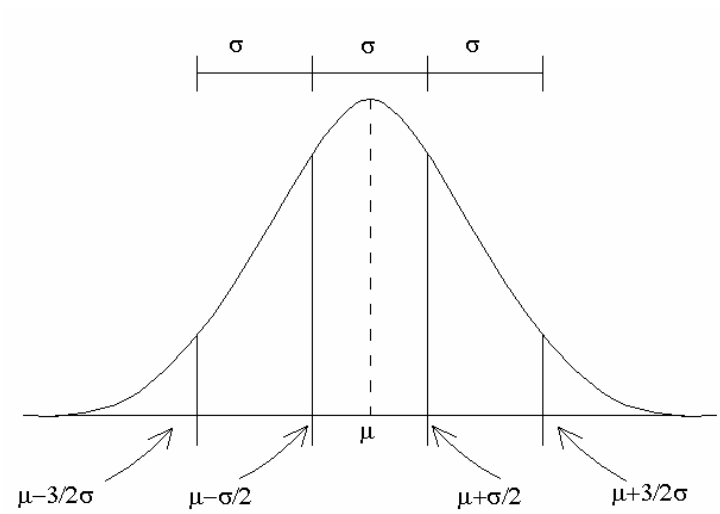


Figure 49. Method of dividing in ranges

Five susceptibility classes were established according to the ranges in table L.

Table L. Susceptibility classes

Classes	Susceptibility	DF value
I	Very low	DF < -3.08
II	Low	-3.08 ≤ DF < -1.62
III	Moderate	-1.62 ≤ DF < -0.16
IV	High	-0.16 ≤ DF < 1.3
V	Very high	DF ≥ 1.3

It may be expected that slope failures would appear in cells having higher discriminant scores (susceptibility levels from III to V).
An extract of the obtained map is reported below.

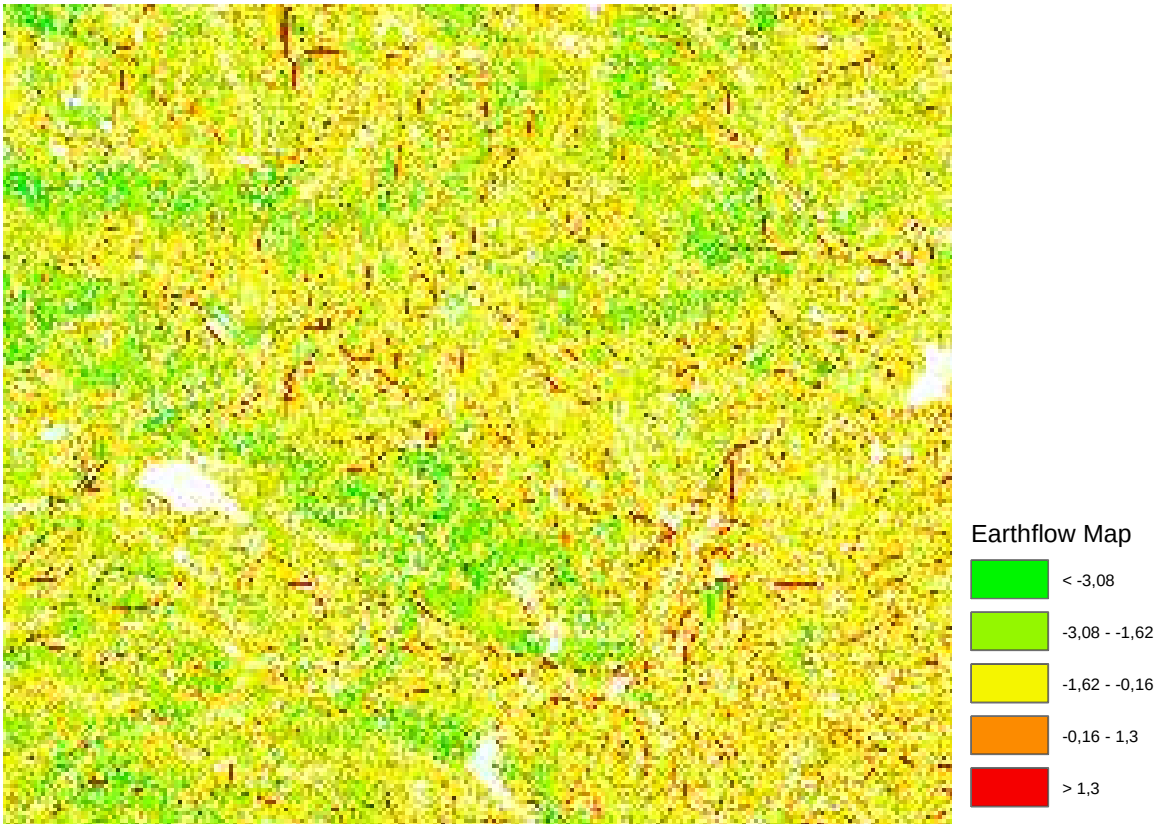


Figure 50. Extract of the map obtained from the discriminant function

Validation of the susceptibility map

In order to verify the susceptibility map, the index of relative landslide density R defined by the (24) has been used. Table LI shows index R for each susceptibility class.

Table LI. Index of relative landslide density for discriminant function DF2 – in number

Susceptibility levels	Landslides (n=2136)	Cells (N=13524686)	R
I (very low)	5	818794	0,51
II (low)	7	1916469	0,30
III (moderate)	302	7207563	3,50
IV (high)	1401	2955508	39,57
V (very high)	421	626352	56,11

R – index defined by (24)
n – number of landslides in a certain susceptibility class
N – number of cells within a certain susceptibility class

As expected, the percentage of presence of landslides in each susceptibility level increases from 'very low' level to 'high' level.

However, as we are dealing with large landslides, a similar calculation has been proposed considering the area of the scarp as the parameter influencing the index R.

In this case, the index takes the following form:

$$R = \frac{a_i / A_i}{\sum (a_i / A_i)} \cdot 100 \qquad (35)$$

where a_i is the area of the scarps observed within a susceptibility class and A_i is the area

occupied by the cells of this class (calculated as the product between the number of cells N_i and the area of the single cell 10x10. Table LII shows the relative landslide density for each susceptibility class.

Table LII. Index of relative landslide density for discriminant function DF2 – in area

Susceptibility levels	Landslides (a=14018004,13)	Cells (A=1352468600)	R
I (very low)	19097	81879400	0,26
II (low)	71384	191646900	0,42
III (moderate)	1512616	720756300	2,35
IV (high)	8857981	295550800	33,50
V (very high)	3556926	62635200	63,48

R – index defined by (35)
a – area occupied by scarps in a certain susceptibility class [m^2]
A – area occupied by the cells of a certain susceptibility class [m^2]

The distribution of the index among the different susceptibility classes seems to adjust better the real situation. We may conclude that the distribution of slope failures observed in these classes indicates that susceptibility levels are consistent.

In addition, effectiveness of the landslide susceptibility map was evaluated relating the areal distributions of observed landslides with susceptibility intervals of 10%, as is usually done in logistic regression.
Generally, logistic regression involves fitting the dependent variable using an equation in the following form:

$$Y = \ln[p/(1-p)] = C_0 + C_2 X_2 + C_3 X_3 + ... + C_n X_n \qquad (36)$$

In the (31) p is the probability that the dependent variable Y is 1 and it is already known from the discriminant analysis.

In order to calculate the parameter p from the discriminant score $DS=Y$, the previous formula has to be inverted as follow:

$$p = \frac{\exp(Y)}{1+\exp(Y)} \qquad (37)$$

While the parameter Y obtained from the discriminant analysis can assume all the possible values, the probability p assumes values in the range [0-1]. Hence, it is possible evaluate the susceptibility in intervals of 0.1 (corresponding to 10%).
In order to verify the effectiveness of the map, two decision rules, introduced by Can et al. (2005), were considered. These decisions rules are:

1. On the map, most of the actual landslides should have to be located in the cells included in high susceptibility classes;
2. On the map, these high susceptibility classes should have to cover small area as possible.

For the purpose, two different curves were drawn (Figure 51). The first (curve a) is landslide susceptibility, subdivided in intervals of 10%, versus the observed cumulative percentage of the scarps' area. In an effective landslide susceptibility map, the actual landslide should be included by the areas having high susceptibility values.

Considering the cut-off value of 0.5, corresponding to the value *Y=0*, approximately 80% of the observed landslides locate in the class of *'susceptible'* (Figure 51). For this reason, it is possible to say that the curve "a" satisfies the rule 1.
The second (curve b) is landslide susceptibility versus cumulative percentage of the overall study area. This curve represents the areal distribution of the susceptibility values in the region. To satisfy the rule 2, the cumulative area obtained from this curve should be as small as possible. Considering the cut-off value of 0.5, the cumulative area of the susceptibility values is obtained as 20%. This means that the curve "b" also satisfies the rule 2.

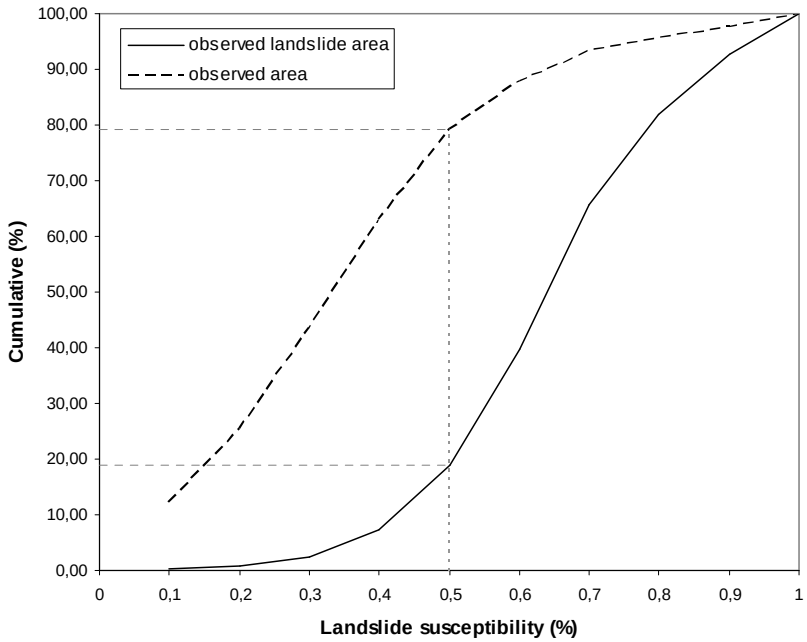


Figure 51. ROC-curves

In the table LIII the data necessary to create the curve a and the curve b are reported. For simplicity, the ten classes from 0.1 to 1 are summarized in 5 classes, so to allow a comparison with the 5 classes of susceptibility defined by the discriminant scores DS.

Table LIII. Observed landslide area and observed area data

	Observed landside area			***Observed area***		
Susceptibility levels	**Partial**	**Cumulative**	**% Cumulative**	**Partial**	**Cumulative**	**% Cumulative**
0-0,2	15	15	0,70	3435467	3435467	25,40
0,2-0,4	139	154	7,21	5095739	8531206	63,08
0,4-0,6	693	847	39,65	3363346	11894552	87,95
0,6-0,8	905	1752	82,02	1055089	12949641	95,75
0,8-1	384	2136	100,00	575045	13524686	100,00

The same calculation has been developed considering the area of the scarps rather than the number. In this case the curve is the landslide susceptibility (calculated as *p* [0-1]) versus observed cumulative percentage of the area of cells that include landslides.

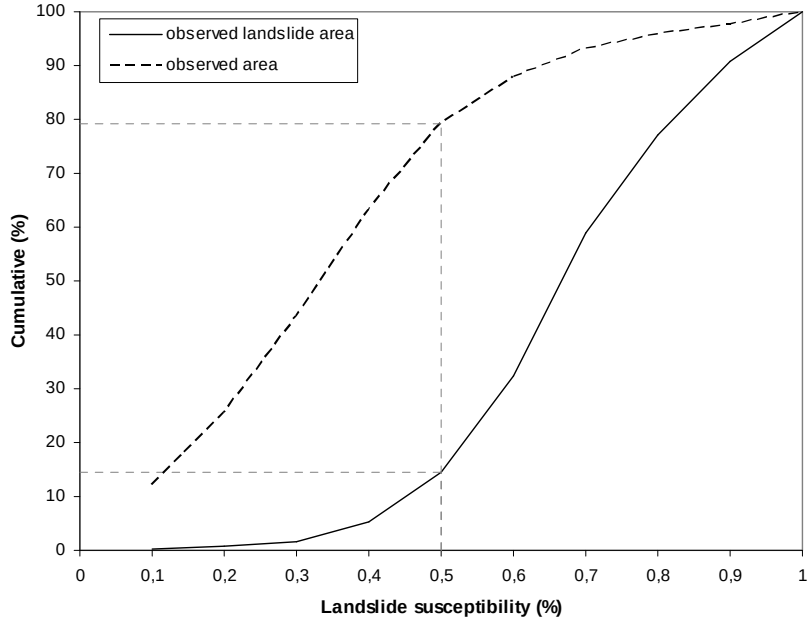


Figure 52. Graph showing the effectiveness of the produced landslide susceptibility map – earthflows

Table LIV. Landslide observed area data

	Observed landside area			***Observed area***		
Susceptibility levels	**Partial**	**Cumulative**	**% Cumulative**	**Partial**	**Cumulative**	**% Cumulative**
0-0,2	99963	99963	0,71	3435467	3435467	25,40
0,2-0,4	629345	729308	5,20	5095739	8531206	63,08
0,4-0,6	3819234	4548542	32,45	3363346	11894552	87,95
0,6-0,8	6242773	10791315	76,98	1055089	12949641	95,75
0,8-1	3226689	14018004	100,00	575045	13524686	100,00

Also in this case, the curve “a” and “b” also satisfy the rule 1 and 2.

SLIDES

Parameters description

The first step is creating a general description table, in order to give informations about the mean value assumed by all the selected parameters in the two groups of failed and unfailed cells. For the same reasons explained in the case of earthflows, we have decided to cut the right and left tails of the frequency distribution of the variable β, reducing its range values to [6-40°]. The table containing the main statistical information about the samples is reported below.

Table LV. Statistics

UNFAILED	H	β	α'	CURVAR	PF	PL	PENDM	ST	A	LONG	SPI	RUGOS
N	1807	1807	1807	1807	1807	1807	1807	1807	1807	1807	1807	1807
Mean	518,64	14,84	97,31	-0,44	0,15	-0,28	14,31	28,06	9082,89	126,16	2303,97	1,05
Std. Error of Mean	6,62	0,18	1,23	0,04	0,024	0,023	0,18	0,32	2124,62	5,58	537,53	0,001
Median	471,00	11,97	98,13	0,00	0,00	0,00	11,60	40,84	600,00	64,40	143,29	1,02
Mode	202,00	8,53	135,00	0,00	0,00	0,00	6,723(a)	40,84	100,00	10,11(a)	11,79(a)	1,01
Std. Deviation	281,71	7,92	52,29	1,76	1,04	0,98	7,98	13,74	90315,51	237,49	22850,04	0,06
Variance	79366,14	62,74	2734,59	3,10	1,08	0,96	63,69	189,03	8156891473,75	56402,33	522124389,36	0,004
Skewness	0,727	1,161	-0,196	-0,79	0,223	-2,230	1,206	-0,319	17,649	8,294	19,685	2,203
Std. Error of Skewness	0,058	0,058	0,058	0,058	0,058	0,058	0,058	0,058	0,058	0,058	0,058	0,058
Kurtosis	0,061	0,535	-1,075	6,230	6,940	17,209	0,892	-1,665	349,740	101,020	460,122	5,068
Std. Error of Kurtosis	0,115	0,115	0,115	0,115	0,115	0,115	0,115	0,115	0,115	0,115	0,115	0,115
Range	1618,00	33,76	180,00	23,00	16,049	15,365	47,082	34,93	2187200,00	4272,59	644674,99	0,40
Minimum	44,00	6,05	0,00	-12,00	-8,049	-11,615	2,281	5,91	100,00	10,04	7,90	1,01
Maximum	1662,00	39,82	180,00	11,00	8,000	3,750	49,364	40,84	2187300,00	4282,62	644682,89	1,40

FAILED	H	β	α'	CURVAR	PF	PL	PENDM	ST	A	LONG	SPI	RUGOS
N	1536	1536	1536	1536	1536	1536	1536	1512	1536	1536	1536	1536
Mean	495,82	13,38	107,85	0,047	-0,021	0,02	13,09	34,16	9094,93	95,42	1924,58	1,04
Std. Error of Mean	5,761	0,140	1,055	0,009	0,006	0,004	0,13	0,27	1431,07	2,12	252,32	0,0009
Median	451,57	12,02	107,10	0,037	-0,018	0,016	11,84	40,84	1383,69	74,68	296,43	1,02
Mode	275,00(a)	10,25450(a)	93,540(a)	0,00	-0,085(a)	0,00	10,158(a)	40,84	300,00(a)	102,46(a)	5,14(a)	1,02
Std. Deviation	225,80	5,49	41,37	0,35	0,24	0,18	5,33	10,71	56086,54	83,15	9888,99	0,038
Variance	50989,225	30,229	1712,146	0,128	0,061	0,035	28,469	114,908	3145701029,67	6915,22	97792142,92	0,001
Skewness	0,611	1,330	-0,089	0,994	-1,660	-0,070	1,367	-1,175	16,452	5,480	11,484	2,759
Std. Error of Skewness	0,062	0,062	0,062	0,062	,062	0,062	0,062	0,063	0,062	0,062	0,062	0,062
Kurtosis	0,027	1,946	-0,980	10,457	19,967	5,441	2,485	-0,202	358,189	52,815	159,214	9,871
Std. Error of Kurtosis	0,125	0,125	0,125	0,125	0,125	0,125	0,125	0,126	0,125	0,125	0,125	0,125
Range	1339,049	32,253	175,674	5,403	3,906	2,385	36,143	34,93	1492766,67	1207,45	188170,68	0,30
Minimum	76,055	6,009	4,227	-2,136	-2,780	-1,308	4,839	5,91	33,33	3,37	5,14	1,01
Maximum	1415,104	38,26260	179,901	3,267	1,126	1,077	40,982	40,84	1492800,00	1210,82	188175,81	1,31

a Different modas exist. The minor value is shown.

Testing for normal distribution

In order to work with normal distributions of the variables, the normalization test has been developed. To observe if each variable has a normal distribution, the histogram of each one has been previously created (cfr. Appendix 2), and then the test of Kolmogorov-Smirnov has been applied. From the histograms, it is possible to observe that some variables have a normal distribution and others no. In order to confirm all that has been observed from the histograms, a KS test has been developed. In the table LVI the values of the KS test and the significance level have been reported. The greatest values of KS are highlighted in orange.

Table LVI. K-S test results – without transformation

Variable	Z de Kolmogorov-Smirnov
H	4,492
β [°]	7,837
A	26,186
α' [°]	4,609
CURVAR	10,882
PF	10,272
PL	12,949
LONG	16,702
PENDM [°]	7,215
RUGOS	13,013
SPI	26,206
ST	21,736

As happened for earthflows, also in this case of slides the log-transformation is necessary for the variable *A*, *LONG*, *RUGOS* and *SPI*. The results of the transformation in term of KS values and histrograms are reported below. A lower value of the KS test is observed in all the transformed variables.

Table LVII. K-S test results – with transformation

	Before transformation		After transformation
Variable	Z de Kolmogorov-Smirnov	Variable	Z de Kolmogorov-Smirnov
H	4,492		
β [°]	7,837		
A	26,186	log A	3,530
α' [°]	4,609		
CURVAR	10,882		
PF	10,272		
PL	12,949		
LONG	16,702	log LONG	1,859
PENDM [°]	7,215		
RUGOS	13,013	log RUGOS	12,434
SPI	26,206	log SPI	3,213
ST	21,736		

For each variable two graphs have been developed (cfr. Appendix 5): the first refers to the population of unfailed cells, the second to the group of failed cells.

Also in this case, it can be observed that the frequency distribution of all the variables in the population of failed cells (cells representing slides) is almost normal, while it doesn't adjust well to a normal distribution in the group of unfailed cells. However, it does not represent a problem for the discriminant analysis, thanks to its robustness.

Selection of independent variables

Having verified the normal distribution of the variables, the next step consists in verifying the indipendence of the variables. As already said, dependent variables must be removed, searching for possible correlation.

A bivariate analysis of correlation will be first done, in order to search for possible relationship between the variables.

Then the principal component analysis (PCA) will be applied, in order to view the membership of each variable to the different factors.

- **Bivariate Pearson correlation**

The table with the coefficient of Pearson is reported below. The variables showing high correlation have the coefficient colored in yellow.

Regarding the correlations, the variables with great Pearson coefficient are reported below. For the meaning assigned to the correlated parameters cfr. the earthflows' case.

- β, *PENDM* and *RUGOS*

- *logA*, *logLONG* and *logSPI*

- *CURVAR*, *PF* and *PL*

Table LVIII. Pearson coefficients of correlation matrix

	H	β [°]	log A	α' [°]	CURVAR	PF	PL	log LONG	PENDM [°]	log RUGOS	log SPI	ST
H	1											
β [°]	0,199(**)	1										
log A	-0,025	-0,073(**)	1									
α [°]	0,079(**)	-0,091(**)	0,043(*)	1								
CURVAR	-0,033	-0,057(**)	-0,175(**)	0,026	1							
PF	0,000	0,026	0,058(**)	-0,031	-0,876(**)	1						
PL	-0,059(**)	-0,075(**)	-0,252(**)	0,013	0,862(**)	-0,511(**)	1					
log LONG	-0,025	0,017	0,909(**)	0,008	-0,193(**)	0,064(**)	-0,276(**)	1				
PENDM [°]	0,196(**)	0,889(**)	-0,016	-0,100(**)	-0,119(**)	0,116(**)	-0,090(**)	0,074(**)	1			
log RUGOS	0,196(**)	0,965(**)	-0,056(**)	-0,087(**)	-0,101(**)	0,052(**)	-0,125(**)	0,021	0,859(**)	1		
log SPI	0,031	0,195(**)	0,952(**)	0,013	-0,193(**)	0,084(**)	-0,256(**)	0,889(**)	0,281(**)	0,190(**)	1	
ST	0,024	-0,352(**)	0,061(**)	0,161(**)	0,026	0,011	0,059(**)	0,000	-0,365(**)	-0,330(**)	-0,046(**)	1

(**) The correlation is significative at 0.01 level (bilateral).
(*) The correlation is significative at 0.05 level (bilateral).

- **Factorial analysis**

The principal component analysis has been developed with the 12 variables considered. In this case the first 4 factors explain nearly the 80 per cent of the variance (cfr. Table LIX).

Table LIX. Total variance explained in the Principal Component Analysis
(Results of the sum of the square saturation of the rotation)

Factor	Value	% of the variance	% accumulated
1	3,098	25,816	25,816
2	2,864	23,866	49,683
3	2,515	20,960	70,643
4	1,157	9,644	80,287

From the results obtained in the table LX it is clear that the total variance of each variable is greater then the 45% (with the variable *ST*), being the variables *CURVAR* and *logA* and *logSPI* the best representative, with more than 97% of the variance, followed by *β*, *logLONG* and *logRUGOS*, with more than 90%.

Table LX. Communalities or total variance of each variable
(Extraction method: Principal component analysis)

Variable	Extraction	Variable	Extraction
H	0,508	PL	0,754
β [º]	0,941	log LONG	0,922
log A	0,976	PENDM [º]	0,883
α' [º]	0,527	log RUGOS	0,914
CURVAR	1,000	log SPI	0,972
PF	0,787	ST	0,452

The rotated matrix with the factors is presented in the table LXI, where the greater loads are colored in yellow.

Table LXI. Varimax rotated component matrix (Extraction method: Principal component analysis)

	Factor			
	1	2	3	4
H	0,290	-0,040	-0,033	0,649
β [º]	0,970	0,016	-0,012	0,022
log A	-0,087	0,980	-0,083	0,021
α' [º]	-0,128	0,038	0,031	0,713
CURVAR	-0,046	-0,098	0,994	0,002
PF	0,017	-0,032	-0,886	-0,015
PL	-0,063	-0,209	0,840	-0,013
log LONG	0,002	0,955	-0,099	-0,019
PENDM [º]	0,933	0,078	-0,072	0,002
log RUGOS	0,954	0,019	-0,058	0,030
log SPI	0,191	0,963	-0,091	0,021
ST	-0,476	0,022	0,000	0,474

The first Factor, representing nearly the 26% of the total variance, is defined by the variables *β*, *PENDM* and *RUGOS*. The second factor, representing nearly the 24% of the total variance, is defined by the variables *logA*, *logLONG* and *logSPI*. The third factor, representing nearly the 21% of the total variance, is defined by the variables *CURVAR*, *PF* and *PL*. The fourth factor, representing nearly the 9% of the total variance, is defined by the variables *α* and *H*. The graph showing the first three factors in a rotated space is presented below.

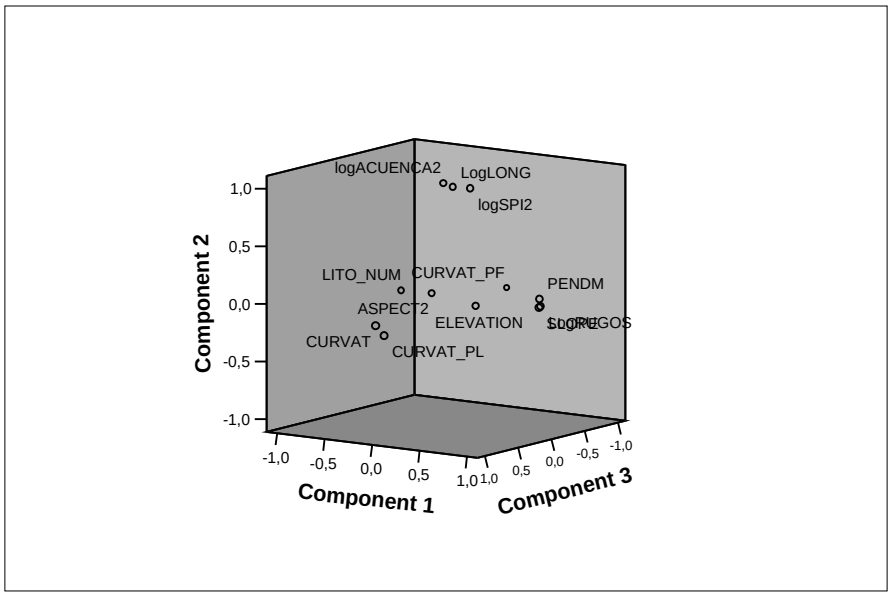


Figure 53. Component plot in a rotated space

After having identified the dependent variables, before rejecting one of these, **One-way test** and **T-test** were developed in order to understand the influence of each variable on stability. The Tables of the results of One-way test and T-test are reported below.

Table LXII. T-test and One-way test results

				One-way test		T-test	
Variable	**Activity state**	**Mean**	**Standard Deviation**	**F**	**Sig.**	**t**	**Sig.**
H	0	518,65	281,72	78,380	0,000	2,599	0,009
	1	495,82	225,81				
β [°]	0	14,84	7,92	225,994	0,000	6,269	0,000
	1	13,38	5,50				
log A	0	2,87	0,68	53,955	0,000	-17,449	0,000
	1	3,25	0,57				
α' [°]	0	97,32	52,29	127,887	0,000	-6,500	0,000
	1	107,85	41,38				
CURVAR	0	-0,44	1,76	1006,129	0,000	-11,591	0,000
	1	0,05	0,36				
PF	0	0,16	1,04	867,622	0,000	7,052	0,000
	1	-0,02	0,25				
PL	0	-0,29	0,98	760,954	0,000	-13,235	0,000
	1	0,02	0,19				
log LONG	0	1,81	0,47	474,214	0,000	-5,540	0,000
	1	1,89	0,27				
PENDM [°]	0	14,31	7,98	236,774	0,000	5,251	0,000
	1	13,09	5,33				
log RUGOS	0	0,02	0,02	174,737	0,000	5,560	0,000
	1	0,02	0,01				
log SPI	0	2,22	0,72	59,528	0,000	-16,015	0,000
	1	2,58	0,59				
ST	0	28,06	13,75	428,603	0,000	-14,345	0,000
	1	34,16	10,72				

According to all these analyses, the same considerations done for the earthflows can be developed also for the slides, in order to define the most significant independent variables to be taken as input of the discriminant analysis.

Obtaining the discriminant function

Also in this case, different discriminant analyses have been developed. The first was performed with the following input variables: *H*, β, α', *PL*, *logLONG*, *PENDM*, *logRUGOS*, *logSPI* and *ST*.

The discriminant function (*DF1*) obtained and the main statistical parameters are shown in Table LVIII. Elevation (*H*), roughness (*logRUGOS*) and slope angle (β) are the most ininfluential variables (the last caught in the analysis), while the most influential variables are the stream power index (*logSPI*) with a standardized coefficient equal to 2.372, the watershed lenght (*logLONG*) with -1.830 and the mean watershed angle (*PENDM*) with -0.605. Slope aspect (α'), plan curvature (*PL*) and soil type (*ST*), with lower weights, were also chosen.

A new discriminant function (*DF2*) was obtained using the same variables except elevation (*H*), roughness (*logRUGOS*) and slope (β). The stepwise method gives nearly equal results, as the Wilks λ increase insignificantly, while the percentage of correctly classification remains almost the same (Table LXIX). The most influential variables are the stream power index (*logSPI*) with 2.378 and the watershed length (*logLONG*) with -1.830. The other variables included in the analysis, with lower weights, were also chosen and no one of the variables was rejected.

A final analysis was performed without the variable aspect (α'), which is the last caught in the *DF2*, in order to evaluate its importance. The stepwise method gives nearly equal results, as the Wilks λ increases insignificantly, while the percentage of correctly classification remains almost the same.

The table LXIX presents the main statistical characteristics of the analyses developed. The best discriminant analysis up to now is shown in yellow.

As the variables chosen in the discriminant analysis are the same caught in the case of earthflows, it is not necessary to carry out other analyses in order to verify the importance of the variables caught in the *DF3*. Actually, it is expected that further analyses give not better results (cfr. the calculation done using the variable *CURVAR* and *PF* rather than *PL* or the calculation made using *logA* rather than *logLONG*).

Table LXIII. DF1, DF2 and DF3 discriminant function results

Discriminant Function / Variables	DF1				DF2				DF3			
	input	used*	rejected	Standardized coefficient	input	used*	rejected	Standardized coefficient	input	used*	rejected	Standardized coefficient
H	*	7		-0.099								
β	*	9		-0.500								
logA												
α'	*	6		0.102	*	6		0.094				
CURVAR												
PF												
PL	*	4		0.494	*	4		0.479	*	4		0.479
logLONG	*	2		-1.830	*	2		-1.830	*	2		-1.836
PENDM	*	3		-0.605	*	3		-0.593	*	3		-0.599
logRUGOS	*	8		0.559								
logSPI	*	1		2.372	*	1		2.378	*	1		2.387
ST	*	5		0.315	*	5		0.314	*	5		0.326
Wilks lambda λ	0.637				0.644				0.646			
% correctly classified general	82.0				81.6				81.9			
% correctly classified Unfailed	85.0				84.3				84.5			
% correctly classified Failed	78.4				78.4				78.8			
Function at group centroid - Unfailed	-0.691				-0.681				-0.678			
Function at group centroid - Failed	0.823				0.811				0.807			

used*: the number indicates the order of capture

The table LXIII presents in yellow the discriminant function considered as the best discriminating. It has been chosen because contains the smaller number of variables which explain the behaviour of the selected cells regarding the slope stability/instability.
The main statistical parameters of the selected function (*DF3*) are reported below:

Table LXIV. DF3 discriminant function results

Variables	Function coefficients		Correctly classified (%)
	Standard	Unstandard	
PL	0,479	0,688	General, 81.9
logLONG	-1,836	-4,775	Partial (unfailed-failed), 84.5-78.8
PENDM	-0,599	-0,088	
logSPI	2,387	3,648	Group centroids
ST	0,326	0,026	Unfailed slopes, -0.678
Constant		0,638	Failed slopes, 0.807

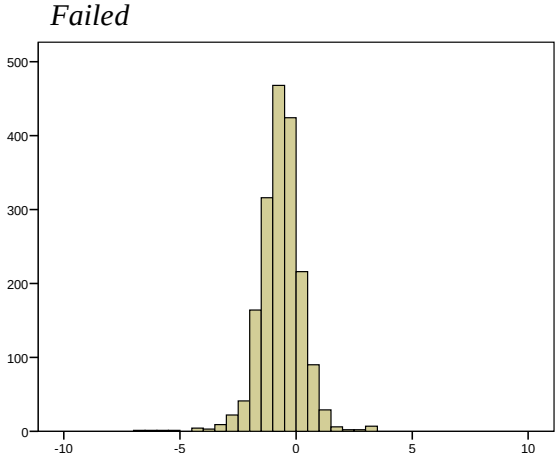
Also in this case, positive discriminant coefficients are associated with failed slopes and negative ones with unfailed slopes.
Therefore, high values of *PL*, *logSPI* and *ST* increase discriminant scores and, consequently, slope instability. High values of *logLONG* and *PENDM*, with a negative coefficient, increase stability. This result is in accord with the mean value assumed by the selected variables in the two populations of unfailed and failed cells, as reported in the descriptive table below.

Table LXV. Mean values of the selected variables

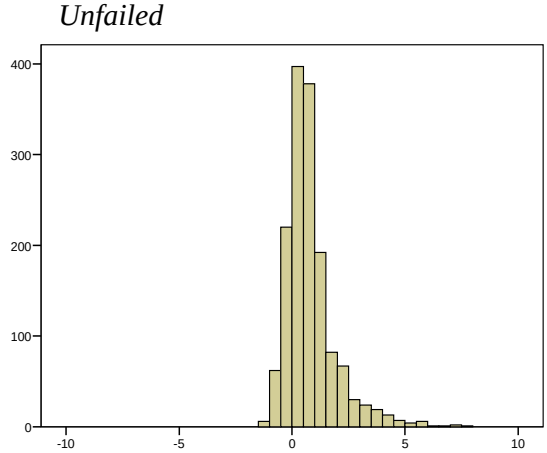
	PL	PENDM	ST	logLONG	logSPI
Activity=0	-0,29	14,31	28,06	1,81	2,22
Activity=1	0,02	13,09	34,16	1,89	2,58

The same observations made for earthflows can be developed for slides. In all the cases the coefficients obtained in the discriminant analysis are in accord with the centroids of unfailed or failed cells.
Unfailed and inunfailedslope populations are characterized respectively by centroid values of C_f = 0.807 and C_{un} = -0.678. The distance between centroids shows that separation between the two groups by the discriminant function has been successfully achieved.

Failed



Unfailed



- **ROC curves**

Also in this case, the ROC curve is used in order to assess the performance of the deterministic analysis developed.
Two ROC curves have been obtained either for the training area (cells used in the disciminant analysis) and for the target area (cells not used in the discriminant analysis), considering only the failed cells.

The ROC curves obtained is reported below. It can be observed that in both the cases the area under the curve (AUC) is nearly equal to 0.9, that represents a good result. It means that the analysis developed has good capacities of predict the occurrence of landslides.

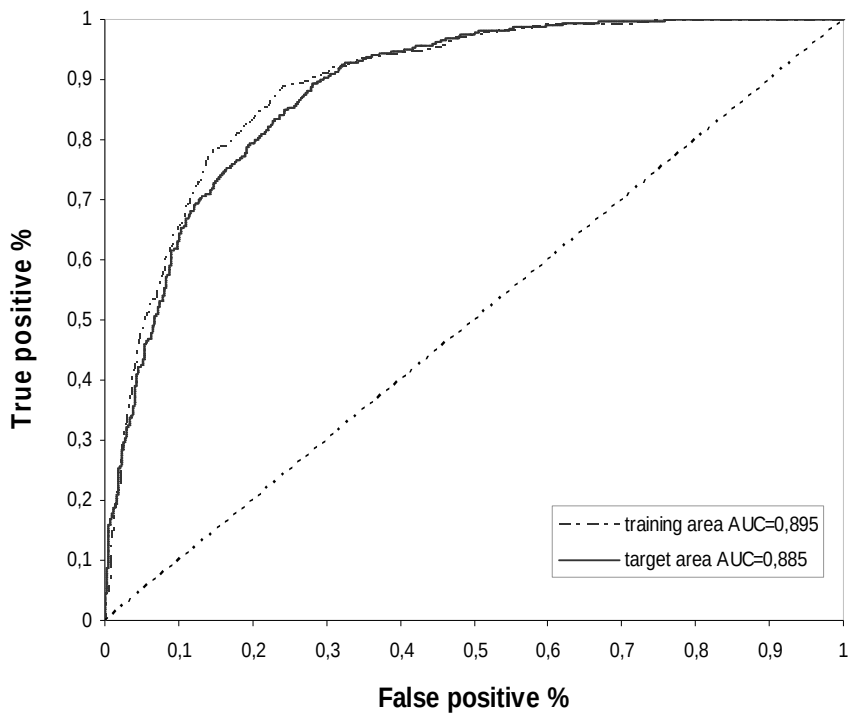


Figure 54. ROC - curves

Testing of the discriminant function
In order to confirm the obtained results, a new sample of unfailed cells has been compared with the populations of failed cells (cells representative of slides).
From these two populations, similar steps as the previous ones have been developed.
The results of the discriminant function obtained, with a Wilks λ of 0.651, have been summarized below.

Table LXVI. Discriminant function result

Variables	Function coefficients		Correctly classified (%)
	Standard	Unstandard	
PL	0,473	0,641	General, 80.6
logLONG	-1,837	-4,708	Partial (unfailed-failed), 83.7-77.0
PENDM	-0,582	-0,084	
logSPI	2,401	3,672	Group centroids
ST	0,336	0,027	Unfailed slopes, -0.676
Constant		0,361	Failed slopes, 0.792

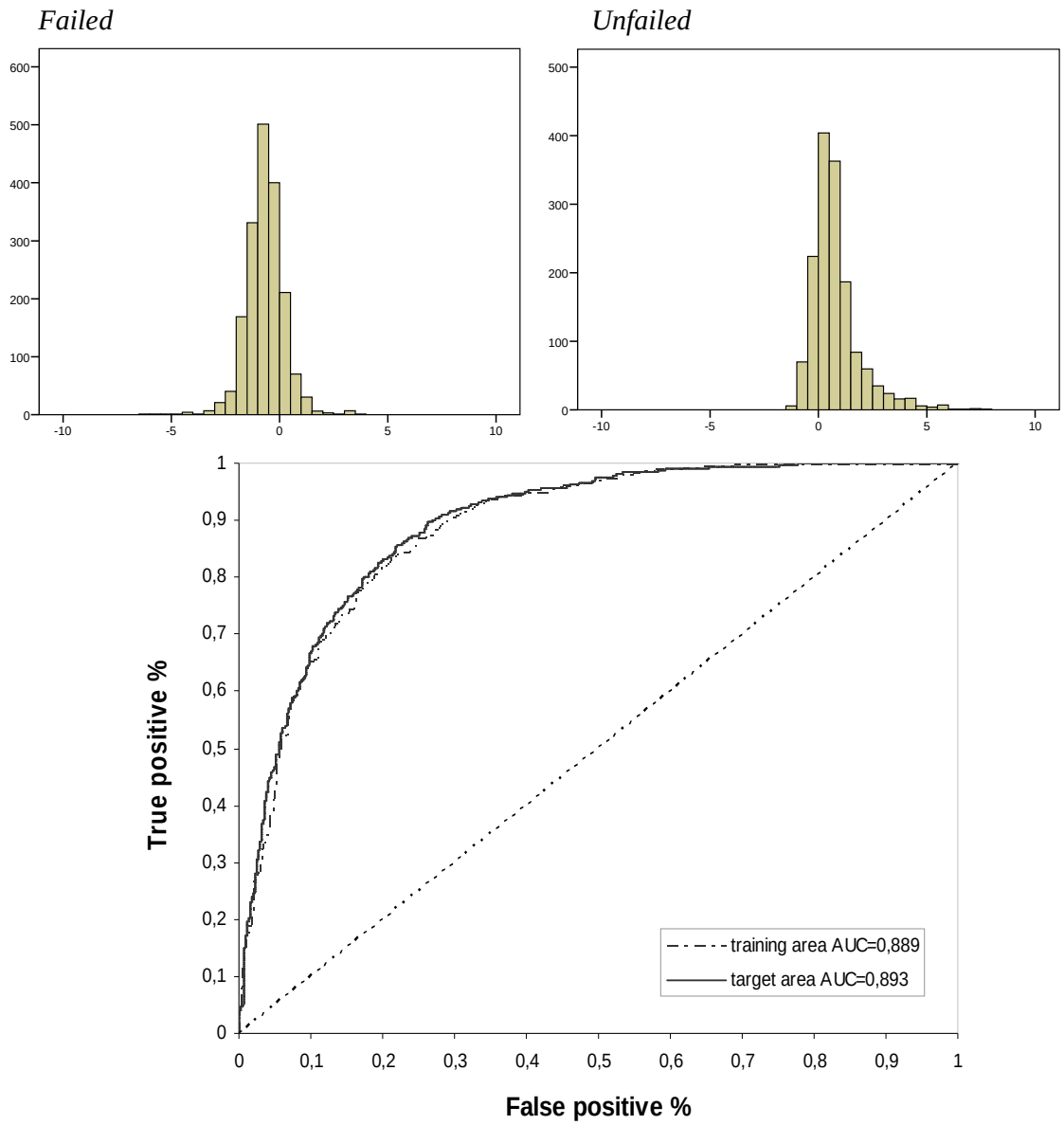


Figure 55. ROC - curves

<u>*Definition of the susceptibility levels and creation of the susceptibility map*</u>

Finally a landslide-susceptibility map was prepared using the discriminant function *DF3* (Cfr. Appendix 4). In this case the disciminant scores are described by the sequent function:

$$DS = 0.638 + 0.688 \cdot PL - 4.775 \cdot \log LONG - 0.088 \cdot PENDM + 3.648 \cdot \log SPI + 0.026 \cdot ST \tag{38}$$

The susceptibility map has been obtained using this function in ArcMap. The resulting discriminant scores range from -6 to 8. In order to best interpret the susceptibility map, discriminant scores were divided into several ranges, based on the variance of the distribution obtained (cfr. Figure 45).

Discriminant scores were divided into five ranges, based on the percentage of unfailed slopes with respect to the whole slope population. The susceptibility classes were established according to these ranges in the table LXVII.

Table LXVII. Susceptibility classes

Classes	Susceptibility	DF value
I	Very low	DF < -3.23
II	Low	-3.23 ≤ DF < -1.83
III	Moderate	-1.83 ≤ DF < -0.43
IV	High	-0.43 ≤ DF < 0.97
V	Very high	DF ≥ 0.97

It may be expected that slope failures would appear in cells having higher discriminant scores (susceptibility levels III to V).

Validation of the susceptibility map

In order to verify the susceptibility map, the index of relative landslide density R has been used. Tables LXVIII and LXIX show index R for each susceptibility class, calculated respectively in number and in area.

Table LXVIII. Index of relative landslide density for discriminant function DF3 – in number

Susceptibility levels	Landslides (n=1581)	Cells (N=13524686)	R
I (very low)	2	1022042	0,18
II (low)	4	1859674	0,20
III (moderate)	122	6676204	1,68
IV (high)	994	3369892	27,15
V (very high)	459	596875	70,79

R – index defined by (24)
n – number of landslides in a certain susceptibility class
N – number of cells within a certain susceptibility class

Table LXIX. Index of relative landslide density for discriminant function DF3 – in area

Susceptibility levels	Landslides (a=11665260,64)	Cells (A=1352468600)	R
I (very low)	13287	102204200	0,15
II (low)	19760	185967400	0,12
III (moderate)	651293,662	667620400	1,12
IV (high)	7127784,477	336989200	24,33
V (very high)	3853135,91	59687500	74,27

R – index defined by (35)
a – area occupied by scarps in a certain susceptibility class [m^2]
A – area occupied by the cells of a certain susceptibility class [m^2]

We may conclude that the distribution of slope failures observed in these classes indicates that susceptibility levels are consistent

Also in this case the effectiveness of the landslide susceptibility map was evaluated relating the areal distributions of observed landslides (curve “a”) and observed area (curve “b”) with susceptibility intervals of 10% (Figures 56). Considering the cut-off value of 0.5, corresponding to the value *Y=0*, approximately 78% of the observed landslides locate in the class of *‘susceptible’* and the cumulative area of the susceptibility values is obtained as 20% (Figure 56).

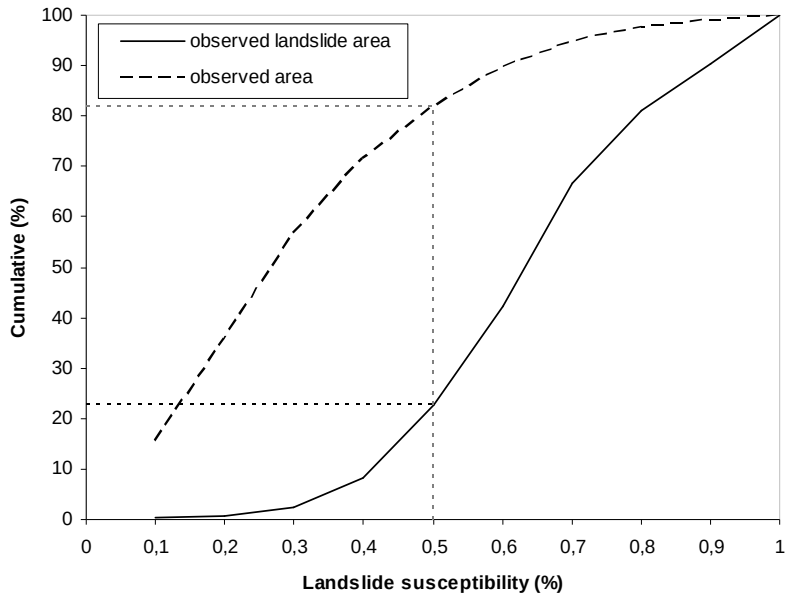


Figure 56. Graph showing the effectiveness of the produced landslide susceptibility map – slides

Table LXX. Landslide observed area and observed area data

	Landslide observed area			***Observed area***		
Susceptibility levels	**Partial**	**Cumulative**	**% Cumulative**	**Partial**	**Cumulative**	**% Cumulative**
0-0,2	11	11	0,65	4828349	4828349	35,70
0,2-0,4	131	142	8,42	4823052	9651401	71,36
0,4-0,6	573	715	42,38	2481437	12132838	89,71
0,6-0,8	655	1370	81,21	1066976	13199814	97,60
0,8-1	317	1687	100,00	324872	13524686	100,00

The same calculation has been developed considering the area of the scarps rather than the number.

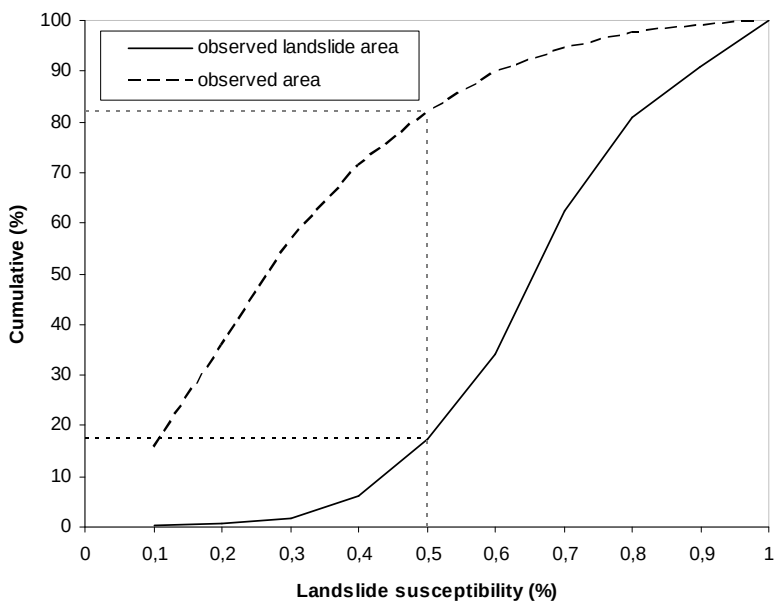


Figure 57. Graph showing the effectiveness of the produced landslide susceptibility map – slides

Table LXXI. Landslide observed area data

	Landslide observed area			*Observed area*		
Susceptibility levels	**Partial**	**Cumulative**	**% Cumulative**	**Partial**	**Cumulative**	**% Cumulative**
0-0,2	70250	70250	0,60	4828349	4828349	35,70
0,2-0,4	654400	724650	6,21	4823052	9651401	71,36
0,4-0,6	3269443	3994093	34,24	2481437	12132838	89,71
0,6-0,8	5447998	9442091	80,94	1066976	13199814	97,60
0,8-1	2223170	11665261	100,00	324872	13524686	100,00

Comparing earthflows and slides

The analyses developed for cells showing the presence of earthflows' scarps and cells showing the presence of slides' scarps give nearly the same results. The discriminant functions obtained have caught the same variables with the same significance and nearly the same coefficients. It means that a clear distinction between the two typologies of movements considered is not possible with the parameters taken into account.

This result is also confirmed applying the discriminant function obtained for earthflows to the case of slides. The index of relative index R, calculated respectively in number and in area, is shown below.

Table LXXII. Index of relative landslide density for discriminant function DF3 – in number

Susceptibility levels	**Landslides (n=1581)**	**Cells (N=13524686)**	**R**
I (very low)	2	818794	0,25
II (low)	5	1916469	0,26
III (moderate)	252	7207563	3,55
IV (high)	926	2955508	31,79
V (very high)	396	626352	64,15

R – index defined by (24)
n – number of landslides in a certain susceptibility class
N – number of cells within a certain susceptibility class

Table LXXIII. Index of relative landslide density for discriminant function DF3 – in area

Susceptibility levels	**Landslides (a=11665260,64)**	**Cells (A=1352468600)**	**R**
I (very low)	13287	81879400	0,21
II (low)	30113	191646900	0,20
III (moderate)	1408250,866	720756300	2,54
IV (high)	7015460,555	295550800	30,80
V (very high)	3198149,00	62635200	66,25

R – index defined by (30)
a – area occupied by scarps in a certain susceptibility class [m^2]
A – area occupied by the cells of a certain susceptibility class [m^2]

These tables highlight that the discriminant function obtained for the earthflows is effective also for slides, as the index R calculated in area and in number give nearly the same results.
It underlines the idea that earthflows and slides can not be distinguished using the parameters taken into account.

The analyses developed up to now have considered earthflows and slides separately, belonging to two different samples to compare with the population of unfailed cells.
A more correct procedure should consider a new sample of movements, without distinction between earthflows and slides. This sample should be analyzed again following all the steps described in the statistical treatment, in order to obtain a unique discriminant function and a unique landslide susceptibility map. This new map should describe the area more prone to develop earthflows or slides.

COMPLEX PHENOMENA: SLIDE-EARTHFLOWS

As the analyses developed considering separately the two categories of earthflows and slides have demonstrated that consistent differences between the two typologies of movements can not be emphasized using the variables taken into account, it is expected that further analyses considering the complex phenomena give similar results. This typology of movement can be thought as a combination of the two categories already analyzed: the movement begins as a rotational slide in the upper part of the slope and then evolves as a earthflow. Hence, the multivariate statistical analysis has not been developed for this type of movements.

5.3.4 Applicability of the obtained Earthflows Map

The map obtained from the discriminant function has to be used with caution. Beside the results in terms of validation-index or effectiveness graphs, it is important to understand the possible applications of this map.

The susceptibility can be thought as a intrinsic characteristic of the terrain, which depends on lithological and morphological conditions. In order to evaluate the areas with potential slope instability, it is necessary to understand the particular geomorphological conditions under which a landslide may occur. In order to do that, the understanding of landslide inventories represents a fundamental step.

Landslide Inventory Maps give information about the spatial distribution of past and present phenomena occurred in a certain area. Regarding this work, the analyses developed allow to evaluate the most influencing variables on slope instability in the study area. Knowing the particular lithological and morphological conditions under which past landslides occurred, it is possible to evaluate areas of potential slope instability, simply verifying if the same conditions subsist in that areas.

Several examples reported below try to explain the perspectives for the applicability of the used approach.

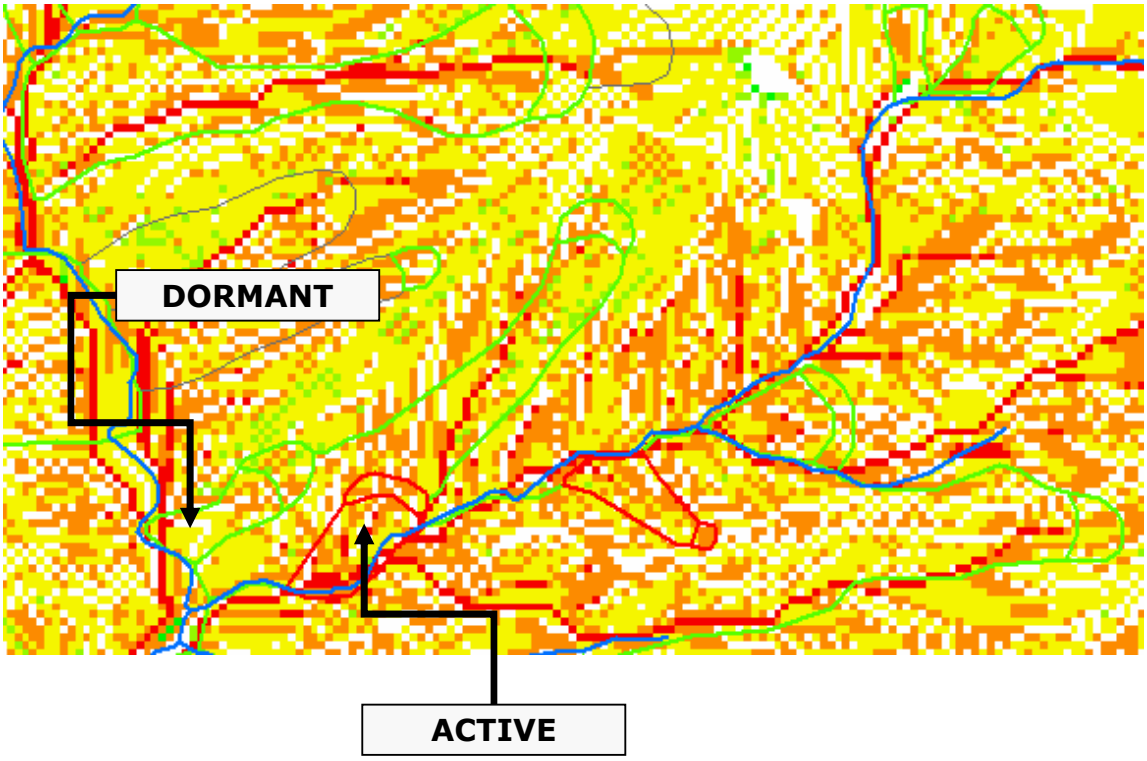


Figure 58. Extract of the obtained map

In the Figure 58 it is possible to observe in particular two phenomena: a dormant landslide and an active landslide. The analysis of the percentage of the cells belonging to the defined classes demonstrate that the active landslide contains a greater number of cells belonging to high and very high susceptibility classes. It can be explained considering that the conditions subsisting in the dormant landslide are farer from the ones which can lead to instability, respect with the conditions under which the active landslide lives. It confirms that the obtained map can be used as a way to interpret the Landslide Inventory.
Moreover, it can be observed that a great concentration of cells belonging to the high and very high susceptibility classes subsist in certain areas. It means that the particular lithological and morphological conditions which have lead to slope instability in the past may subsist in areas without any recorded movements. In this case, the map can be used as a way to individuate areas of potential slope instability, a way to understand if a need of investigation is required in certain areas that are defined stable by the Inventory Map.

Figure 59. Extract of the obtained map

Another interesting example is reported in Figure 59. Several movements have been observed in this area. However, due to the concentration of cells belonging to high and very high susceptibility classes, the same conditions seem to be present also in other areas, which represent therefore zones of potential slope instability.

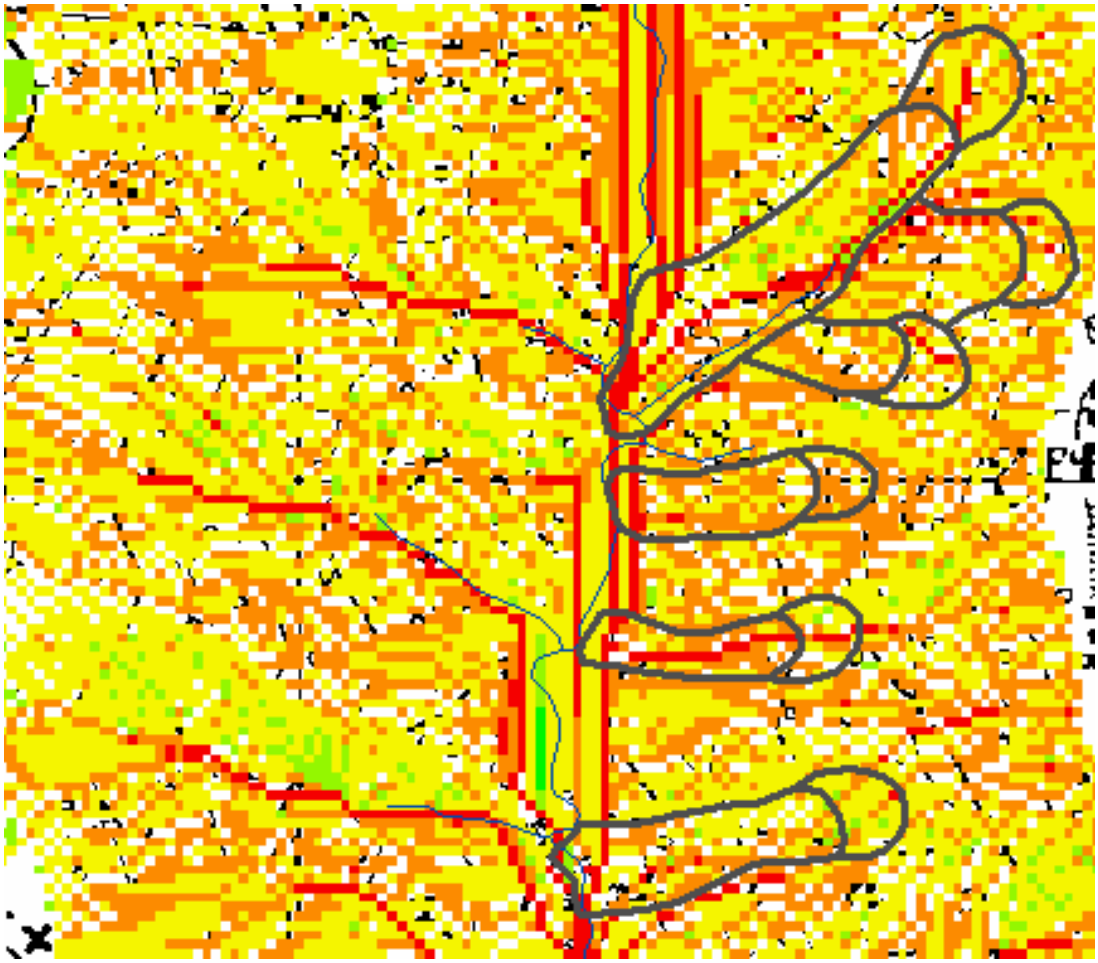


Figure 60. Extract of the obtained map

A drawback of statistical multivariate analyses is the strong dependence of their reliability from the types and quality of the chosen parameters. The Figure 60 underlines this important concept. Observing the right and the left slopes, as the obtained map gives almost the same results in terms of percentage of cells belonging to the different classes, it seems that the same lithological and morphological conditions characterize them. However, the Inventory Map has recorded movements only on the right slope. It can be explained considering that the analysis developed in this work has employed only variables related to the terrain's surface conditions, without any information about the subsurface water conditions or about the structural geology. The conditions in terms of positions of the geological layers may be different on the two slopes.

The example reported in Figure 60 shows the presence of hollows in which the geomorphologist while creating the Inventory Map has recognized creep phenomena. These phenomena are characterized by plastic deformations of the material. From the figure, it can be observed that few of the hollows have a great concentration of cells belonging to the high and very high susceptibility classes, while other present a greater concentration of cells belonging to low and moderate susceptibility classes. It means that the potentiality of developing instability is different. The hollows more filled of material present such plastic deformations that the creep may evolve in failures. The other hollows, instead, live under conditions that, even though characterized by deformations, are far from instability.

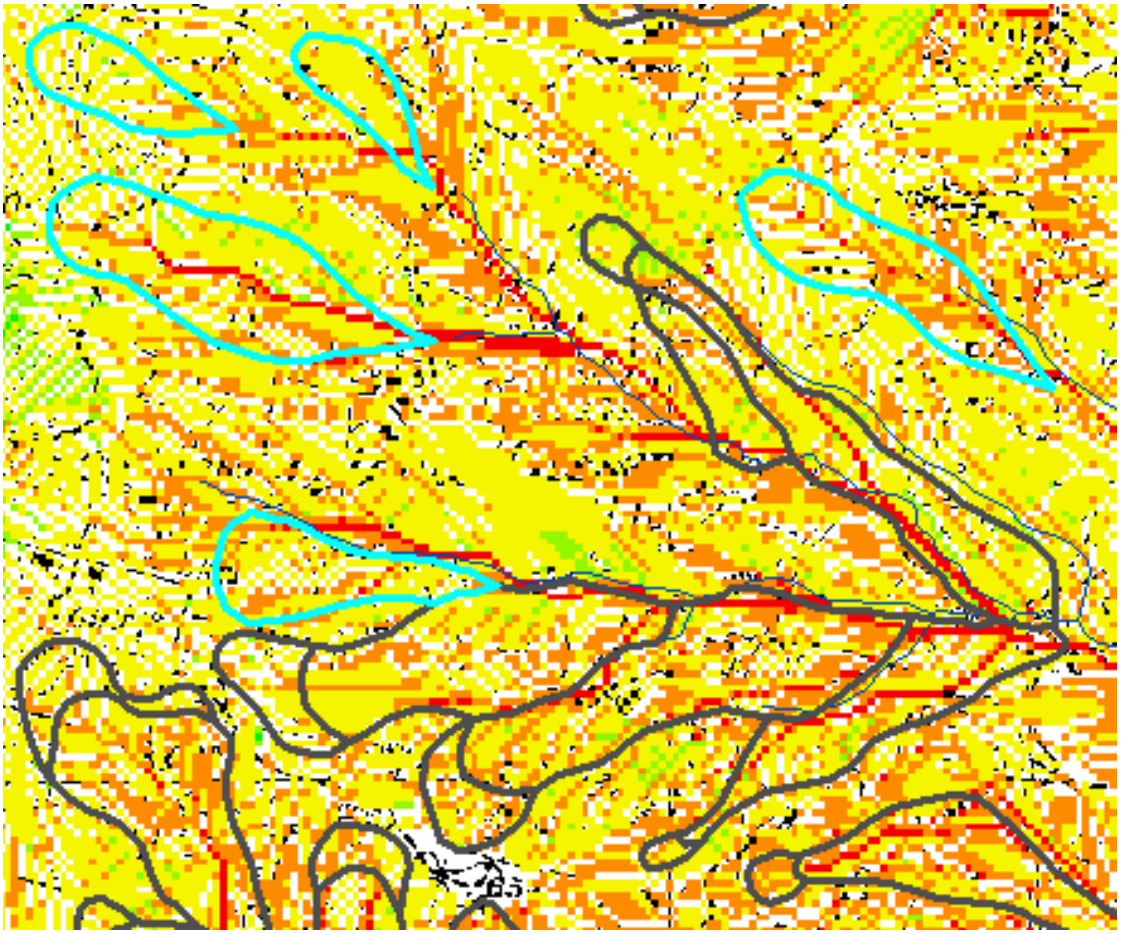


Figure 61. Extract of the obtained map

The last example (Figure 62) show how the correct applicability of the obtained map and its understanding may be improved using information about damage. The circles indicate the position of investigations effectuated to evaluate the presence or absence of damage. The circles in bold represent the presence of damage, while the other indicate an absence of damage. Observing the active landslide in figure, it can be observed a great concentration of red cells on the lower part of the slope. It means that the landslide, defined as active in the Inventory map, may be characterized by movement only in the lower part, remaining dormant in the upper part. This condition may be explained considering the strong influence that the river exercises at the toe of the slope. Adding information on damage to the obtained map, the described conditions are confirmed. Damage are recorded only at the toe of the slope.

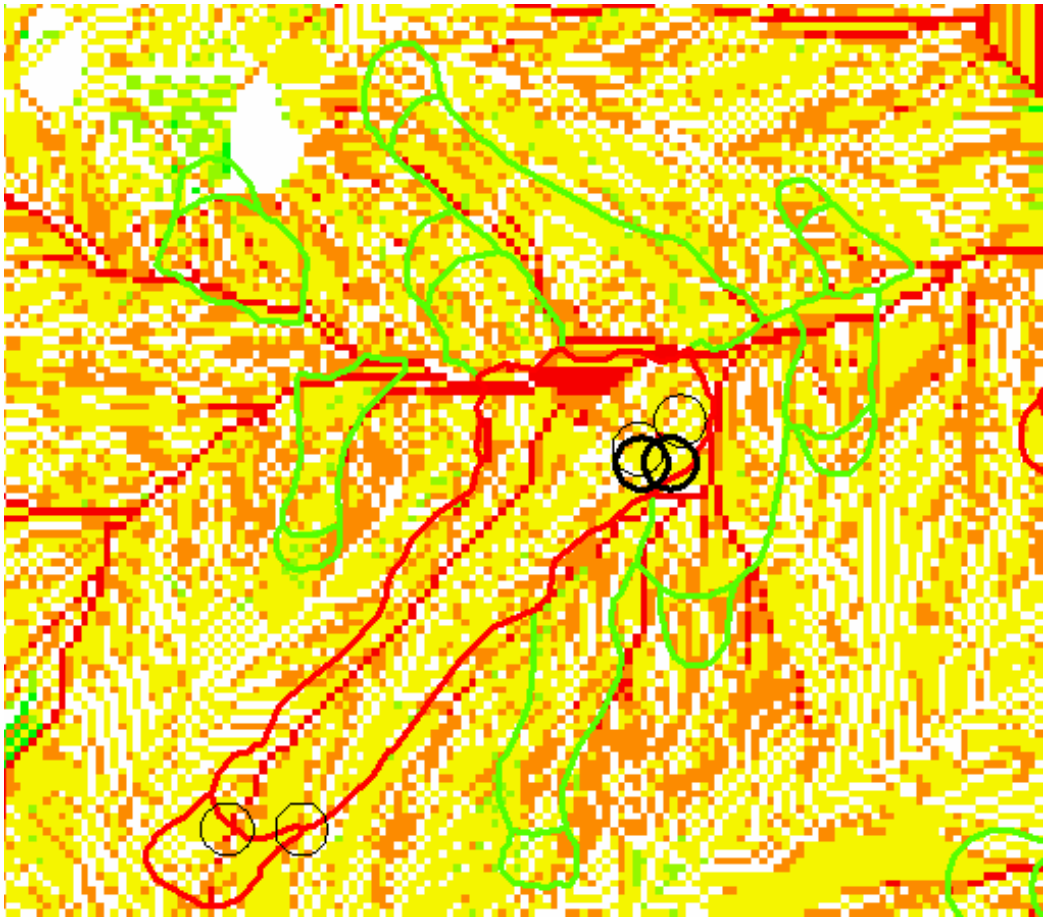


Figure 62. Extract of the obtained map

CHAPTER 6

CONCLUDING REMARKS

Over the past years, government and research institutions have invested considerable resources in assessing landslide susceptibility and hazard, and in attempting to produce maps portraying their spatial distribution. Several different methods and techniques for evaluating landslide susceptibility, hazard and risk have been proposed or tested. Among the different approaches developed in literature to assess landslide susceptibility, the statistical multivariate analyses have proved very useful.
The majority of papers dealing with multivariate analysis discuss specific attempts at the evaluation of landslide susceptibility and hazard, trying to identify the most influential parameters on landsliding. However, these works investigate all the mass movements in order to create a landslide susceptibility map in limited areas, while very few studies have applied statistical procedures to assess large landslide susceptibility over large areas.
In this investigation we used multivariate analysis to study large landslide distribution over the Italian Benevento Province, on a 1:25000 working scale.
Two attempts have been developed, following the statistical multivariate procedure used by Baeza&Corominas (2001) to study shallow landslide susceptibility in a study area in the Spanish Eastern Pyrenees.

The first attempt consists in the application of the statistical multivariate technique to characterize the actual landslide distribution, in particular to distinguish the two populations of active and dormant earthflows. The variables used in the analysis in order to carry on the statistical treatment are representative of the behavior of the moving mass and are not related to the scarps. These variables have been classified as belonging to different groups, according to the type of information provided: Digital Elevation Model; Geometry; Geology; Activity state; Hydrographic network.
The results obtained are inconsistent, demonstrating that a clear distinction between the two groups is not possible with the parameters taken into account. All the variables used in the analysis have been captured automatically. The lack of data and their quality influence considerably the reliability of the results.
Actually, the spatial landslide distribution is the result of the interaction of many factors, some of which are difficult to incorporate in susceptibility analysis. A reliable and accurate susceptibility assessment depends on the proper identification of these factors. The inclusion or omission of some of the factors may change significantly the capability of susceptibility assessment.

The second attempt consists in the application of the statistical analysis to assess terrain susceptibility to produce earthflows, using parameters related to the occurrence of landsliding from the scarp areas. Also in this case, the variables adopted have been classified as belonging to different groups, according to the type of information provided: Digital Elevation Model; Geometry; Watershed dimension; Geology; Landslide. All the variables have been captured automatically.
The results obtained in this case are consistent, as shown below.

- The discriminant function has allowed us to correctly classify nearly the 80% of

landslides. It may be considered as a good result, although it may be improved using also data gathered directly in the field.

- The results demonstrate that the variables selected as the most influential factors on stability are related to the watershed dimension, the superficial run-off, the lithology and the plane curvature. In particular high values of plan curvature (*PL*), stream power index (*SPI*) and lithology (*ST*) increase slope instability, while high values of watershed length (*LONG*) and watershed mean angle (*PENDM*) increase stability. According to this, failures are expected on slopes with gentle watershed, where the ability of groundwater flow to reach the potential landslide site is relevant. The importance of the *SPI* is particularly high, demonstrating the relevance of the erosive power of water flow on slope instability. Regarding the soil type (*ST*), the lithology showing the greatest index of landslide density is the most prone to landsliding and thus the most contributing to the slope instability. Thus it is expected that first-time failure earthflows occur mostly in flysh deposits.

- The area under the curve (AUC) is nearly equal to 0.9, that represents a good result. It means that the analysis developed has good capacities of predict the occurrence of landslides.

- The index of relative landslide density used to assess the effectiveness of the obtained map demonstrates that slope failures would appear in cells having higher discriminant scores. In particular, five classes were established according to ranges based on the standard deviation of the distribution obtained: I (very low); II (low); III (moderate); IV (high); V (very high). The presence of landslides in each susceptibility level increases from 'very low' level to 'high' level. The distribution of slope failures observed in these classes indicates that susceptibility levels are consistent.

- The effectiveness of the landslide susceptibility map was evaluated relating the areal distributions of observed landslides with susceptibility intervals of 10%, as is usually done in logistic regression. It is necessary demonstrate that: (i) on the map the most of the actual landslides should have to be located in the cells included in high susceptibility classes; (ii) on the map, these high susceptibility classes should have to cover small area as possible. In this work, approximately 80% of the observed landslides locate in the class of *'susceptible'* and the cumulative area of the susceptibility values is obtained as 20%, so that the two rules are well satisfied.

The same analysis has been developed considering the movements classified as slides. The stepwise method gives nearly the same results. The discriminant function obtained, the variables caught as the most influential factors on slope instability, the susceptibility classes defined and the methods employed to evaluate the effectiveness of the susceptibility map demonstrate that a clear distinction between the two typologies of movements considered is not possible with the parameters taken into account. Some factors which play a significant role in the distinction between the two typologies of movements are probably missing in these analyses. The results may be improved by adding to the data set information relative to factors that are better linked to the processes that led to the occurrence of an earthflows rather than a slide.

As the basis of the two analyses is the same Landslide Inventory Map, which was constructed from aerial photography, the examples stress also the importance of a correct interpretation of landslide distribution to produce reliable susceptibility analyses. Several studies have shown

that differences between the interpretations carried out by different observers can be very large (Carrara et al., 1992; Dunoyer and Van Westen, 1994) and that uncertainties in landslide inventories restrict the applicability of susceptibility maps. Actually the appropriate completion of the landslide inventories still rely on the skill of the specialists, while the identification and interpretation of landslide features should be more evident and less subjective.

Moreover, multivariate techniques provide quantified evaluation of the simultaneous influence of different factors and therefore a more realistic and objective approach to the assessment of landslide susceptibility. The susceptibility can be thought as an intrinsic characteristic of the terrain, which depends on lithological and morphological conditions. In order to evaluate the areas with potential slope instability, it is necessary to understand the particular geomorphological conditions under which a landslide may occur. In order to do that, the understanding of landslide inventories represents a fundamental step.
Landslide Inventory Maps give information about the spatial distribution of past and present phenomena occurred in a certain area. Regarding this work, the analyses developed allow to evaluate the most influencing variables on slope instability in the study area. Knowing the particular lithological and morphological conditions under which past landslides occurred, it is possible to evaluate areas of potential slope instability, simply verifying if the same conditions subsist in that areas.

Finally, it is important to stress the importance of GIS. The reliability of statistical multivariate depends mostly on the amount and quality of the variables chosen in the analyses. Because many factors can play a role in the occurrence of mass movements, the analysis is complex and requires a large number of input variables. The development of Geographic Information Systems (GIS) has enhanced the capabilities for susceptibility assessment over large region. GIS are considered as a faster and more efficient acquisition and processing of those geological-geomorphological data which are both relevant in assessing landslide susceptibility and mappable at effective cost over wide regions. The performance of neighborhood operations with the GIS allows extraction of morphological and hydrological parameters from Digital Elevation Models (DEM), that otherwise would be difficult to obtain. The main goal is the automatic capture of most of the parameters in relation to the occurrence of slope failures.

INDEX

REFERENCES

Ardizzone F, Cardinali M, Carrara A, Guzzetti F, Reichenbach P. 2002. Impact of mapping errors on the reliability of landslide hazard maps. *Natural Hazards and Earth System Science* **2**: 3-14.

Atkinson, P.M. and Massari, R. 1998. Generalised linear modelling of susceptibility to landsliding in Cetral Apennines, Italy. Computers and Geosciences 24: 373-385

Ayalew L, Yamagishi H. 2005. The application of GIS-based logistic regression for landslide susceptibility mapping in Kakuda-Yahiko Mountains, Central Japan. *Geomorphology* **65**: 15-31.

Baeza C. 1994. *Evaluaciòn de las condicines de rotura y de la movilidad de los deslizamientos superficilaes mediante el uso de tècnicas de analisis multivariante.* Doctoral thesis, Universidad Politécnica de Catalunya.

Baeza C, Corominas J. 2001. Assessment of shallow landslide susceptibility by means of multivariate statistical techniques. *Earth Surf. Process. Landforms* **26**, 1251-1263.

Baum, R. L.; Coe, J.A.; Godt, J.W.; Harp, E.L.; Reid, M.E.; Savage, W.Z.; Schulz, W.H.; Brien, D.L.; Chleborad, A.F.; McKenna, J.P. and Michael, J.A. (2005) Regional landslide-hazard assessment for Seattle, Washington, USA. *Landslides*, 2 (4): 266-279.

Baynes, F.J. and Lee, E.M. (1998). Geomorphology in landslide risk analysis, an interim report. *Proc.8th Congress of the Int. Assoc. Eng. Geologists*. Editors P Moore and O Hungr, Balkema, 1129-1136.

Bonham-Carter GF, Agterberg FP, Wright DF. 1990. Weights of evidence modelling: a new approach to mapping mineral potential. In *Statistical Applications in the Earth Sciences*, Agterberg FP, Bonham-Carter GF (eds). Geological Survey of Canada, Paper 8–9: 171–183.

Brabb, E.E., E.H. Pampeyan and M.G. Bonilla. (1972). Landslide susceptibility in San Mateo County, California. *U.S. Geol. Surv., Misc. Field Studies*, Map MF-360. Scale 1:62,500.

Brabb, E. E.: 1984, Innovative approaches to landslide hazard and risk mapping, *IV InternationalSymposium on Landslides*, Vol. 1, Toronto, pp. 307–323.

Can T, Nefeslioglu HA, Gokceoglu C, Sonmez H, Duman TY. 2005. Susceptibility assessment of shallow earthflows triggered by heavy rainfall at three catchments by logistic regression analyses. *Geomorphology* **72**: 250-271.

Carrara, A. Catalano, E., Sorriso-Valvo, M., Reali, C., and Osso, I.: 1978. Digital terrain analysis for land evaluation, *Geologia Applicata e Idrogeologia* **13**, 69–127.

Carrara A. 1983a. Multivariate models for landslide hazard evaluation. *Mathematical Geology* **15**(3): 403–426.

Carrara A. 1983b. Geomathematical assessment of regional landslide hazard. In *Proceedings of the 4th International Conference on the Applications of Statistics and Probability in Soil and Structural Engineering*. Augusti G, Vannucchi G (eds). (Università di Firenze) Pitagora Editrice, Bologna; 3–27.

Carrara A. 1984. Landslide hazard mapping: aims and methods. *Bureau de Recherches Gèologiques et Minière* **83**: 141–151.

Carrara A., Cardinali M., Detti M, Guzzetti F, Pasqui V, Reichenbach P. 1991. GIS techniques and statistical models in evaluating landslide hazard. *Earth Surface Processes and Landforms* **16**: 427-445.

Carrara A., Cardinali M., Guzzetti F. 1992. Uncertainty in assessing landslide hazard and risk. *ITC Journal* **2**: 172-183.

Carrara, A., Cardinali, M., Guzzetti, F., and Reichenbach, P.: 1995, GIS technology in mapping landslide hazard, In: A. Carrara and F. Guzzetti (eds), *Geographical Information Systems in Assessing Natural Hazards*, Kluwer Academic Publishers, pp. 135–175.

Cascini L, Bonnard Ch, Corominas J, Jibson R, Montero-Olarte J. 2005. Landslide hazard and risk zoning for urban planning and development. *Landslide Risk Management – Hungr, Fell, Couture & Eberhardt (esd).*

Chung CF, Fabbri A. 1993. The representation of geoscience information for data integration. *Nonrenewable Resources* **2**(2): 122–139.

Chung CF, Leclerc Y. 1994. A quantitative technique for zoning landslide hazard. In *International Association for Mathematical Geology Annual Conference* (IAMG'94) Quebec, Canada: 87–93.

Clerici A, Perego S, Tellini C, Vescovi P. 2002. A procedure for landslide susceptibility zonation by the conditional analysis method. *Geomorphology* **48**: 349-364.

Coe JA, Godt JW, Baum RL, Bucknam RC, Michael JA. 2004. Landslide susceptibility from topography in Guatemala. In: Lacerda, Ehrlich, Fontoura & Sayao (eds.), *Landslide: Evaluation and Stabilization*, Taylor & Francis Group, London, ISBN 04 1535 665 2.

Corominas J, Alonso E. 1990. Geomorphological effects of extreme floods (November 1982) in the Southern Pyrenees. In *Proceedings of the 2nd Symposium of Hydrology in Mountainous Regions Artificial Reservoirs: Water and Slopes*. International Association of Hydrological Sciences, Lausanne, **194**; 295–302.

Corominas J. 1992. Landslide risk assessment and zoning. In *Planning the Use of the Earth's Surface*, Cendrero A, Lütting G, Wolf FC (eds). Springer-Verlag: Berlin; 141–173.

Corominas J, Baeza C, Salueña I. 1992. The influence of geometrical slope characteristics and land use on the development of shallow landslides. In *Proceedings of the 6th International Symposium on Landslides* D. H. Bell (ed.). 919–924.

Corominas J, Moya J. 1999. Reconstructing recent landslide activity in relation to rainfall in the Llobregat River basin, Eastern Pyrenees, Spain. *Geomorphology* **30**: 73–93.

Cruden, D. M. and Varnes, D. J.: 1996, Landslide types and processes', In: A. K. Turner and R. L. Schuster (eds), *Landslides: Investigation and Mitigation*, TRB Special Report, 247, National Academy Press, Washington, pp. 36–75.

Dai, C.F. and Lee, C.F. (2001). Terrain based mapping of landslide susceptibility using a geographic information system: a case study. *Canadian Geotec J. :*38, 911-923

David M, Dagbert M, Beaucheimin Y. 1977. Statistical analysis in geology: Correspondence analysis method. *Quarterly Journal of theColorado School of Mines* **72**: 1–45

Dewitte O, Chung CJ, Demoulin A. 2006. Reactivation hazard mapping for ancient landslides in West Belgium. *Natl Hazards Earth Syst Sci* **6**: 653-662.

Dillon,W. R. and Goldstein,M.: 1986, *Multivariate Analysis. Methods and Applications*, JohnWiley and Sons Ltd., 565 pp.

Duman TY, Can T, Gokceoglu C, Nefeslioglu HA, Sonmez H. 2006. Application of logistic regression for landslide susceptibility zoning of Cekmece Area, Istanbul, Turkey. *Environ Geol* **51**: 241-256,

Dunoyer M, van Westen CJ. 1994. Assessing uncertainty in interpreting landslides from air-photos. *ITC Journal* **3.**

Evans, N.C. and King, J.P. (1998). *The natural terrain landslide study.Debris avalanche susceptibility*. Technical Note TN 1/98, Geotechnical Engineering Office, Hong Kong

Fell, R., Ho, K.K.S., Lacasse, S. and Leroi, E. (2005). A framework for landslide risk assessment and management. In *Landslide Risk Management*, Editors O Hungr, R Fell, R Couture and E Eberhardt, Taylor and Francis, London,3-26

Flentje, P, Stirling, D and Chowdhury, R, (2007). Landslide Susceptibility and Hazard derived from a Landslide Inventory using Data Mining – An Australian Case Study. *Proceedings of the First North American Landslide Conference*. Vail, Colorado, June 2007. Abstract accepted, paper submitted.

Gallart F, Clotet N. 1988. Some aspects of the geomorphic processes triggered by an extreme rainfall event: The November 1982 flood in The Eastern Pyrenees. *Catena supplement* **13**: 79–95.

Gómez, H. and Kavzoglu, T. 2005. Assessment of shallow landslide susceptibility using artificial neural networks in Jabonosa river basin, Venezuela. Engineering Geology, 78: 11-27

Greenway DR. 1987. Vegetation and slope stability. In *Slope Stability*, Anderson MG, Richards KS (eds). John Wiley: Chichester; 187–231.

Guadagno FM et al. *Carta delle frane della Provincia di Benevento*. Università degli studi del Sannio.

Guzzetti F, Carrara A, Cardinali M, Reichenbach P. 1999. Landslide hazard evaluation: a review of current techniques and their application in a multiscale study, Central Italy. *Geomorphology* **31**: 181-216.

Hansen MJ. 1984. Strategies for classification of landslides. In *Slope Instability*, Brunsden D (ed.). John Wiley: Chichester; 1–25.

Hartlen J, Viberg L. 1988. Evaluation of landslide hazard. In *Proceedings of 5th International Symposium on Landslides*. Bonnard CH(ed.). Balkema: Rotterdam; **2**: 1037–1057.

Hungr O, Corominas J, Eberhardt E. 2005. Estimating landslide motion mechanism, travel distance and velocity. In: Hungr, Fell, Couture & Eberhardt (eds.), *Landslide Risk Management*, Taylor & Francis Group, London, ISBN 04 1538 043 X.

Jones FO, Embody DR, Peterson WL. 1961. *Landslide along the Columbia river valley, Northeastern Washington* US Geological Survey Professional Paper 367.

Kienholz, H. (1978) Map of geomorphology and natural hazards of Grindelwald, Switzerland, scale 1:10,000. *Artic and Alpine Research*, 10: 169-184.

Lebart L, Morineau A, Fenelan JP. 1982. *Traitement des don´ees statistiques*. Dunad: Paris.

Lee, S.; Ryu, J.H.; Lee, M.J. and won, J.S. 2006. The application of artificial neural networks landslide susceptibility mapping at Janghung, Korea. Mathematical Geology, 38: 199-220

Leroi, E., Bonnard,C., Fell, R. and McInnes, R (2005) Risk assessment and management. In *Landslide Risk Management*, Editors O Hungr, R Fell, R Couture and E Eberhardt, Taylor and Francis, London, 159-198

Leroueil, S. (2001) Natural slopes and cuts: Movements and failure mechanisms. *Geotechnique*, 51, No 3, 197-243.

Luzi L. 1995. *GIS for slope stability zonation in the Fabriano area, central Italy* PhD thesis, International Institute for Aerospace Survey and Earth Sciences (ITC), Enschede.

Montgomery, D.R. and Dietrich, W.E. 1994. A physically based model for the topographic control on shallow landsliding. Water Resources Research, 30: 1153-1171

Montgomery, D.R. Sullivan, K. and Greenberg, H.M. 1998. Regional test of a model for shallow landsliding. Hydrological Processes, 12: 943-955

Moore ID, Grayson RB, Ladson AR. 1991. Digital terrain modeling: a review of hydrological, geomorphological and biological applications. *Hydrol Process* **5**: 3-30.

Muñoz JA. 1985. *Estructura alpina i herciniana a la vora sud de la zona axial del Pirineu Oriental* Doctoral thesis Facultad de Geologia, Universitat de Barcelona.

Muñoz JA, Mart´ınez A, Verg´es J. 1986. Thrust sequences in the Spanish Eastern Pyrenees. *Journal of Structural Geology* **8**(3/4): 399–405.

Naranjo JL, van Westen CJ, Soeters R. 1994. Evaluating the use of training areas in bivariate statistical landslide hazard analysis – a case study in Colombia. *ITC Journal* **3.**

Neuland H. 1976. A prediction model of landslips. *Catena* **3**: 215–230.

Nilsen TH, Wright RH, Vlasic TC, Spangle VE. 1979. *Relative slope stability and land-use planning in the San Francisco Bay Region, California* US Geological Survey Professional Paper 944.

Novoa M. 1984. Precipitaciones y avenidas extraordinarias en Catalunya. *Jornadas de trabajo sobre Inestabilidad de laderas en elPirineo* **I**: 1.1–5.1.

Okimura T. 1983. A prediction system for the site of probable surface failure of mountain-slope by topographical factors. *Proc. JapanSociety of Civil Engineers*; **338**: 131–138.

Oyagi N. 1984. Landslides in weathered rocks and residual soils in Japan and surrounding areas: A state of the art report. In *Proceedings of the 4th International Symposium on Landslides*. University of Toronto: Downsview; **3**: 1–31.

Quinlan, J.R., C4.5: Programs for Machine Learning. (1993).

Santacana N, Baeza C, Corominas J, De Paz A, Marturià J. 2003. A GIS-Based Multivariate Statistical Analysis for Shallow Landslide Susceptibility Mapping in La Pobla de Lillet Area (Eastern Pyrenees, Spain). *Natural hazards* **30**: 281-295.

Savage, W.Z., Godt, J.W., and Baum, R.L., (2004), Modeling time-dependent slope stability: *Proceedings IX International Symposium on Landslides*, Rio de Janeiro, Brazil, June 27-July 2, 2004, p. 23-38.

Smith TC. 1988. A method for mapping relative susceptibility to debris flow, with an example from San Mateo County. In *Landslides,Floods and Marine Effects of the Storm of January 3–5, 1982, in San Francisco Bay Region, California*, US Geological Survey: United States; Professional Paper; **1434**: 185–194.

Soeters R, van Westen CJ. 1996. Slope instability recognition, analysis and zonation. In Turner&Schuster (eds.) *Landslide Investigation and Mitigation.* TRB Special Report **247**: 129-177, National Academy Press, Washington D.C.

SPSS Inc. 1988. *SPSS-X Users Guide*, 3rd edn. Gorinchen.

Süzen ML, Doyuran V. 2004. Data driven bivariate landslide susceptibility assessment using geographical information systems: a method and application to Asarsuyu catchment, Turkey. *Engineering Geology* **71:** 303-321.

van Westen CJ. 1994. GIS in landslide hazard zonation: a review with examples from the Andes of Colombia. In *Mountain Environment and Geographic System*. Price M, Heywood DI (eds). Taylor and Francis: London; 135–165.

van Westen CJ, Seijmonsbergen AC, Mantovani F. 1999. Comparing landslide hazard maps. *Nat Hazards* **20**: 137-158.

van Wescten, C.J. and Terlien, M.T.J. 1996. Deterministic landslide hazard analysis in GIS: a case study from Manizxales (Colombia). Earth Surface Processes and Landforms, 21: 853-868

Varnes D.J. (1984). Landslide hazard zonation: a review of principles and practice. UNESCO Press, Paris, 63 pp

WP/WLI (Working Party on World Landslide Inventory). 1993. A suggested method for describing the activity of a landslide. *Bull Int Assoc Eng Geol* **47**: 53-57.

Yesilnacar E, Topal T. 2005. Landslide susceptibility mapping: a comparison of logistic regression and neural networks methods in a medium scale study, Hendek region (Turkey). *Engineering Geology* **79:** 251-266.

Yin KL, Yan TZ. 1988. Statistical prediction models for slope instability of metamorphosed rocks. In *Proceedings of the 5th International Symposium on Landslides*, Bonnard CH (ed.). Balkema: Rotterdam; **2**: 1269–1272.

Xie, M.; Esaki, T. and Zhou, G. 2004. GIS-based probabilistic mapping of landslide hazard using a three-dimensional determministic model. Natural Hazards, 33. 265-282

Zhou, G.; Esaki, T.; Mitani, Y.; Xie, M. And Mori, J. 2003. Spatial probabilistic modelling of slope failure using an integrated GIS Monte Carlo simulation approach. Engineering Geology, 68: 373-386