TECHNISCHE UNIVERSITÄT DRESDEN

Faculty of Environmental Sciences Institute for Cartography

MASTER THESIS

Landslide hazard in Central Asia

Understanding the relationship between slope instabilities, tectonic and geomorphology using satellite data and integration into a landslide susceptibility model

Submitted by: **Laura Natalie Barbosa Mejia**
Born on: 18.08.1990 in Ibagué, Tolima - Colombia

Submmited for the academic degree of
Master of Science (M.S.c)

Submmision on 15.10.2018

Supervisors Prof. Dr. habil. Elmar Csaplovics
Institute of Photogrammetry and Remote Sensing, TUD

DRS. B.J. Barend Köbben
Institute for Geo-information Sciences and Earth Observation (ITC)

External Supervisors Dr. Louis Andreani
Helmholtz Institute Freiberg for Resource Technology

Dr. Richard Gloaguen
Helmholtz Institute Freiberg for Resource Technology

Herewith I declare that I am the sole author of the thesis named ***„Landslide Hazard in Central Asia. Understanding the relationship between slope instabilities, tectonic and geomorphology using satellite data and integration into a landslide susceptibility model“*** *which has been submitted to the study commission of geosciences today. I have fully referenced the ideas and work of others, whether published or un-published. Literal or analogous citations are clearly marked as such.*

Dresden, 15th of October 2018

Laura Natalie Barbosa Mejia

Abstract

Central Asia is highly exposed to natural disasters that annually lead to substantial deaths and economic losses. Some of them can be prevented through a better understanding of the geological and geomorphological processes. Landslides are a common event in the Pamir and Tien Shan, triggered by climatic conditions, landscape features, and tectonic activity. To understand which factors are more relevant to the landslide occurrence and identify regional areas that should be studied in detailed in future works, this study proposes to analyse the spatial association between different factors and the known landslides occurrences. We then use this knowledge in order to model the landslide susceptibility for the area which encompasses the Tadjik basin, South Western Tien Shan and Western Pamir. The weight of evidence (WOE) uses the log-linear form of the Bayesian probability to assign weights to the predictive factors and relate them to the landslide occurrence. The Logistic Regression (LR) is a modification of linear regression that fits a sigmoid curve equation to a dependent binary variable and a certain number of independent variables. Random Forest (RF) implements the Bayesian tree combined by the idea of bagging and random feature selection to grow a forest of many trees. The receiver operating characteristic (ROC) is used to evaluate the performance of the model along with the model error. The best model is selected for each of the methods, and the differences among them are discussed. Finally, the results are compared with previous works from the area.

Zusammenfassung

Zentralasien ist stark von Naturkatastrophen betroffen, welche jährlich mit Todesfällen und wirtschaftlichen Verlusten einhergehen. Einige dieser Folgen könnten durch ein besseres Verständnis der geologischen und geomorphologischen Prozesse vermieden werden. Erdrutsche sind häufige Ereignisse im Pamir und Tien Shan, welche durch klimatische Bedingungen, Landschaftsmerkmale und tektonische Aktivitäten ausgelöst werden. Um zu verstehen welche Faktoren für ein Erdrutschereignis relevant sind und um regionalen Gebiete zu identifizieren, die in zukünftigen Arbeiten detailliert untersucht werden sollten, schlägt diese Studie vor den räumlichen Zusammenhang zwischen verschiedenen Faktoren und den bekannten Erdrutschereignissen zu analysieren. Wir nutzen dieses Wissen, um die Erdrutschanfälligkeit für das Gebiet südwestlich des Tadjik-Becken von Tien Shan und West-Pamir umfassend, zu modellieren. Die verwendete weight of evidence-Methode (WOE) ist eine loglineare Form der Bayes'schen Wahrscheinlichkeit, um den prädiktiven Faktor zu wichten und mit dem Erdrutschereignis in Beziehung zu setzen. Die Logistische Regression (LR) ist eine Modifikation der linearen Regression, die eine sigmoidale Kurvenausrichtung an eine abhängige binäre Variable und eine bestimmte Anzahl unabhängiger Variablen anpasst. Random Forest (RF) implementiert den Bayes'schen Wahrscheinlichkeitsbaum, kombiniert mit der Idee des Baggings und der zufälligen Merkmalsauswahl, um einen Wald mit vielen Wahrscheinlichkeitsbäumen wachsen zu lassen. Die Receiver operating characteristics (ROC) wird verwendet, um die Leistung des Modells, unter Berücksichtigung des Modellfehlers, zu bewerten. Für jede der Methoden wird das beste Modell ausgewählt und die Unterschiede zwischen ihnen diskutiert. Schließlich werden die Ergebnisse mit früheren Arbeiten in der Region verglichen.

Contents

List of Figures

List of Tables

List of abbreviations

- EM-DAT - International Disaster Database
- GSHAP - Global Seismic Hazard Assessment Program
- HAR - High Asia Refined analysis project
- CAFD - Central Asia Fault Database
- RGI - Global inventory of glaciers outlines
- EMCA - Earthquake Model Central Asia
- DEM - Digital elevation model
- LS - Lanslide susceptibility
- WOE - Weight of evidence
- RF - Random Forest
- LR - Logistic Regression
- GLOF's - Glacial Lake Outburst Floods
- MPT - Main Pamir Trust
- DFZ - Darvaz Fault Zone
- GIS - Geographical information systems
- NDVI - Normalized difference vegetation index
- DEM - Digital elevation model
- TW - Topographic wetness index
- TPI - Topographic position index
- SR - Surface roughness
- ERR - Elevation relief ratio
- SI - Surface index
- ROC - Receiver operating characteristic curve
- AUC - Area under the curve
- VIFs - Variance inflation factors
- OOBE - Out-of-bag error

Chapter 1

INTRODUCTION

The occurrence of different natural hazard like earthquakes, landslides, extreme temperatures and rainfalls worries the communities located in vulnerable areas not only for the structural vulnerability of the cities but also for the economic vulnerability of the country. Natural disasters cause many deaths and injuries per year around the world. According to the EM-DAT (The international disaster Database), just in 2016, 342 disasters triggered by natural hazards caused a total of 8.733 fatalities and 569.4 million people were affected. The number of deaths reported for 2016 were the second lowest since 2006, however, the number of people affected was the highest, resulting in an estimated cost of 154 billion US Dollars (Guha-Sapir *et al.* , 2017).

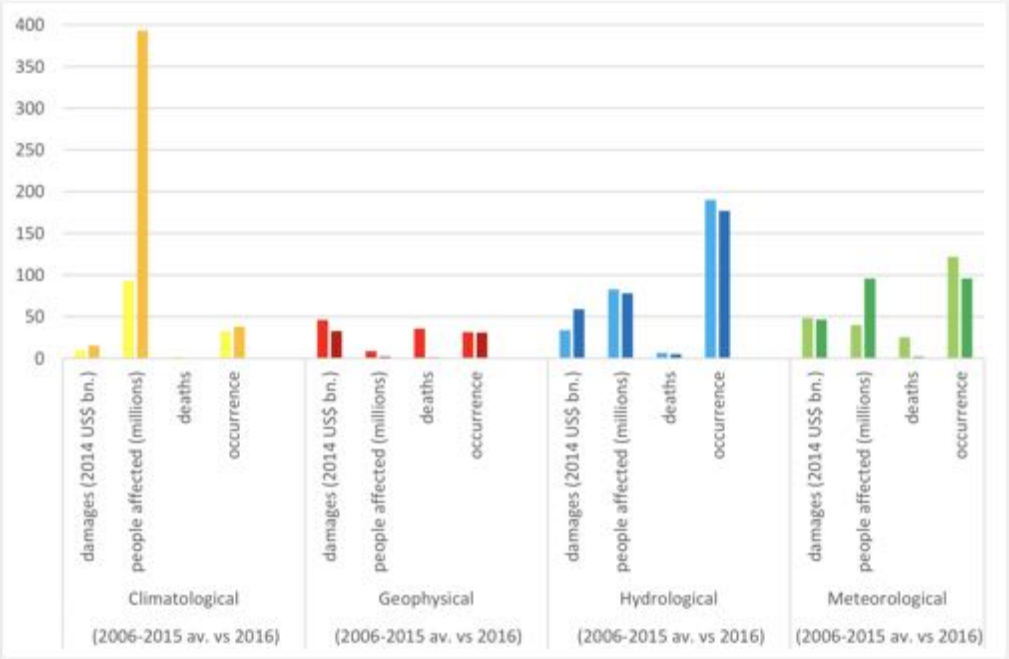


Figure 1.1: *Natural disasters impact. Comparison between 2016 and 2006 to 2015 data. Climatological group: Drought, glacial lake outburst, wildfire; Geophysical group: earthquake, mass movement (dry),volcanic activity; Hydrological group : Flood, landslide, wave action; Meteorological group: storm,extreme temperature, fog. source: (Guha-Sapir* et al. *, 2017)*

The most common natural disaster during the period of 2006 to 2016 are related to hydrological processes that include floods, landslides and wave action with an average of 50.5% of events and an 2006-2017 annual average of 6 657 deaths (Guha-Sapir *et al.*, 2017). While hydrological events occur more frequently, the rate of morality is lower for the events related to geophysical events like earthquakes, mass movements or volcanic activity (figure 1.1). Earthquakes are the most common geophysical event, that caused 1315 deaths in 2016.

The spatial distribution of the worldwide natural catastrophes, provides evidence that for 2016, 46% of the events happened in Asia; followed by Americas, Africa, Europe and Oceania. The hydrological disasters are the most frequently reported in Asia. During 2016 a total of 12905 people were affected by floods or landslides in Central Asia, of which 12750 were affected by a single flood in Tajikistan. (Guha-Sapir *et al.*, 2017). Moreover, landslides are a common mountainous process and Asia represents the dominant geographical area where those events takes place (Froude & Petley, 2018). The distribution of landslides occurrence around the world (figure 1.2) and the associated fatalities emphasize the importance of Central Asia with annual deaths between 100 and 5000 in the period of 2007 to 2013 (Kirschbaum *et al.*, 2015).

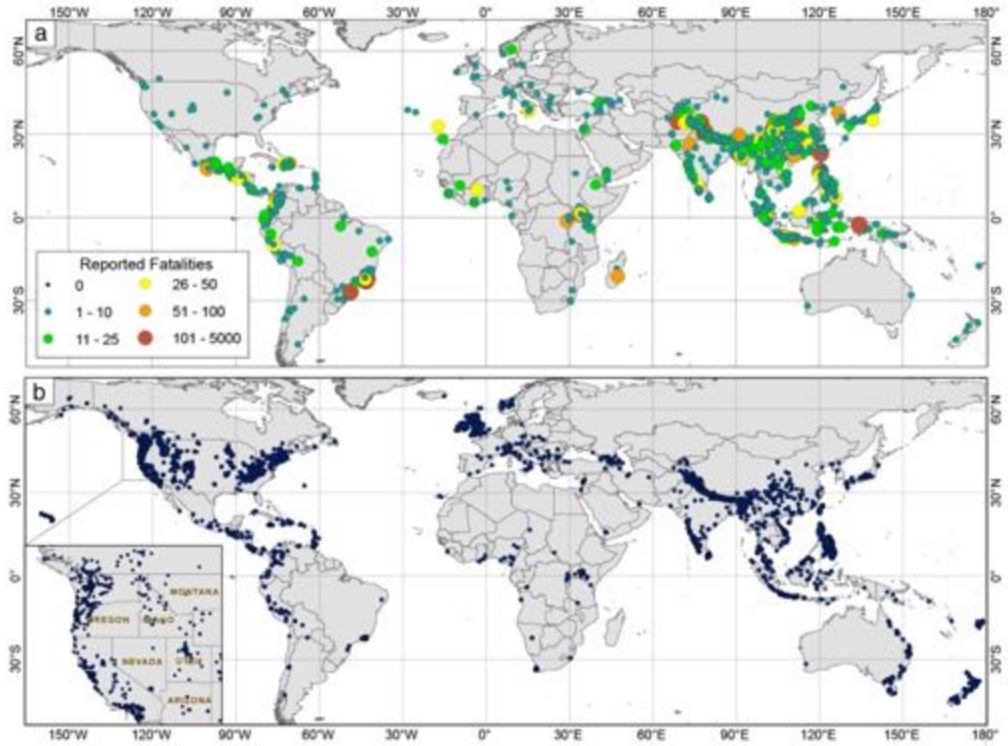


Figure 1.2: *Global map of reported landslide events from 2007-2013 in the GLC source: (Guha-Sapir* et al.*, 2017)*

On the other hand, earthquakes remain as important natural disasters in Asia (Guha-Sapir *et al.*, 2017). The continent is exposed to a high geophysical hazard (figure 1.3) due to its location in one of the most active tectonic areas of the world. Seismic activity has been reported in Central Asia since 1887 and 1910 when the city of Almaty in Kazakhstan was affected two magnitude 7-8 earthquake rupture. Countries like Kyrgyzstan, Tajik-

istan, Turkmenistan and Uzbekistan report an important number of earthquakes though time, some of them with magnitudes of seven to eight in the Richter scale (Thurman, 2011a).

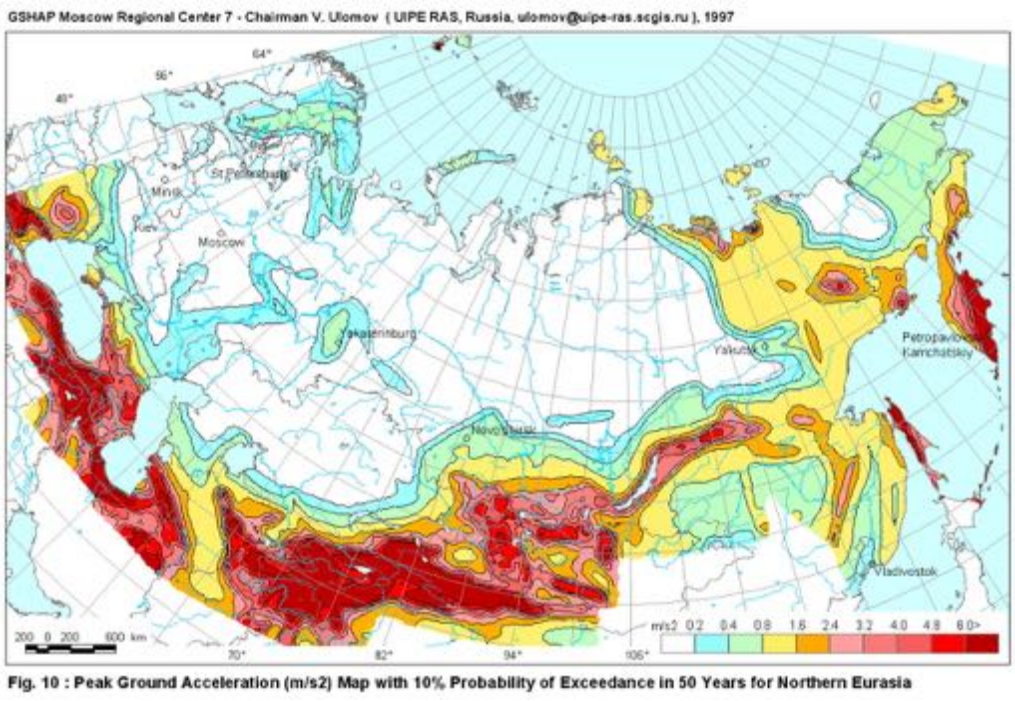


Figure 1.3: *Peak ground acceleration (m/s^2) map with 10% of exceedance in 50 years for Northern Eurasia source: (Ulomov, 1999)*

Secondary effects of earthquakes are also a relevant factors in Central Asia. Seismic events can directly trigger or accelerate other natural disasters like landslides, rockslides, mudflows, soil liquefaction and the rupture of glacial lakes and outburst flood (GLOFs). The most famous example of this type of interaction between natural hazards is the 1949 Khait earthquake in Tajikistan (Evans *et al.*, 2009). The Khait area is located near the southern limit of the Tien Shan Mountain and the northern limit of the Tajik Depression within the northern edge of the Pamir salient; which marks the active indentation of the India Plate into Eurasia. An earthquake occurred on July 10 and triggered many loess flows and rockslides in the area, affecting 7200 people located in more than 20 villages.

The Natural Disaster Risk in Central Asia (Thurman, 2011b) report that a significant portion of disasters that impacts the area could have been avoided if the vulnerability to natural hazard is reduced, however, in order to implement plans to decrease the degree of vulnerability, susceptibility maps should be created.

Creating such susceptibility maps require a deep knowledge of the factors associated to the type of natural disaster. Landslides can be triggered by a single factor like an extreme precipitation event or by a combination of factors like earthquake-triggered landslides, however, they are not spatially constant and change from one area to another. Global parameters like slopes, lithology, soil distribution, precipitation and peak ground acceleration are used to model global landslide susceptibility (Nadim *et al.*, 2006) whereas others parameters like fault presence, distance to streams or aspect could result

in a more relevant information in a regional or local scale.

In order identify areas which are most at risk in Central Asia, a landslide susceptibility map is created by the implementation of different statistically-based approaches. During the study some of the most used parameters are used among to geomorphological indices with the aim to characterize the landscape, to better understand the surface processes and their impact on mass movements.

1.1 Problem Statement

The Central Asia region consists of the former Soviet republics of Kazakhstan, Kyrgyzstan, Tajikistan, Turkmenistan and, Uzbekistan. It is an area with varied geographical domains like the Tien Shan and Pamir mountains as well as the deserts of Kyzyl and Taklamakan. Moreover, the weather in the region is affected not only by the Indian monsoons but also by the westerlies that influence the amount of precipitation inside the area (Pohl *et al.* , 2015). On the other hand, the geological setting is dominated by active tectonics that leads to an important number of earthquakes with different depths and magnitudes (Käßner *et al.* , 2016; Negredo *et al.* , 2007; Liu *et al.* , 2017; Ischuk *et al.* , 2013). Also, a variety of geomorphological environments like glacial process and high mountain erosional processes influence the landscape among calm sedimentary deposition in flatlands (Fuchs *et al.* , 2015).

Central Asia has been studied since the USSR time in order to characterize and determine the geological potential of the area in terms of mineral resources, consequently, detailed geological maps are available in paper-based and Russian language as well as an important number of publications related to the geodynamic setting of the area and geophysical analysis (Leith & Simpson, 1986; Burtman & Molnar, 1993). However, this is not the case for the others factors that are related to landslide triggering.

The variety in conditions of the area, limit the understanding of which factors influence more in the landslide triggering. The aim of this study is to understand the relationship between an important number of trigger factors and the landslide occurrence as well as to determine which of them places a major role in the landslide-triggering in the area.

1.2 Research Significance

Some studies have been performed in the area at country scale or more regional as a first attempt to understand the factors that trigger landslides (Gruber & Mergili, 2013; Saponaro *et al.* , 2015; Mergili & Schneider, 2011) using different approaches and databases. Previous studies attempt to understand the surface processes by the implementation of methodologies like risk indicator using GRASS GIS (Gruber & Mergili, 2013) or landslide factor analysis (Havenith *et al.* , 2015b). However, recent techniques based on statistical approaches are more frequently used and reported good and accurate results.

This study aims to assess the landslide susceptibility on a regional scale by the implementation of three different statically-based approaches: The Weight of evidence, logistic regression as one of the most used method, and random forest, a machine learning tech-

nique. Methodologies never implement for the study area.

Apart from the computation of landslide susceptibility maps based on new and more accurate techniques, the creation of different thematic information is a contribution to the knowledge of the area, known for the scarcity of information regarding to landslide triggering factors. Different geomorphological indices are computed with the aim to represent better the landscape and compensate the lack of information. Also, some of the geomorphological indices have never been used in the literature for the landslide susceptibility mapping. The aim of this work is thus to test their applicability and evaluate their effectiveness for modelling landslide susceptibility.

The study covers a large area, however, the landslide susceptibility assessment is performed in fine resolution (30m), forcing the implementation of a workflow that allows to work with big datasets. The developed workflow also rely exclusively on open source softwares such as Python and R for processing the data and Qgis for handling geographical informations. One of the aims is to achieve a workflow that can be replicated by populations in Central Asia, which often face limited software and computational means.

1.3 Research Objectives

The aim of this research is to determine the relationship between the mass movement and the possible trigger factors.The research objectives are listed below and a brief explanation of the methodology used to achieve them is presented.

- **To create a reliable landslide catalogue using remote sensing techniques and previous catalogues.**

  To achieve the research objectives, first data integration of previous landslide catalogues is implemented. To complete the missing areas, manual delimitation of using google earth imagery is done. A field work also allowed to check existing maps and identify new events.

- **To understand the interaction and spacial associations between different factors and mass movements.**

  The factors that influence slope stabilities the most are defined based on the results of the landslide susceptibility models. The approaches are implemented by the creation of different variables combinations - models-, and a sequential process is applied in order to find the combination of variables that leads to the best model.

- **To determine the relationship between the mass movements and the seismotectonic setting for the study area.**

  Different geomorphological indexes are computed in order to understand the different landscape features of the area. Some of the indexes reflects the response of the landscape to the tectonic uplift of the studied region. The relationship between

each of the variables and the landslide catalogue are explored by the analysis of spatial associations and landslide densities.

In order to complete the above-mention tasks, the following research questions should be answered:

- How different are the mass movements (size, frequency) along the area and how different are they from neighbouring regions?
- How do the frequency and size of mass movements correlate with geologically active mapped disturbances and seismicity?
- Are there any areas where mass movements occur but are aseismic or have low seismic activity?
- How do the mass movements correlate with geology (for example, the occurrence of evaporites on overlay fronts) and precipitation?
- What are the interactions /relations with the morphological parameters of the area (relief)?
- Along which tectonic disturbances do mass movements occur frequently?

Chapter 2

STUDY AREA

This chapter introduces the principal characteristics of the study area. First, the geological framework and climate, which tightly control natural hazard such as landslides are described . The second part of this chapter focusses on landslides. First the basic concepts about landslides are introduces. Then, a brief review of past events and previous works done in the area is provided. Finally, the landslide catalogue used in this study is introduced.

The area of study area is located in Central Asia between 66.2100N°to 71.9800N°and 36.5100E°to 39.9800E°. It covers parts of the countries of Tajikistan, Uzbekistan, Afghanistan, and Kyrgyzstan and has a total surface area of 188 316 km^2. Dushanbe is the only capital city located inside the study area; however, other populated regions like Urgut and Samarqand in Uzbekistan are included (figure 2.1).

The South Tien Shan and Western Pamir mountain ranges are the most relevant geographical features along with the Tadjik depression, where the capital city of Tajikistan is located. The landscape is characterised by a substantial variation in the topography as well as important altitudinal changes from 258 to 6049 m a.s.l.. The altitude increases rapidly from west to east ruling the changes in temperature, characteristic of the slopes and the depth of the valleys.

The Panj is the main river in western. It forms a deep valley which incises the Pamir Plateau and its different levels of alluvial terraces have been used to understand the evolution of the area concerning geology and geomorphology and the relationship between climate and tectonics (Fuchs *et al.* , 2015). Some landslide has dammed the river in the past creating problems to the communities and the infrastructure (Strom, 2010). The Vakhsh river flows in the valley between the Pamir and the Tien Shan. It is dammed for hydroelectric power generation north to the Nurek town, giving the name to the dam and it is also used for irrigation of local agricultural. The valley of the Vakhsh is also characterized by the occurrence of landslides that constantly endanger the infrastructure.

The Zarafshon (Zeravshan) is the main river in north.western. The source of the Zarafshon (Zeravshan) rivers is located near the highest point in the area located at 5621 m. a.s.l in Kyrgyzstan. The river flows in a East-West trend valley characterized by a a high presence of populated area. Also, a significant number of landslides are reported, affecting the populated areas located in it.

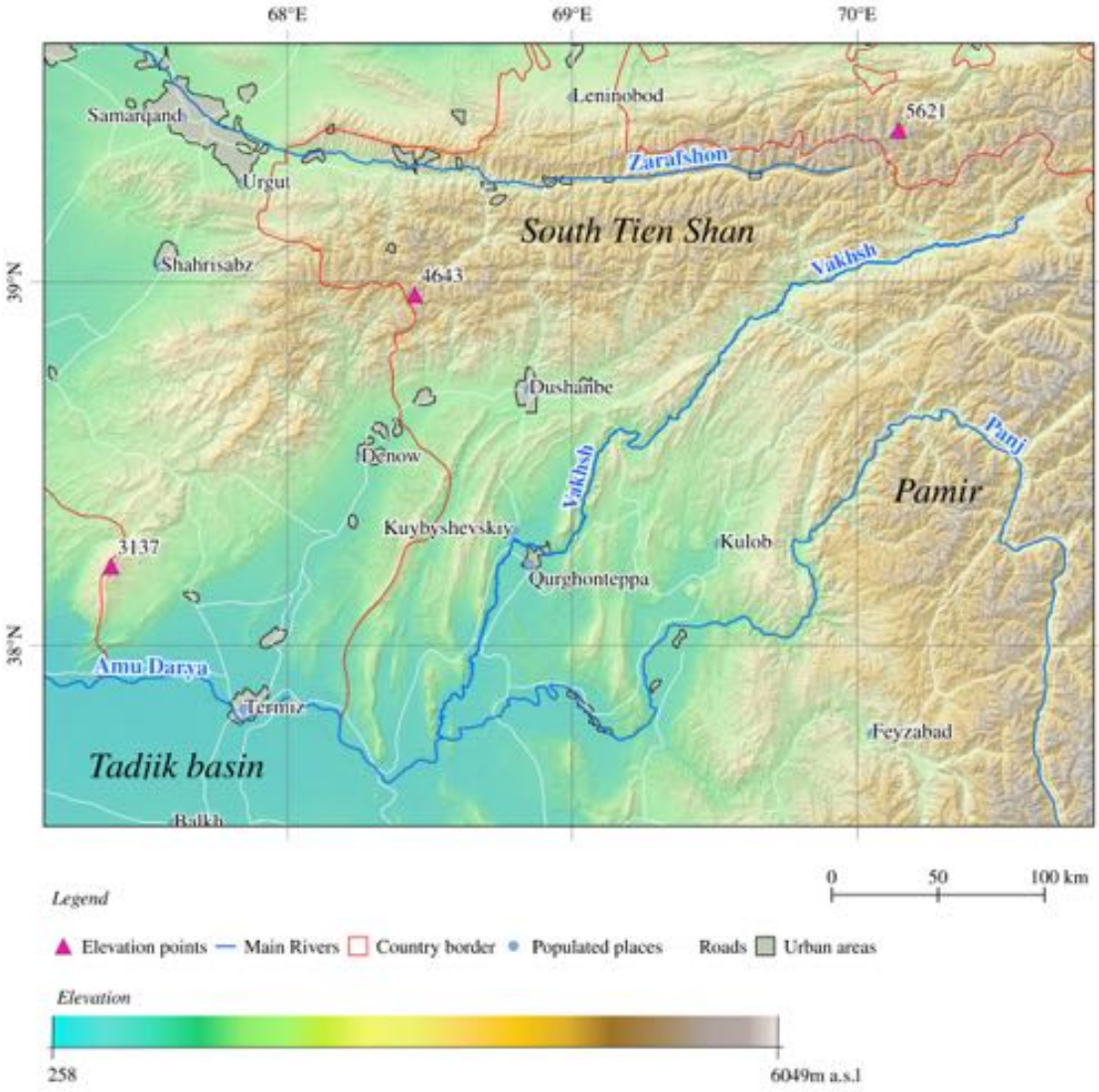


Figure 2.1: *Map of the location of the study area. Country borders are represented by red lines. Main populated and urban perimeters are presented in blue dot and gray polygons. Roads are denoted as white lines. Summits are marked by pink rectangles.*

2.1 Geology and Tectonics

The area is located at the northwestern end of the India-Asia collision zone. Consequently, different tectonic units or terranes characterize the area. The northern part consists of a reactivated Paleozoic orogen formed by a successive arc and micro-continental collision from Devonian to Triassic times. The South Tien Shan mountain (figure 2.2) consists of Paleozoic basement overlain unconformably by Jurassic to recent sedimentary units. Three Paleozoic sutures can be identified in the South Tien Shan, however, just two of them are present in the area. The Turkistan Suture is the northern suture in the area as well as the northern limit of the Turkistan-Alai complex characterized by an overthrust from the north by the oceanic and arc units of the South Fergana basin (Brookfield, 2000). The basement of the Turkistan-Alai units is unknown and alternating shales and quartz-sandstones with Ordovician to Silurian graptolites compose the unit. Then it

pass up into a dominantly carbonate section of Upper Silurian to Devonian and Lower to early Middle Carboniferous carbonates (Rogozhin, 1993). The transition to the Zarafshan unit is marked by a highly deformed and non-continuous ophiolitic belt (The Zirabulak unit) and coincides roughly with the Zeravshan fault zone (Kurenkov & Aristov, 1995).

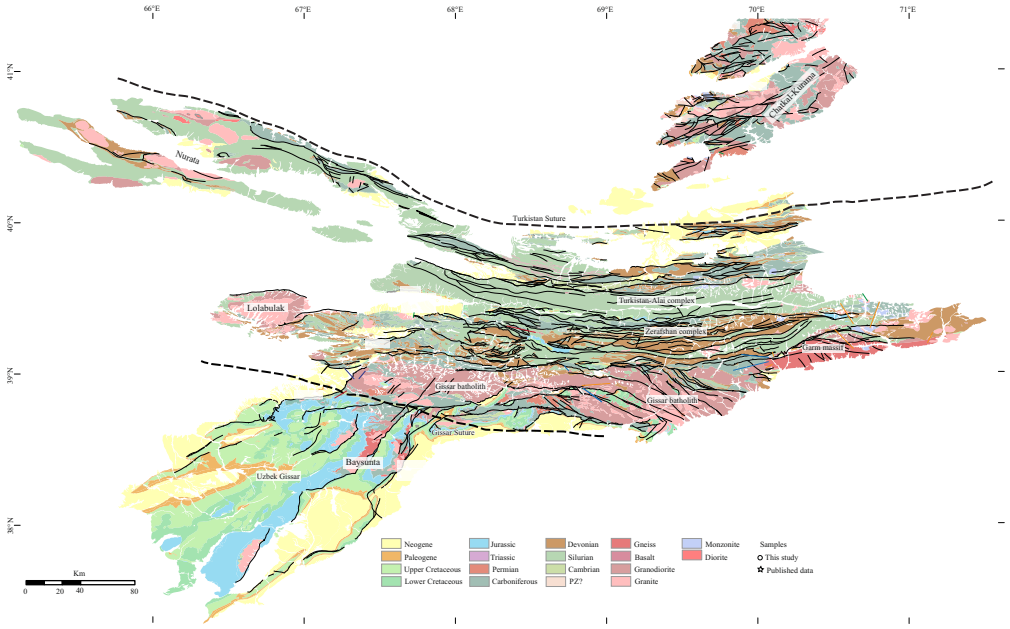


Figure 2.2: *Map of the regional geology of the Tien Shan. The sedimentary units are classify by age, while the igneous rocks are grouped by the type of rock. source: Ratschbacher (2018) personal communication.*

The Zarafshan complex consists of turbidites interbedded with sedimentary bloc melanges derived from the Turkestan-Alai zone to the north and thin, mature Cambrian to Ordovician passive margin clastic materials, overlain by thick Silurian turbidite and thick Lower to Middle Devonian carbonates (Brookfield, 2000).

Huge late Paleozoic batholiths dominate the Gissar unit. They present Carboniferous (Khasanov, 1975) and Permian (Baratov, 1966) cooling ages and represent marked differences in morphology, composition and texture. It intrudes thin Ordovician to Devonian sections that consist of mature deep shelf or slope turbidites passing up into massive shelf limestones (Tulyaganov in (Brookfield, 2000)) as well as the unconformably overlain Carboniferous volcanic rocks (andesite and dacite) and the associated marine volcaniclastics and thin limestones. The Carboniferous sequence continues with coarse non-marine clastic rocks affected by intense folding at the end of the Middle Carboniferous, accompanied by numerous granitoid intrusions and regional metamorphism ((Baratov, 1966)).The South Gissar unit or Gissar suture is an ophiolitic suture dominated by greenschist metamorphics, ophiolites and melanges .This suture contains only Carboniferous and younger oceanic material and is too young concerning the Late Carboniferous collision within the Tien Shan to be anything other than a marginal basin or narrow rift basin (Brookfield, 2000).

The southernmost unit in the Tien Shan is the Baysunta unit; consist of a Protozoic metamorphic core (metapelites) unconformably overlain by Lower Carboniferous continental conglomerates, sandstones, acid volcanics (quartz porphyries and dacites) and tuffs, with some marine limestones in the middle, overlain by Upper Carboniferous marine sediments that begins with submarine spilitic basalts, which underlie a coarsening upwards section of interbedded sandstones, siltstone, and shales with occasional conglomerates and limestones. These are overlain by latest Carboniferous conglomerates with limestones lenses and Upper Permian non-marine conglomerates (Brookfield, 2000).The Gram Massif is located in the eastern part of the Tien Shan and represents the only exposed Precambrian continental metamorphosed crust.

On the other hand, the Eastern part of the area is characterized by a series of sutures, magmatic belts and crustal blocks accreted to the Eurasian plate during the Paleozoic to Mesozoic times (Burtman & Molnar, 1993) that correspond to the evolution of the Pamir mountain belt (figure 2.3). The main tectonic sutures in the Pamir, separate three distinct terranes: The Northern, Central and Southern Pamir, although, in the area, just the Northern Pamir is present. These terrane represent the Paleozoic suture zone between the Central Pamir and the rest of Asia, and consists of oceanic Carboniferous igneous and sedimentary rocks (ophiolites) presented as a section of tholeiitic basal tectonically overlaid by gabbro and ultrabasic rock; or as melanges of serpentinite overlain by pillow basalt of tholeiitic composition (Darvaz Range). The upper part of the sections contains limestones, island-arc volcanic rocks or terrigenous sediments depending on the location of the section. Carboniferous and Permian conglomerate and limestones overlie these ophiolites. These group of rock as well as the late Paleozoic mafic and ultramafic metamorphosed volcanic rocks located in the south of the terrane, indicate the existence of a Carboniferous ocean basin denominated the Akzhilga zone. On the other hand, metamorphosed Precambrian and Paleozoic rocks are located in the south of the unit and they are interpreted as continental fragments that collided with Asia probably during the Permian time after the closure of the Akzhilga ocean basin (Burtman & Molnar, 1993).

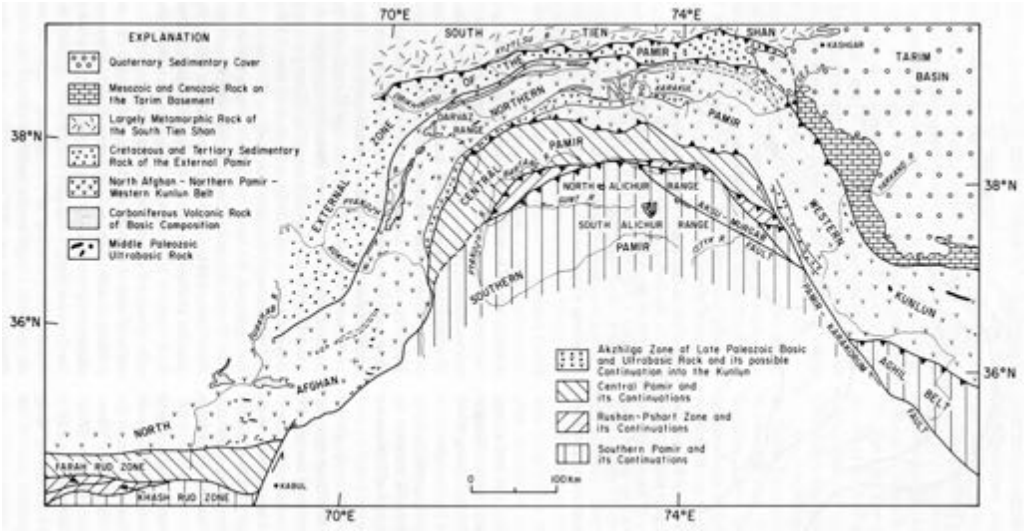


Figure 2.3: *Map of the simplified geology of the Parmir. The lithological information is divided by geodynamic units. Modify from: Burtman & Molnar (1993).*

The territory between Pamir and Tien Shan, is represented by the Tadjik Depression. It was an area of marine and continental sedimentation in the Cretaceous and Paleogene that suffered a disruption of the sedimentation pattern facies during the Late Cenozoic a product of crustal shortening during the convergence between the Pamir and the Tien Shan that resulted in an overthrust of the Pamir massif onto the Tien Shan during the India and Eurasia collision (Burtman & Molnar, 1993). The Early Cretaceous facies corresponds to a red-coloured clastic fluviatile sedimentary rocks covered by strata of intercalating marine and non-marine deposits, while the Late Cretaceous is characterized by a marine environment. The Paleogene sequence starts with marine deposits followed by an early Oligocene marine transgression (Burtman, 2000) and apparently the sedimentary basin was continuous from the Tadjik Depression in the west across the Alai region to the Tarim Basin to the east (Davidzon et al., 1982). The Neogene deposits are siltstones, sandstones and mainly gravel to boulder conglomerates that include large brecciated carbonate blocks interpreted as rock-avalanches deposits. Molasse deposits are also located in the front of the main Pamir thrust system (MPT) (Arrowsmith & Strecker, 1999).

Finally, the quaternary deposits are mainly coarse pediment-cover gravels, glaciall terraces and alluvial-fan gravels along with glacial tills and landslide deposits. Arrowsmith & Strecker (1999) used those deposits to constrain the spatial and temporal distribution of the late Quaternary deformation along the Trans Alai Range front in the Alai valley.

The area is characterized by an important number of active faults (figure 2.4). Normal faulting is reported by Schurr *et al.* (2014) along the western edge of the Pamir mountains. On the other hand, left-slip faults systems are mapped in the south of the area following two different predominant directions. The Badakhshan Fault zone is composed by 10 different traces intersecting each other following a general N-S orientation; in contrast, The Alburz-Marmul Fault zone located at the south of the Tadjik depression presents a predominant E-W orientation.

The Gunt Shear Zone (dextral strike slip) (Schurr *et al.* , 2014), as well as the Henjvan Fault are located at the Pamir. They are reported as strike-slip faults by Schurr *et al.* (2014); Ruleman *et al.* (2007). Some landslides area located in the north part of the Henjavan Fault. Additionally, an important number of thrust are located in the area; they are considered as the main structures related to the high seismicity because of the complex subduction interaction among the South Tien Shan, Pamir and Tadjik depression. The predominant tendency of these structures is E-W to SW-NE and an important vertical displacement of basement rocks reflects its activity; for example the Illiac fault shows as much as 3km vertical displacement (Leith & Simpson, 1986). Others thrust in the area are represented by the Tanymas, Sarez-Murghab and Pamir Thrust systems - Vakhsh thrust System being the northernmost one of the most important in the area as well as the Darvaz Fault Zone (sinistral transpressive)

The neotectonic activity is dominated by the northward propagation of the Indian plate inducing east-west striking mountain ranges. Crustal shortening is mainly accommodated at the MPT by subduction beneath the frontal part of the orogen where most of the seismicity occurs. The lateral margins of the orocline display strike-slip motion of -12 mmyr-1 along the western Darvaz Fault Zone (DFZ) (Mohadjer *et al.* , 2010).

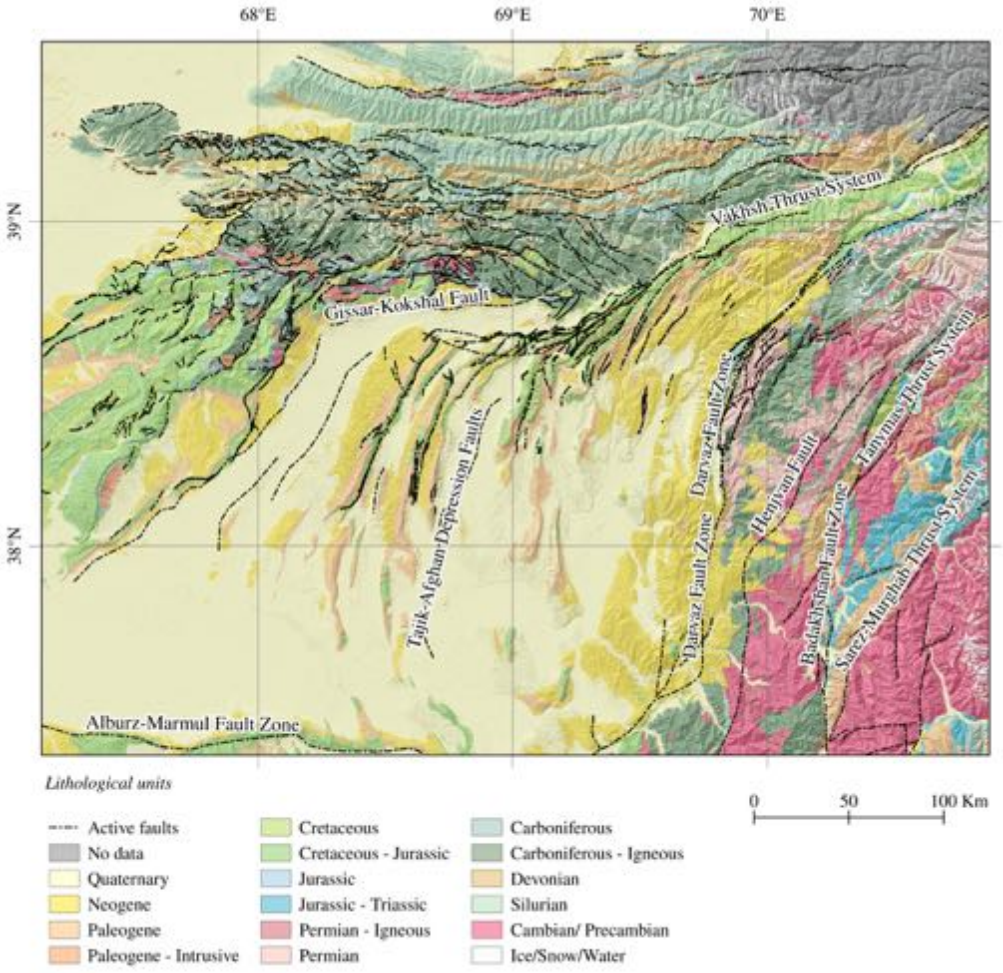


Figure 2.4: *Map of the geology of the study area divided into 16 units. The division is made based on the age and the type of rock. 1. Quaternary, 2. Neogene, 3. Paleogene, 4. Paleogene-Intrusive, 5. Cretaceous, 6. Cretaceous-Jurassic, 7. Jurassic, 8. Jurassic-Triassic, 9. Permian-Igneous, 10. Permian, 11. Carboniferous, 12. Carboniferous-Igneous, 13. Devonian, 14. Silurian, 15. Cambrian/Precambrian, 16. Glacier areas. Active faults represented by dotted lines.*

The seismicity in the northern margin of Pamir that interacts with the Tien Shan is characterized by shallow-intermediate earthquakes. Fault plane solutions are determined suggesting a large component of thrust faulting and roughly north-south crustal shortening. The plane solutions suggest southward under thrusting of the Ferghana Basin beneath the South Tien Shan (Burtman & Molnar, 1993).

On the other hand, the Nurek reservoir area presents superficial earthquakes, mostly between 2 to 8 km, while deeper seismicity (> 20 km) is identified for the rest of the area (Leith & Simpson, 1986). Schurr *et al.* (2014) suggest that the seismicity in the Pamirs is spatially partitioned and occurs along well-defined zones. The Pamir's northern perimeter is outlined by a band of earthquakes; whereas its eastern and western flanks exhibit only sparse and more diffuse seismicity. On the other hand, the south margin of the Pamir that interacts with the Hind Kush is marked by intense intermediate depth seismicity

(90-250 km). This area has been interpreted as a continental subduction by (Burtman & Molnar, 1993) and supported by seismic tomography (Negredo *et al.* , 2007).

2.2 Climate

The climate in the area is temperate, semi-arid to arid and continental with a hot summers and cold winters. The position of the area coincides with the transition between the atmospheric circulation systems of the Indian Summer Monsoon (ISM) and the Westerlies, implying a particular climatic setting in terms of erosion, rainfall patterns and temperatures because the region is highly sensitive to variation in atmospheric circulation patterns. From a global scale analysis, Aizen *et al.* (2001) concluded that the area is weakly influenced by the Siberian anticyclonic circulation and moderately influenced by a southwest cyclonic circulation that brings warm moist air masses into the region. The moist air masses increases the precipitation falls during winter, however, the maximum of precipitation occurs during spring (figure 2.5), while the second maximum precipitation occurs in autumn. Contrasting, the formation of a thermal low during summer causes a decrease in precipitation in August and September.

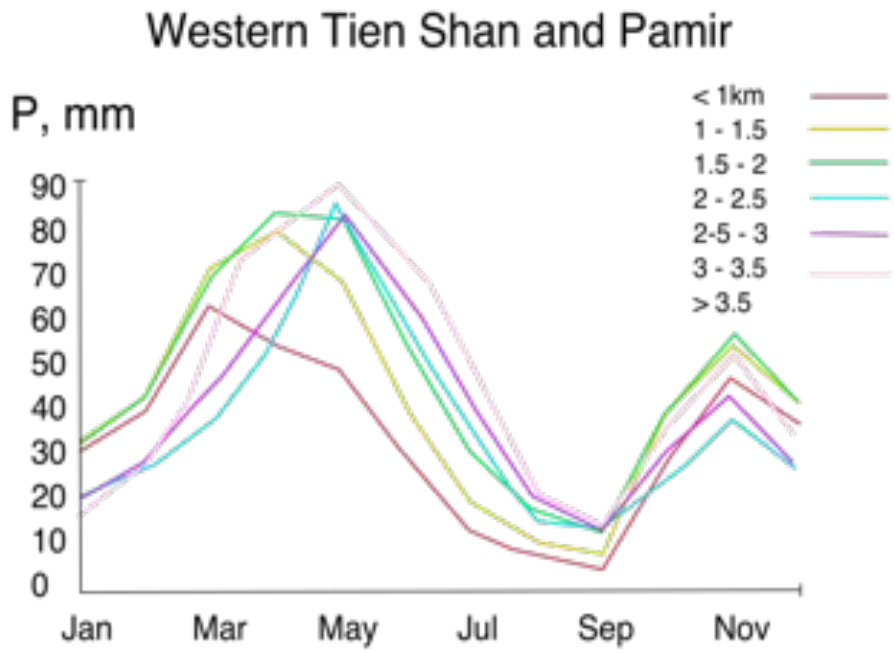


Figure 2.5: *Annual variation in precipitation in altitudinal zones. Data source: Reference book of Climate USSR, Kirgiz SSR (1988) source: Aizen* et al. *(2001).*

The seasonal variation of the precipitation in the area is studied by (Pohl *et al.* , 2015) based on the analysis of harmonic time series (HANTS) using the High Asia Refined analysis project (HAR) database. The first general conclusion of the study is that there are strong differences in the average winter and summer precipitation distribution in the Pamir area. The area is divided into main orographic barriers because they would intercept the moisture supply among others. The Western Pamir and the Alay mountains (South Tien Shan) are characterize by a significant amount of precipitation in winter,

while the Western Pamir receive almost no precipitation in summer (figure 2.6).

An important number of glaciers are located in this area, however, many studies report there their retreat, favoring the development of lakes in the glacier forefields or in subsiding areas, increasing the probability of mass wasting or glacier lake outburst flood (GLOF) events in the surroundings. Mergili & Schneider (2011) identify 6 very hazardous lakes and 34 hazardous lakes in the south-western Pamir, some of them dammed by landslide deposits or older moraines.

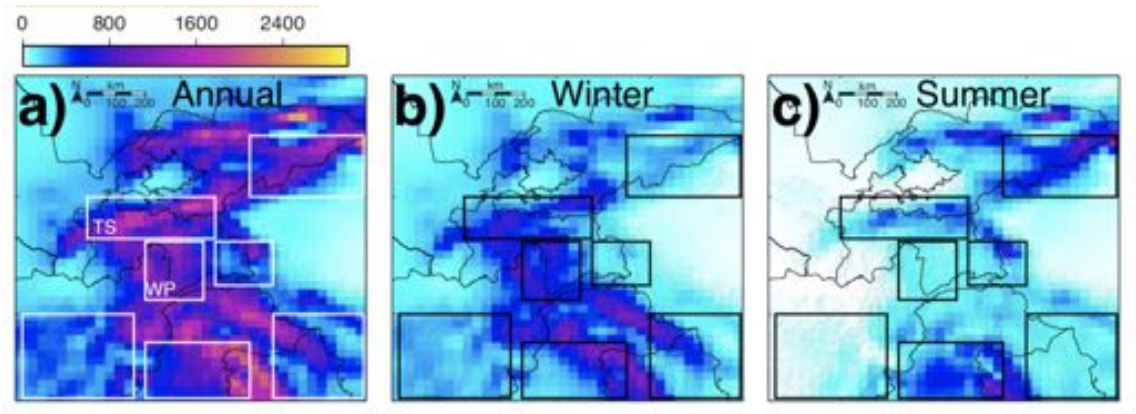


Figure 2.6: *Precipitation distribution of the applied HANTS method for the annual, winter and summer precipitation patterns. TS: Tien Shan, WP: Western Pamir source: Pohl* et al. *(2015).*

2.3 Landslides

2.3.1 *Landslide definition and classification*

The term **landslide** is defined as a general term to describe the downslope movement of soil, rock, and organic materials under the effect of gravity and also the landform that results from such movement (Highland *et al.*, 2008). This term refers to several different things from mudflows to rock avalanches and it is broadly used as a non-technical word to describe any or all relatively rapid forms of **mass wasting** (Tarbuck *et al.*, 2014). **Mass wasting** is a surface process that must not be confused or related to erosional processes because mass wasting does not require a transporting medium such a water, wind or glacial ice (Tarbuck *et al.*, 2014).

Specific terms are used to describe the morphology of a landslide (figure 2.7). The *crown* of a landslide is the undisplaced material still in place and adjacent to the highest part of the *main scarp*; a steep surface on the undisturbed ground caused by the movement of the material away from the undisturbed ground. In addition, these areas are characterized by the presence of *crown cracks* parallel to the crown, product of an instability in the terrain. Similarly, the *head* of the landslide is the upper part of material located along the contact between the displaced material and the main scarp and can or not, be limited by *minor scarps* that divided the moved material, produced by differential movement within the displacement.

On the other hand, the *toe* of a landslide is usually the curved margin of the displaced material and it is the most distant part from the main scarp. This area present *transverse cracks* formed for the compression of the material at the *toe of surface rupture* as well as *radial cracks* near to the end of the toe as a product of the material divergence (Highland *et al.*, 2008).

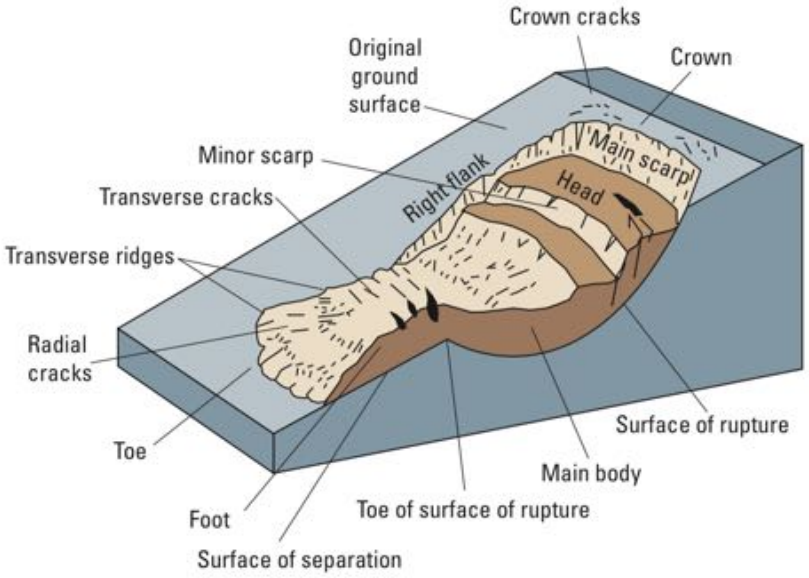


Figure 2.7: *Illustration of the most commonly used labels for the parts of a landslide. source: Varnes (1978) in Highland* et al. *(2008).*

The *surface of rupture* is the lower boundary of the displaced material below the original ground surface and the *toe of surface of rupture*. The toe of the surface of rupture is the intersection (usually buried) between the lower part of the surface of rupture and the original ground surface. The morphology of the surface rupture determines the type of mass wasting and it is related to the material that is moved.

The displaced material that overlies the surface rupture, between the main scarp and the toe of the surface rupture is the *main body*. The area where the main body is located is denoted as the *depletion zone* as the material lies below the original ground surface. In contrast, the *accumulation zone* is where the material lies above the original ground surface. The material that overlies the original ground surface and is located beyond the toe of surface of rupture is denoted as the *foot* of the landslide (Highland *et al.*, 2008).

Landslides are classified based on the type of movements on the surface of rupture and the type of material moved. The most accepted classification was proposed by Varnes (1958). New classifications are present (Hungr *et al.*, 2014) based on the Varne's proposal (figure 2.8), but only bring minor changes.

The first type of movement or landslide type is **Falls**. This is the simplest type and begins with the detachment of soil or rock from a steep slope along a surface with little or no shear displacement. The material subsequently descends mainly by falling, bounc-

ing, or rolling. Additionally, Highland *et al.* (2008), propose a variation of the rockfall denominate **topple** that consist of the rotation, out of a slope, of a mass of soil or rock around a point or axis below the center of gravity of the displaced mass (figure 2.9).

TYPE OF MOVEMENT	TYPE OF MATERIAL			
	BEDROCK		SOILS	
FALLS	ROCKFALL		SOILFALL	
FEW UNITS SLIDES	ROTATIONAL SLUMP	PLANAR BLOCK GLIDE	PLANAR BLOCK GLIDE	ROTATIONAL BLOCK SLUMP
MANY UNITS		ROCKSLIDE	DEBRIS SLIDE	FAILURE BY LATERAL SPREADING
	ALL UNCONSOLIDATED			
	ROCK FRAGMENTS	SAND OR SILT	MIXED	MOSTLY PLASTIC
DRY FLOWS	ROCK FRAGMENT FLOW	SAND RUN	LOESS FLOW	
			RAPID EARTHFLOW	DEBRIS AVALANCHE / SLOW EARTHFLOW
WET		SAND OR SILT FLOW	DEBRIS FLOW	MUDFLOW
COMPLEX	COMBINATIONS OF MATERIALS OR TYPE OF MOVEMENT			

Figure 2.8: *Classification of landslides based on the type of movement and the type of material. source: Varnes (1958).*

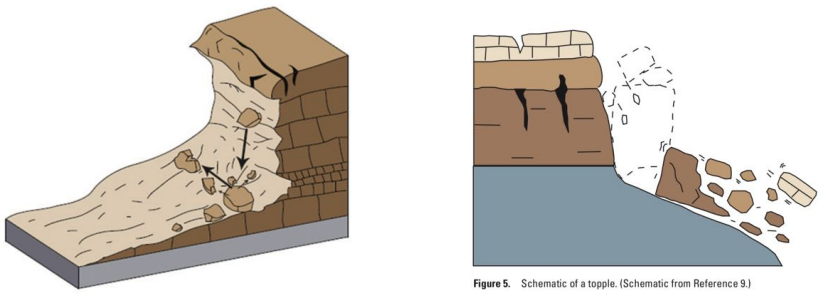


Figure 2.9: *Ilustration of the different types of rockfalls. Left: Illustration based on Varnes (1958). Right: Illustration based on Highland* et al. *(2008). source: Highland* et al. *(2008).*

The second type of mass-wasting is denoted as a **slide**. It is described as a material that moves as a discrete block. In this type of movement, a distinct zone of weakness separates the slide material from the more stable underlying material. Two basic types of slides are recognized based on the morphology of the surface of rupture: *rotational slide* where the surface of rupture is a concave-upward (spoon-shaped). The descending material exhibits a rotational movement and *translational slide* where the material moves along a relatively flat surface such as a joint, fault or bedding plane Highland *et al.* (2008) (figure 2.10).

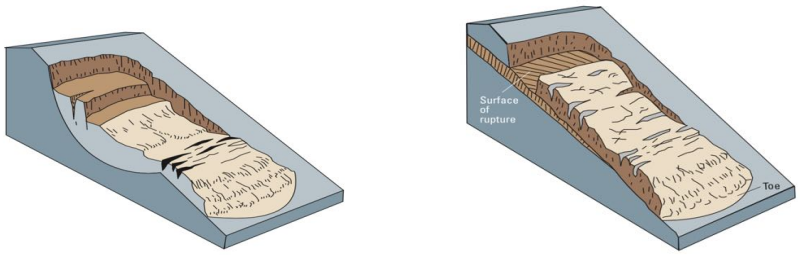


Figure 2.10: *Illustration of the different types of slides. Left: Illustration of a rotation slide. Right: Illustration of a translational slide. source:* Highland et al. (2008).

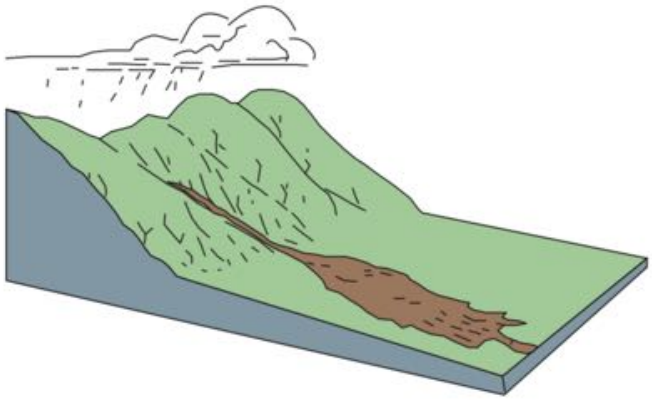


Figure 2.11: *Illustration of a flow. source:* Highland et al. (2008).

And the third is described as a spatial continuous movement in which the surface of shear is short-lived, closely spaced and usually not preserved (figure 2.11). The materials in a **flow** behave as a viscous liquid that flows downslope. The most studied flows are debris flows, that occurs due to rapid melting of snow in semi-arid mountain areas, where a large quantity of soil and regolith are washed into nearby stream channels because of the lack of vegetation and the particle size. In contrast, mudflow is most often formed on hillsides in humid areas during the time of heavy precipitation or snowmelt. Most of the big magnitude landslides cannot be classified under a single type as because a combination of movements, materials, and triggers are related. Those landslides are classified as *complex* because of the combination of multiple landslide types.

2.3.2 Historical Landslides

In the area of study, some historical landslides are reported and mapped based on their magnitude and the impact they produced in the landscape and the population. One of the most famous events was related to the Kait Earthquake. The main Khait earthquake occurred on July 10, 1949, with a M7.4 (magnitude calculated using surface waves). The exact location of the epicenter is uncertain, but three different locations were reported, all three within 8.5 km from Khait. The focal depth has been estimated at between 16 and 20 km by Rautian & Leith (2002) and many loess flows and rockslides were widespread in the epicentral area as well as cracks in rock slopes. In the Yaman valley, hundreds of loess landslides coalesced to form a massive loess flow with an estimated volume of 245M m^3 that traveled up to 20km on a slope of only 2° and killed an approximately 4000

people located in 20 villages (*kishlaks*) (Evans *et al.*, 2009). In the adjacent valley, the Kait landslide (rockslide) (figure 2.12) was transformed into a very rapid flow (̃30 m/s) by the entrainment of saturated loess into its movement. Evans *et al.* (2009) performed simulation in the Khait landslide and estimated a volume of 75 M m^3 and ca. 800 causalities. A total of approximately 7200 people were killed by earthquake-triggered landslides in the epicentral region.

Figure 2.12: *View of the Khait rockslide trigerred by the 1949 M7,4 Khait earthquake. The scar on Chokhran mountain shows the origin of the slide, which overwhelmed the village of Khait source: (Evans* et al. *, 2009).*

The Gissar earthquake occurred on January 23, 1989, South of Dushanbe, capital of Tajikistan. It triggered a series of earth-floss in loess and killed an least 200 people and buried hundreds of houses. The mass wasting was related to extensive liquefaction, which had developed from a horizontal acceleration of about 0.15g (Ishihara *et al.*, 1990). The largest landslide, called "Okuli", had an estimate volume of $20 * 10^6 m^3$. It is the results of two independent slides triggered in the north, which then merge into the main stream of the mudflow. The scarps of many of the landslides are located along a water channel installed on the shoulder of the hills, with the sliding surface located at a depth of about 15m within the saturated part of the 30m thick loess deposit. Ishihara *et al.* (1990) assumed that the the mass wasting was triggered during the earthquake, but the materials had been saturated over many years by the stream. Theory supported by observation of muddy water oozing from the earth-flow (Havenith & Bourdeau, 2010).

On the other hand, the effects of the landslides on the area are not only limited to the damage caused by the mass displacement; but, frequently, landslide dams area created (Strom, 2010). In the Panj River valley, it is possible to find partially or complete eroded rockslide dams with a volume higher than $1km^3$. An example is the Shids rockslide dam formed during prehistoric times (figure 2.13); it is composed of Proterozoic granite and gneiss by now almost completely incised by the Panj river, however, 20-30 m high dam still exist and a remnant lake extends about 15 km upstream (Strom, 2010).

Another landslide dam in the area is the Shiva lake. It is located in the Afghan part of the Pamirs (figure 2.14).The lake Shiva is an evidence of the impact of the large landslides in the landscape. It is located 16 km from the south-west of Khorog along the valley of Arakht in the border between Afghanistan and Tajikistan. The natural dam is 1.6 km wide. It is composed by material of at least three landslides and many rock glacier. The dam has never been overtopped because the water supplied by the catchment travelled through the dam's permeable material forming a stream (figure 2.14).The lake is determine as a no immediate hazard area, but a partial collapse of the dam due to retrogressive and piping erosion cannot be excluded (Schenider *et al.* , 2013).

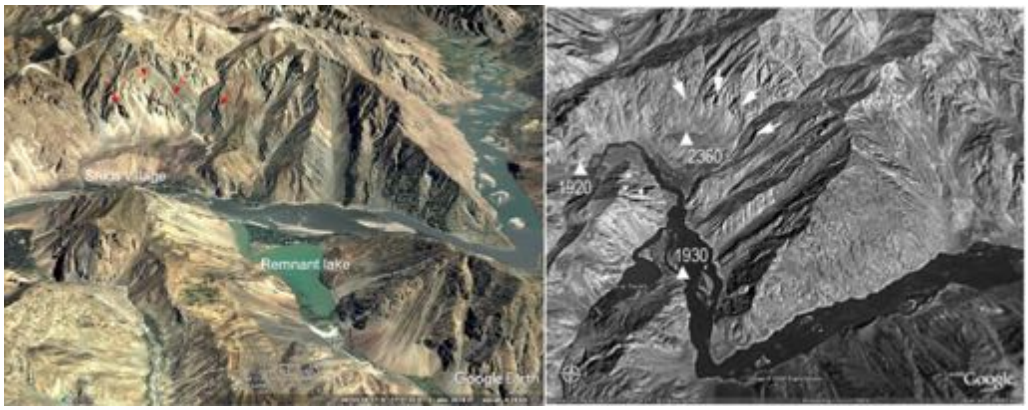


Figure 2.13: *Google earth panoramic view of the Shids rockslide and remnant lake in the Panj River source: (Strom, 2010)*

Similarly, the Tien Shan mountains are also highly affected by bedrock landslides of which many of them caused river's damming like the Iskandrkul, the largest landslide-dammed water body in the area. The Iskanderkul-Daria river is the source of the lake. The rockslide of nearly 1 km^3 is related to the collapse of a mountain slope composed of Paleozoic sedimentary rocks. The actual river erodes it and an up to 50-70 m deep gorge formed in the landslide body with an impressive waterfall in its central part. The possibility of a breach is plausible because the dam undergoes intensive backward erosion. Others small lakes had been formed in the valley, but they had been partially filled, and, finally, drained (Strom, 2010).

The Yashinkul lake is located in the north-east corner of the study area. It is a rockslide about $50 * 10^6 m^3$ in volume, that had converted into rock avalanche and filled the valley of the Alichur River, the source of the Gunt River. The dam is characterized by dual a structure with semi-rounded moraine-like boulders inside, overlaid by angular gneiss blocks and debris. An explanation for this structure is that the lake was firstly dammed by the end moraine, later cover by a rockslide, however, (Strom, 2010) consider the whole dam as a pure rockslide. The dam is stable and the outflow is artificially controled to increase power production of the hydraulic power plants at the lower part of the Gunt Valley (figure 2.15).

A recent event is the Aini blockage occurred in 1964 (Strom, 2010), when 20 million cubic meters of debris blocked the Zeravshan valley and created a dam up to 150 m high and 1 km long the stream, however, the situation was controlled and become one of the

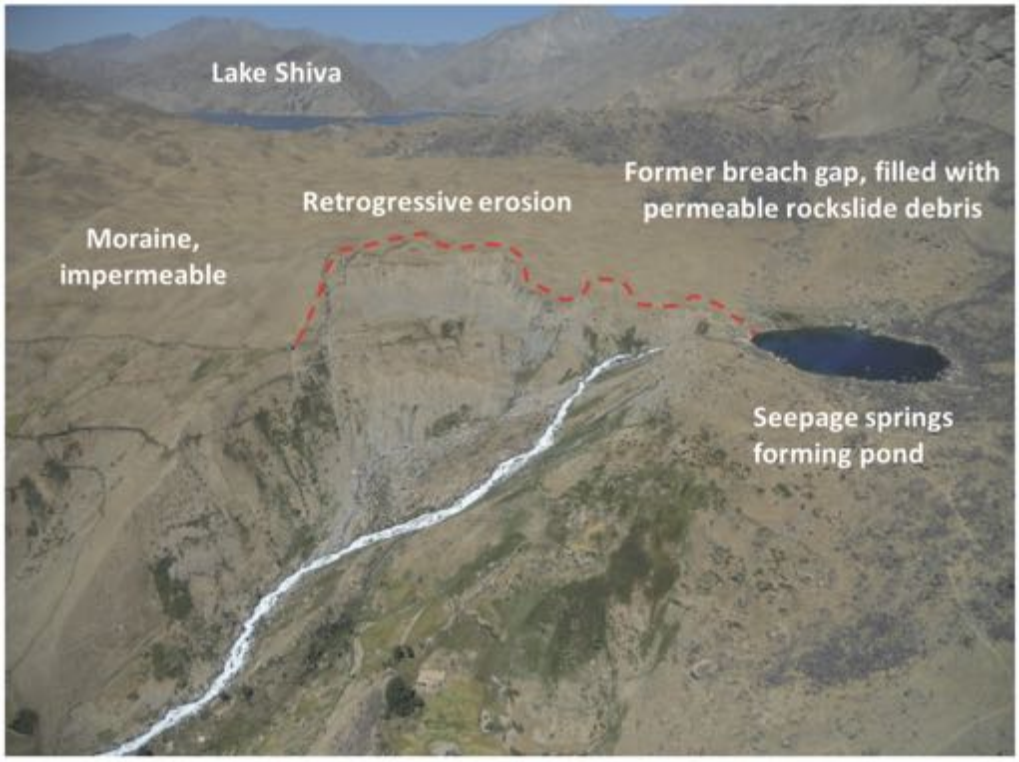


Figure 2.14: *Lake Shiva. The depression spring (circular lake of approximate 200m) and headwater of Arakht torrent in the foreground. To the left, the lake Shiva. source: Schenider* et al. *(2013).*

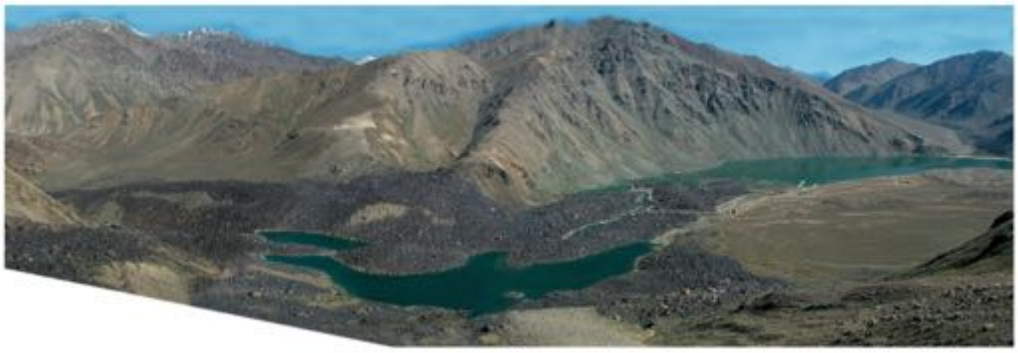


Figure 2.15: *Overview of the Yashilkul rockslide dam before the spillway construction source: Strom (2010).*

first examples of the successful prevention of the rock-slide dam breaching disaster .

2.3.3 Previous works

Previous regional scale analysis has been done in the area in order to identify multi-hazards and risk indicator; Gruber & Mergili (2013) used GRASS GIS to implement a model framework that includes high-mountain processes like rock slides, ice avalanches, periglacial debris flows and lake outburst floods in an area of 98 300 km^2 in the Pamir (Tajikistan - figure 2.16). The objective of the model framework is to help gain an idea

about possible hazards and risks based on the analysis of parameters like elevation, glaciers location, lakes information (lake type, lake drainage, calving of ice, lake area, lake evolution), mean annual air temperature, permafrost susceptibility, seismic susceptibility, exposure and communities location.

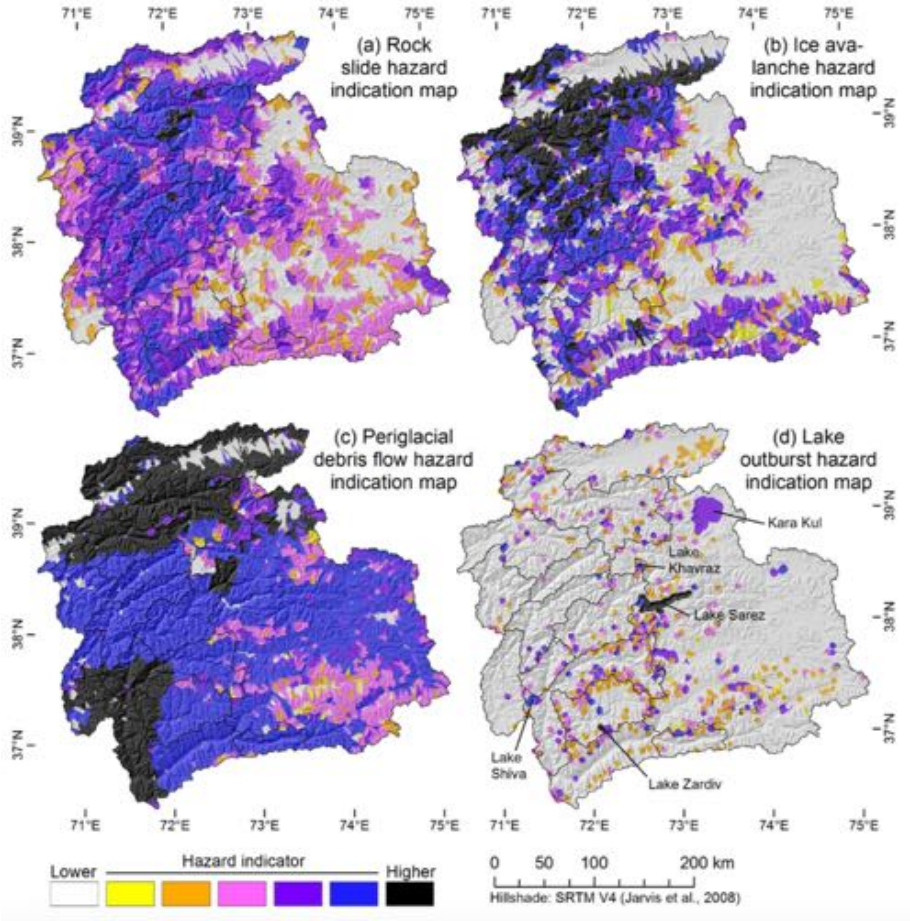


Figure 2.16: *Distribution of the hazard indicator for the analyzed high-mountain processes. Approach : Model framework. source: Gruber & Mergili (2013).*

The results of the study show that the periglacial debris flow susceptibility/hazard is the most common in the area studied by Gruber & Mergili (2013) (figure 2.17), in contrast to the ice avalanch and lake outburst susceptibility, due to their confinement to glaciers and lakes respectively. On the other hand, rock slide susceptibility displays intermediate patterns in terms of the total area because it is associated with limited locations occupied by very steep slopes Gruber & Mergili (2013).

Similarly, the Earthquake Model Central Asia (EMCA) project (Saponaro *et al.* , 2015) made a contribution to the landslides susceptibility mapping in Central Asia by the computing of a weight of evidence model using seismic intensity as a trigger mechanism in order to increase the understanding about the role of the earthquakes as triggering factors (figure 2.17). The seismic intensity is expressed through the observed macro-seismic intensity (MSK 64). The territories of Kyrgyzstan and Tajikistan area characterized by returned intensities of VIII and IX which are expected in the future, while for Uzbekistan

an intensity of VIII is foretelled. The model takes into account parameters like slope gradient, slope aspect, profile curvature, geology, distance from faults and seismic intensity.

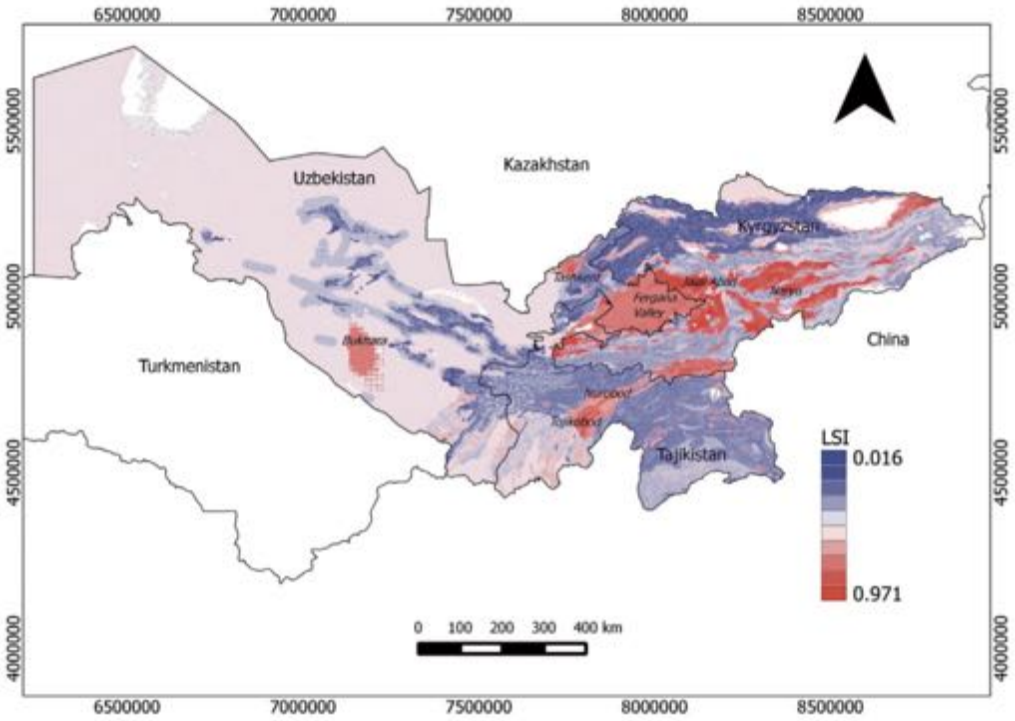


Figure 2.17: *Landslide susceptibility index (LSI) map for Kyrgyzstan, Tajikistan and Uzbekistan. Model accuracy level greater than 70%. Approach: Weight of evidence. source:* Saponaro et al. (*2015*).

The most recent study in the area was performed by Havenith *et al.* (2015b). The study used the Tien Shan geohazard database collected by the author, that includes a more complete landslide catalogue as well as an earthquake catalogue. The study assigned landslide susceptibility based on the landslide factor analysis using morphological parameters, geological information, river distance, precipitation, earthquake and fault distance as thematic variables and two different models were computed. The first one includes four factors (morphological, geological, river distance and precipitation), while the second one is a combination of the first model plus the seismo-tectonic influence (figure 2.18). The performance of the model is statistically analyzed based on the scarp and landslide densities obtained, however, it cannot be considered as a prediction model. The resulting landslides susceptibility assessment reproduces well the trend of the observed landslides activity taking into count the large extent of the area that almost covers an entire mountain range.

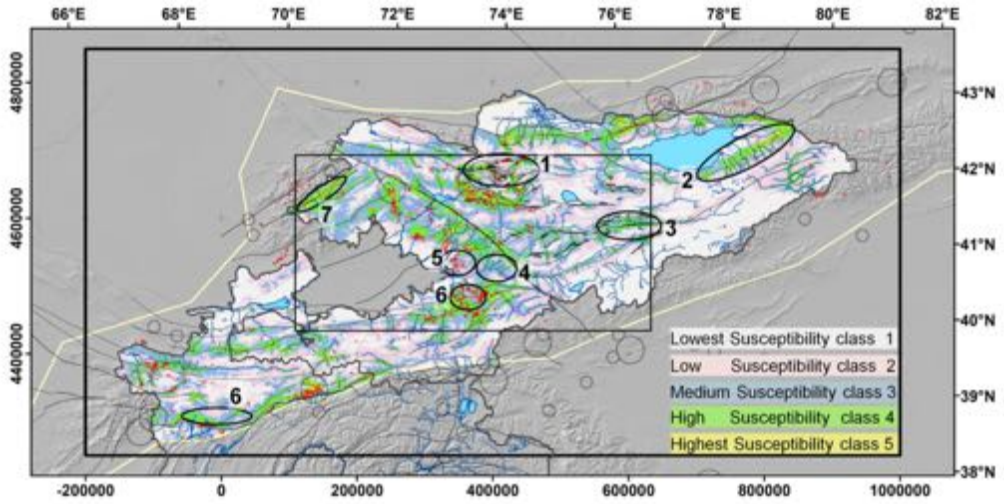


Figure 2.18: *Landslide susceptibility map considering morphological, geological, hydrological, climatic and seismo-tectonic parameters. Landslides are outlined in red. Black ellipses outline problematic zones where either significantly over- or underestimated the observed landslide density. Approach: Landslide factor analysis. source:* Havenith et al. *(2015b)*

.

2.3.4 *Landslide catalogue*

A landslides catalogue or landslide inventory is a collection of information related to where, when and why landslides occurred. Other relevant information can be associated like type of landslides, area or other morphological information of the landslide, number of people affected, deaths, material losses, geomorphological features, rain intensity, earthquake magnitude and so on, depending on the purpose of the catalogue and the sources of information. The techniques used to create the landslide catalogue depends not only on the purpose of the inventory but also the extent of the study area, the scale of the base maps, resolution, characteristics of the available imagery and the resources available to complete the work (Guzzetti *et al.*, 2012).

The traditional method to collect landslide data is based on the aerial or satellite images interpretation, topographic maps analysis, and field inspection. The result of this methodology is often subjective, incomplete, time-consuming and resource intensive. However, it is still used for areas where the data is scarce or when the area to cover is too extensive. On the other hand, new resources like google earth allow a faster delimitation of landslides and the creation of a polygon based catalogue through digitization.

The study area does not have a complete landslide catalogue; however, the amount of recorded events is sufficient to perform the landslide susceptibility assessment. The **Global Landslide Catalog** (GLC) (Kirschbaum *et al.*, 2015, 2010) is a point dataset with a total of 18 points in the study area. The attribute table associated is complete and gives information about the country, triggering factor, type of movement, fatalities, date and so on. On the other hand, the **Tien Shan Geohazard Database** (Havenith *et al.*, 2015a) is a polygon database created by the compilation of small datasets that used different techniques like manual delimitation, supervise classification and compilation of events reported in newspapers or local databases, as well as fieldwork. In total 701 polygons area

taken from the Tien Shan Geohazard database; however, the available database doesn't contain information related to the type of movement, date of occurrence or triggering factors.

A final polygon landslide catalogue was created by the integration of the GLC, the Tien Shan Geohazard database and manual delimitation of landslides from Google Earth Imagery to include the Western Pamir and the Tadjik basin in the analysis. Finally, 0.02% of the area is covered by a total of 1003 polygons with areas in the range of $0.23km^2$ to $16.9km^2$, from which, landslides with less than $0.5km^2$ are predominant (figure 2.19).

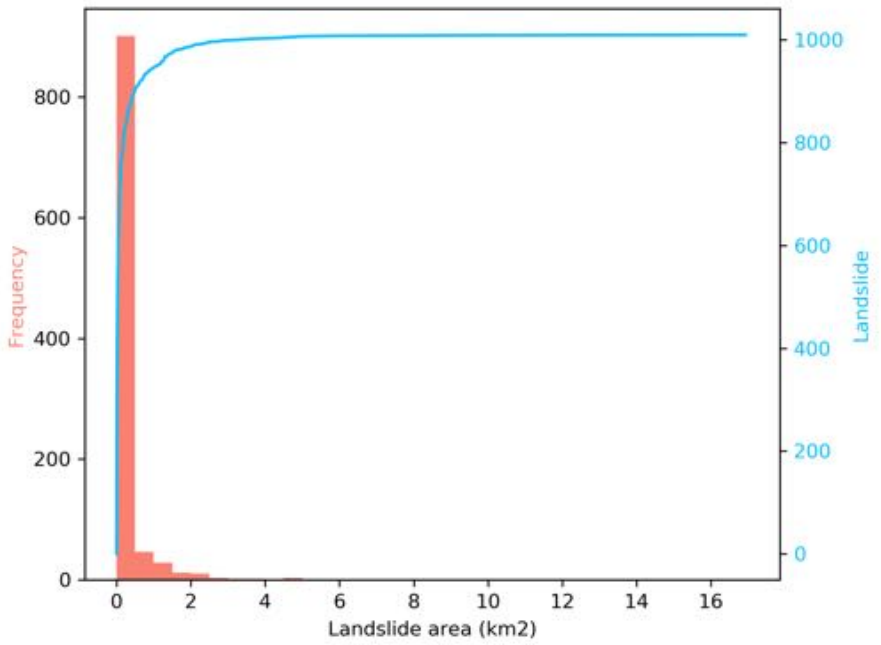


Figure 2.19: *Area distribution of the landslides in the study area. Left axis: Frequency values, right axis: Individual landslides area.*

The distribution of the landslides based on the catalogue is shown in figure 2.20. The area with most landslides mapped is located in the valley of the Zerafshon river, east to the cities of Samarqand and Urgut. This area also identified by Havenith *et al.* (2015a) as hosted a major relatively recent important mass movements events. Also, some landslides dam related to rockslides were reported in the surroundings by Strom (2010). Other areas associated with a relevant concentration of landslides are located north and southeast of Dushanbe respectively. The north area of Dushanbe was the focus of a study in 1989 where an earthquake of $M = 5.5$ (Gissar Earthquake) triggered a debris slide (flow) killing hundreds of people in a village close to Dushanbe (Havenith *et al.*, 2003). On the other hand, the third area is located north to the Nurek dam in the Vakhsh River valley; Havenith *et al.* (2013) reported different landslides events that affected the area. The most recent event is the Baipaza landslide located downstream the Nurek dam which has been active since 1968. The first displacement was reported because the landslide partially blocked the Vakhsh river. Since the first movement, reactivations are recorded in 1992 triggered by a heavy rain and in 2002 triggered by the $M = 7.4$ Hindu Kush Earthquake. The last reactivation caused problems to the Nurek and Baipaz hydropower

plants. Other relevant spots are observed along the Vakhsh thrust system (Main pamir thrust) where intermediate size landslides are located. Also, a concentration of small landslides are presented throughout the Darvaz fault system.

The landslide catalogue created for the study does not have a description of the type of movement; however, from a literature review and further field observation it can be concluded that the predominant landslides type are rock falls, rock slides, rock flow loess/debris flow and complex landslides. They are landslides characterized by a detachment of the material along a surface with little or no shear displacement and the subsequent movement of the material downhill. Depending on the type of material involved, the accumulation zone can reach meters to kilometres from the crown.

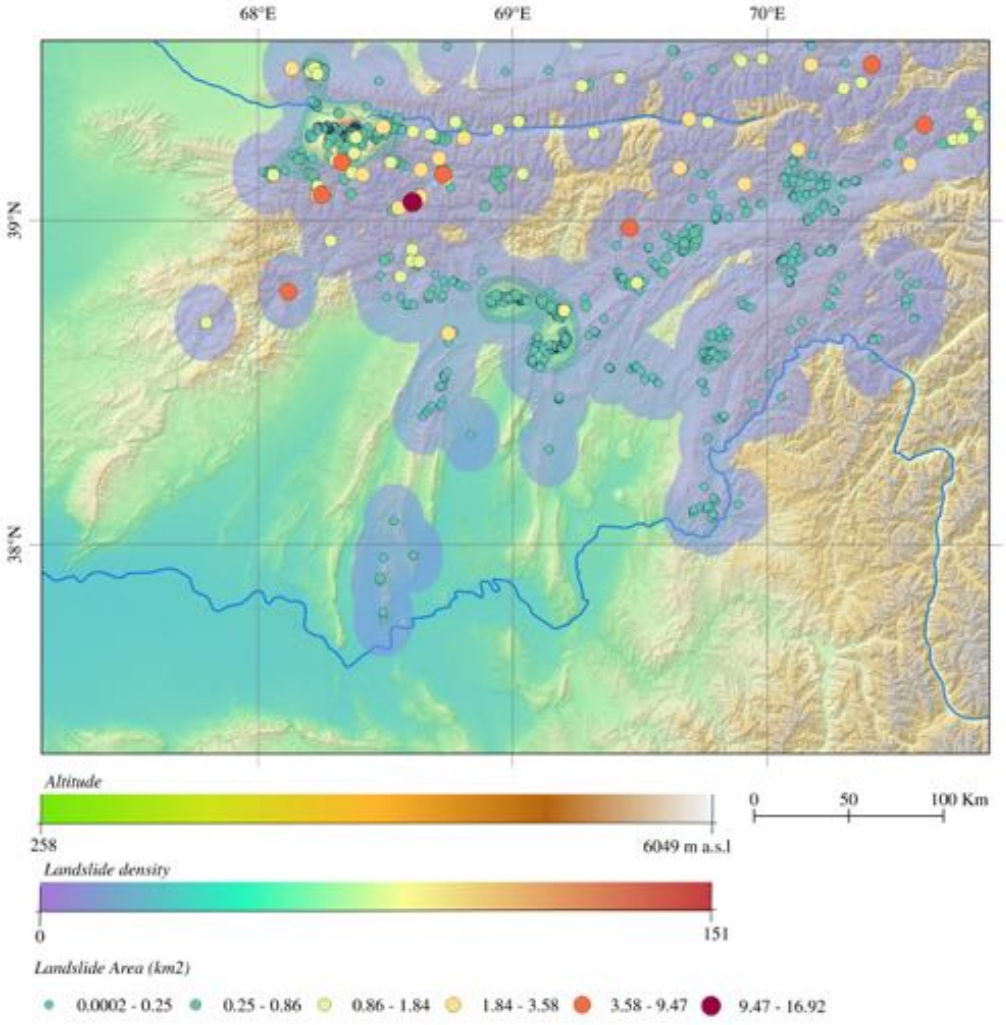


Figure 2.20: *Landslides distribution map. Landslides are indicated by dots which increasing size depending of the area.*

2.3.5 *Field observations*

A field campaign took place in the Tien Shan area. During which, some landslides were included in the landslide catalogue. Also, field observations of the predominant type of landslides in the area were collected and how they relate to the landscape.

Rock falls are commonly observed in the valleys. An substantial amount of material is mobilized and deposited at the bottom of the valley created an obstruction of the channel. An example was observed in the Yagnob river (figure 2.21), where the unstable material completely filled the valley, dammed the river and blocked the road. The vicinity surrounding the landslides is characterized by high slopes with rock exposure. other smaller rock flows are observed too.

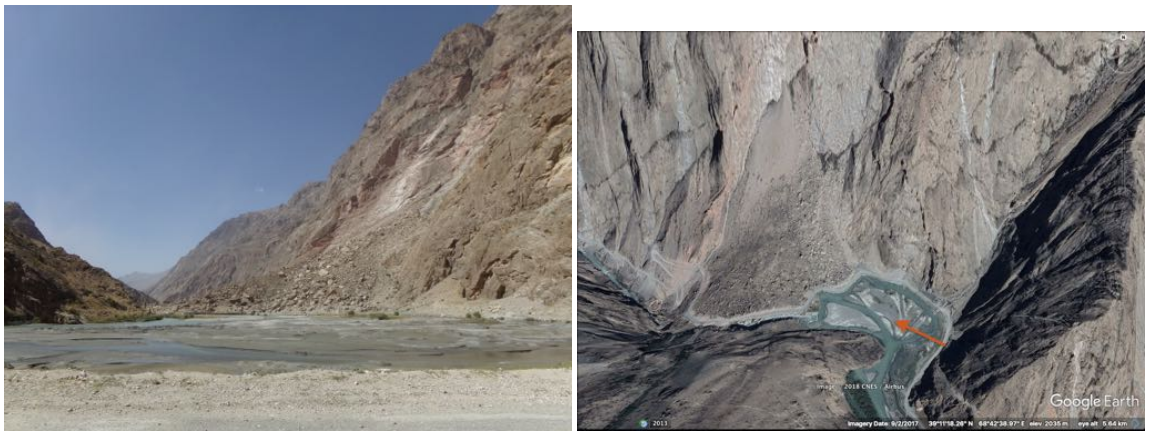


Figure 2.21: *Rockfall damming the Yagnob River. Left: Field picture. Right: Google Earth view. The red arrow indicates the direction from which the photo was taken.*

Figure 2.22: *Rock landslide that dam one of the Seven lakes in Tadjikistan. Left: Field picture. Right: Google Earth view. The red arrow indicates the direction from which the photo was taken.*

The area of the seven lakes (Marquzor lakes) in Tien Shan is characterized by a succession of landslides which created several dams along a single river. A big rock complex landslide took place in the lower part of the slope and produced a complete fill of the narrow valley (figure 2.22). The size of the mobilized material is mixed. Big blocks are identifid on the surface and also small gravel is located in the surroundings. The source

of the material in the bottom of the valley comes from both sides of the slope and landslides crowns can be identified in the surroundings.

The Iskander lake is considered one of the prettiest lakes in Tajikistan. It is formed by the dam of the Iskanderkul-Daria river by a massive landslide. The crown of the landslide is easy to identify in the figure 2.23 because of change in the concave slope and the morphology of the surface. A continuous mark is observed in the along the slopes surrounding the lake. It corresponds to the paleo-level of the lake when it was completely dammed.

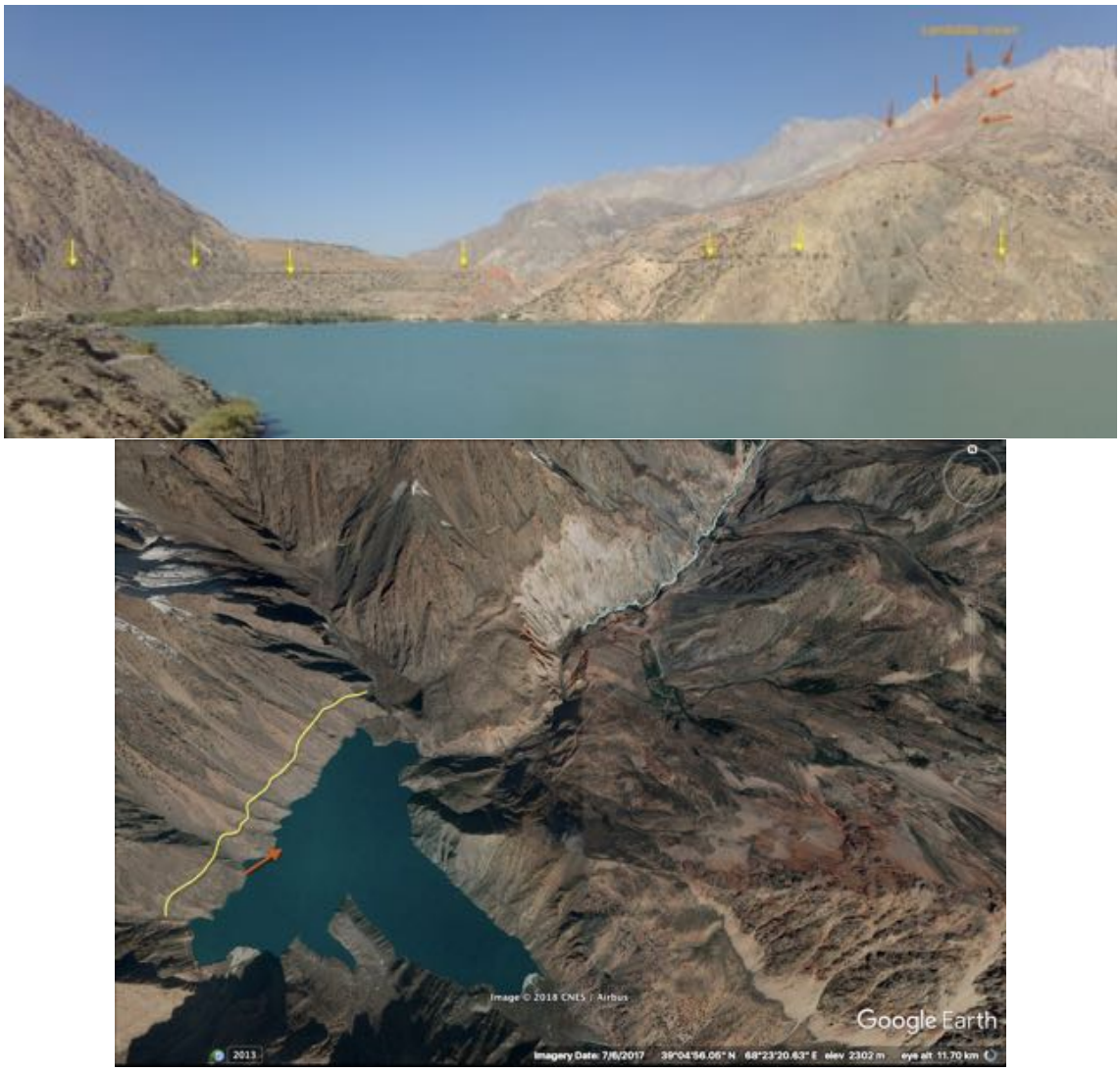


Figure 2.23: *Landslide that dam the Saratogh river to form the Iskander lake. Above: Field picture. The yellow arrows indicate the paleo-level. Oragen arrows indicate the crown of the landslide. Below: Google Earth view. The red arrow indicates the direction from which the photo was taken. The yellow line indicates the pale-level of the lake before started draining.*

One of the lithological units characterized by the presence of landslides is the Cretaceous/Paleogene units composed by calcareous rocks. Interbedded layers of gypsum

associated to the folding of the strata favor the detachment and sliding of massive rock bodies (figure 2.24).

Figure 2.24: *Mass wasting related to the Crecteceous/Paleogene sequences. Left: Rock falls located in the backslope of a cuesta in the Cretaceous/Paleogene sequence near the Khodzhagarib town. Pink arrows enhance the location of the blocks. Right: Field picture of a landslide located in the Cretaceous/Paleogene sequence.*

Chapter 3

INSTABILITY FACTORS

This chapter introduces the most common instability factors used in the landslide susceptibility assessment and the different thematic variables selected for the implementation of the different landslides susceptibility models during this work. For each thematic variable, the data-source, the method, and computation are described, and the interaction between the variable, and the landslides are discussed.

3.1 Review of the thematic variables

Mass wasting or landslides are the result of erosional processes related to landscape evolution.As previously defined, a mass wasting occurs under certain circumstances that require interaction among different factor related to materials, geomorphology, meteorological, geological and human conditions.Based on the classification of landslides, it is clear that different type of materials can be moved during a landslide. These materials must to follow characteristic like fractured material or saturated soil combined with slope gradient, in order to be affected by the gravity. Based on these, some thematic variables can be analysed in order to identify areas which are most susceptible to landslide occurrence.

Reichenbach *et al.* (2018) studied different thematic variables used for landslides susceptibility models based on a total of 596 different inputs from the literature. The number of thematic variables used vary from 2 to 22 with an average of nine variables for every single model. The author summarized the most used thematic variables and grouped them in five cluster (figure 3.1).

Morphological variables are obtained by processing terrain elevation data and have proven particularly effective in predicting the spatial distribution of landslides or the lack of landslides (Marchesini *et al.*, 2014). Reichenbach *et al.* (2018) concluded that authors prefer "simple" morphologic measures that include elevation, relief, slope, aspect, and curvature because these variables are simple to calculate and DEM are nowadays easily available. As expected, slope and aspect are the most used thematic variables to represent the relationship between the landscape features and the gravitational forces, while, factors like curvature can determine the velocity and the path of the waste movement. The elevation is another frequently input even though others morphometric variables could represent better the relation between altitude and landslide occurrence. Reichenbach *et al.* (2018) states that "simple" morphometric variables may not be the best way to capture the morphometric signature of landslides, instead complex variables that de-

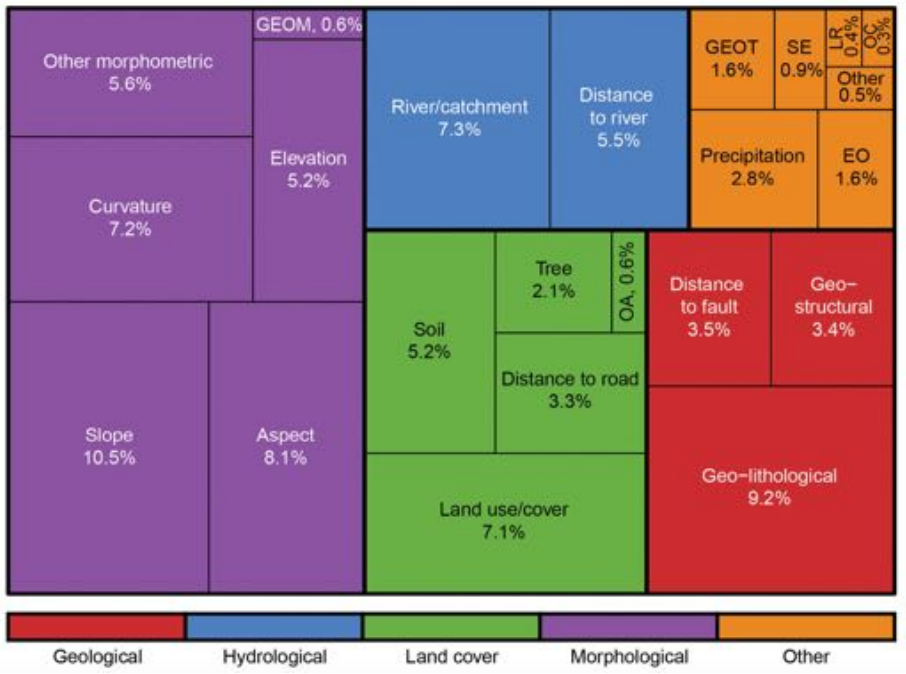


Figure 3.1: *Treemap char showing the portion of the original thematic variables as listed in the articles reviewed. Legend: EO, Earth observation; GEOM, geomorphological; GEOT, geotechnical; LR, landslide related; OA, other anthropic; OC, other climatic; SE, seismic. source:* Reichenbach et al. (2018)

scribe in an overall way the morphology of an entire slope are also good descriptors of landslide terrain; however, they are not commonly used in the literature, primarily due to the lack of specialized software.

Geological information is used to understand the materials exposed to denudation and gravitational movement. Some types of rocks are more easily eroded than others depending of the chemical composition and the size of the minerals that form it; thus it is one of the most used inputs. Whereas, information related to tectonic activity as distance to fault or geo-structural information is commonly used too, probably interpreted as triggering factors or as an indicator for the degree of fracturing of the rock.

Landcover information is mainly represented by land use/cover where the type of surface and vegetation are discriminated, while the less used but also important, is the composition of the soils which determines the characteristics of the materials that could be moved by a landslide. Other parameters in the clusters are the distance to roads which relates to the footprints of the human activity in the landscape as well as tree presence and other anthropogenic information.

Two **hydrological** variables are commonly used. The river/catchment variable is widely used to understand how landslides are spatially distributed concerning the catchment areas. The study of the catchment enables the discriminate between areas that are affected by regional processes like tectonic: also, characterized slopes under the same erosional conditions. On the other hand, the distance to river aims to recognize patterns on the landslide distribution associated to the river channel.

Finally, Reichenbach *et al.* (2018) groups **other parameters** like precipitation (one of the most common triggering factors), earth observations, geotechnical information, seismic data, landslide related, other climatic and others grouped a less relevant miscellaneous group.

In this study, a total of 17 thematic variables were created and used as a predictors of landslides occurrence for the present work. They are grouped in 5 thematic groups based on the purpose of the variable and its significance (table 3.1).

Table 3.1: *Thematic variables*

Thematic group	Factor	Significance	Datasource
Geology	Lithology	Rock type association.	Geological maps
Climatic and hydrological	Precipitation	Rainfall landslides triggering.	HAR
	Glacial distance	Glacial and periglacial geomorphological processes.	RGI
	Elevation Above Channel	Influence of the gradient and potential energy.	DEM
	Distance to river	River dynamic impact .	DEM
	Topographic wetness index	Saturation conditions.	DEM
Land Cover	NDVI	Slope instabilities in relation to presence or absence of vegetation.	Landsat 8
Geomorphology	Slope	Potential energy.	DEM
	Aspect	Effects to sun/wind exposition.	DEM
	Topographic position index	Separation between ridges, valley bottoms and fat areas.	DEM
	Surface Roughness	Erosional processes.	DEM
	Elevation Relief Ratio	Characterization of the landscape.	DEM
	Surface Index	Discrimination between erosional and steady-state landscapes.	DEM
	Local Relief	River incision.	DEM
	EigenValues	Curvature of valleys and ridges.	DEM
Tectonic	Distance to Fault	Effects of seismicity and fracturation.	CAFD
	Seismo-zones	Density/recurrence of earthquakes.	EMCA

Variables like elevation and curvature are not included in the study because it is considered that others of the geomorphological parameters can represent better the relation of the altitude or the surface characteristics and the landscapes. On the other hand, the soil information is not available at a detailed scale, thus, the NDVI is used to discriminate vegetated areas. The river/catchment analysis is not implemented because of the size of the area; however, geomorphological parameters like elevation relief ratio reflect the influence of different processes at the catchment level. Variables like EigenValues, topographic position index (TPI), surface index (SI), surface roughness (SR) and local relief (LR) are less or never explore to the implementation of landslide susceptibility models.

The relationship between the different thematic variables and the landslides are analysed first, based on the normalized landslide density for standard classes in the variables, calculated by the equation 3.1. Values of 1 indicate an average landslide density. Values less than 1 represent a class characterized by landslides less than the average, while, more than 1 represent a higher landslides density than the average (Havenith *et al.*, 2015b). The landslide density is contrasted by the number of pixels per class.

$$LandslideDensity = \frac{NLandslides_{class}}{NLandslides} * \frac{NLandslides_{variable}}{Npixels_{class}} \tag{3.1}$$

Based on (Carranza & Sadeghi, 2010), an analyse of the spatial distribution of the landslides with respect to the predictive variables is implemented by the construction of cumulative relative frequency distribution graphs. First, the cumulative relative frequency distribution for the variable (CRF_v) is plotted as well as the cumulative relative frequency distribution of the landslides in the variable (CRF_L). The differences between the two is understood as the spatial distribution. A positive spatial association is determined when the CRF_L plots above CRF_v, on th contrary, if CRF_L plots below CRF_v, a negative spatial association is determined. The differences ($CRF_L - CRF_v$), indicates for every value of the variable, how much different the frequency of landslides are.

3.2 Geology

A substantial number of geological work has been done in the area since the Soviet Union times in order to determine the geodynamics and its relation to the mineral occurrence. All those studies have focus either on local scales that do not allow the creation of regional cartography, or on regional scales where lithological units have been generalized based on their geodynamic unit or age.

A complete-paper based geological atlas is available for the area as well as the description of the geological units in 1:200.000 scale. The area is entirely covered by 34 maps in .tiff format and its corresponding legend is available in Russian at the Federal State Budgetary Institution A.P. Karpinsky Russian Geological Research Institute (FGUP VSEGEI) (2018).

On the other hand, some digital geological cartography is available for the Tien Shan, Central Pamir, North Pamir, and South Pamir and Afghanistan. This information is available as a shapefile, but it is not public. The metadata related to the creation of the dataset is not presented, making quality evaluation difficult. Although, by the comparison of the georeferenced paper-based maps and the datasets it is possible to recognize similitude in the polygons shapes, geological codes and description, thus, we assume that they were digitalized from the geological maps from the (Federal State Budgetary Institution A.P. Karpinsky Russian Geological Research Institute (FGUP VSEGEI), 2018).

The digital datasets are not homogeneous concerning the scale of production, thematic completeness or consistency. Additionally, the datasets present an important number of topological errors and do not cover the whole area. Some dataset overlaps each other creating inconsistencies as well as some gaps between the datasets. In order to use

this information an intense and time-consuming preprocessing must to be done.

The spatial accuracy of the different datasets is determined by the visual comparison between the shapes of the polygons and the geological maps. The spatial accuracy for Tien Shan, North Pamir, Central Pamir, and South Pamir are very good (90-95% of the polygons fits). Nevertheless, accuracy cannot be determined for the Afghanistan dataset because no information about the data source is available. On the other hand, the thematic accuracy varies from one dataset to another. The North Pamir dataset has a very good accuracy, based on the comparison of the codes in the attribute tables and the one assigned in the geological maps. Similarly, the Central Pamir has a good thematic accuracy because some of the polygons have different codes related to the ones in the geological map. In contrast, the Tien Shan and South Pamir dataset have a bad thematic accuracy, that means that less than 50% of the codes are related to the geological map. The former has no geological codes; the attribute table consists only names of geological time periods except for the quaternary units; while the later used geological codes that do not fit with the international standard for geological coding.

The precision of the dataset is assumed as the scale of the map from which the information was digitalized. For the Tien Shan, North, Central, and South Pamir, the precision is 1:200.000; however, for the Afghanistan database this information is unknown; although, the detail of the information in this dataset is similar to the other datasets, so it is possible to conclude that the precision of all dataset is not different one from the other.

The spatial consistency of the datasets is determined by the topological rules in the digitalized polygons. The most common geometrical errors in the datasets are polygons with holes and silver polygons; Only the South Pamir dataset contained information on self-intersection between polygons. After running topological analysis; we found that all the datasets have a bad spatial consistency. On the other hand, the thematic consistency is analysed based on the geological description in the attribute table and its relation to the map description. For the Tien Shan database, just a few descriptions are storaged and corresponds to less than the 9% of the total information. The bad thematic accuracy of this data set is related to the dataset purpose because it was created to analyse the igneous rock in the area (Käßner *et al.* , 2016). For North and Central Pamir presents 60% and 70% of the descriptions are similar to those collected by the (Federal State Budgetary Institution A.P. Karpinsky Russian Geological Research Institute (FGUP VSEGEI), 2018). In consequence, it is possible to say that the thematic accuracy of this dataset is acceptable. The South Pamir and Afghanistan dataset have a good and very good thematic accuracy. The first one contains 88% of the descriptions according to the based maps and the second one includes all the descriptions such as the lithology as a general field, but also essential information such as descriptions of the materials.

Finally, the completeness of the dataset is evaluated based the requirements to use the data. The geological description is the most important attribute related to the objective of the project. The Tien Shan dataset presents bad completeness because the information associated with the polygons is not enough to fill the project needs. Similarly occurs to the Central Pamir and South Pamir dataset where more information is reported still it is not enough, for this reason, it is considered a dataset with acceptable completeness. In contrast, The North Pamir and Afghanistan datasets have good completeness and just a few processing will need to use them as an input.

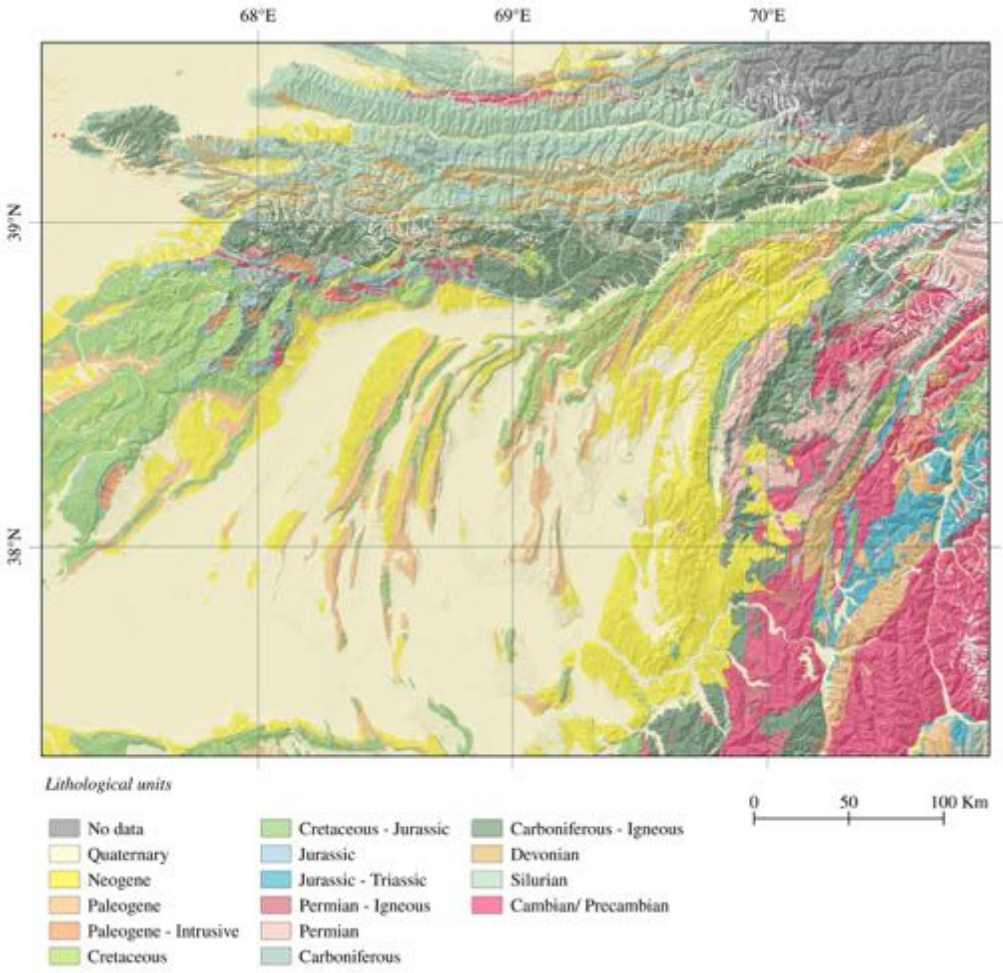


Figure 3.2: *Geological map classify in 16 geological units based on the lithology.*

The final map for **lithological** information was obtained after a data integration process using Qgis as GIS software, topological errors were also corrected as part of this. The lithological information is grouped based on the age of formation and the type of rock (figure 3.2).

The Quaternary unit (1) is composed by pebbles, loess, sandy loams and quaternary alluvium located in the flat areas, generally associated with alluvial plains or terraces. The Neogene unit (2) is characterized by siltstone, clays, sandstones, and conglomerates identified in the souring and in the mountain ranges inside the Tadjik basin; even though the materials are susceptible to move, the slopes where they are located are not representative enough to present a high landslide density (figure 3.3). The Paleogene units are divided in a sedimentary unit (3) (Limestones, sandstones, clays, dolomites, marls, conglomerates and gypsum) and Paleogene intrusions (4) represented by granites. The sedimentary unit is predominantly located the mountain ranges inside the Tadkid depression; while the intrusions are located in the surroundings of the Gissar Suture (Tien Shan).

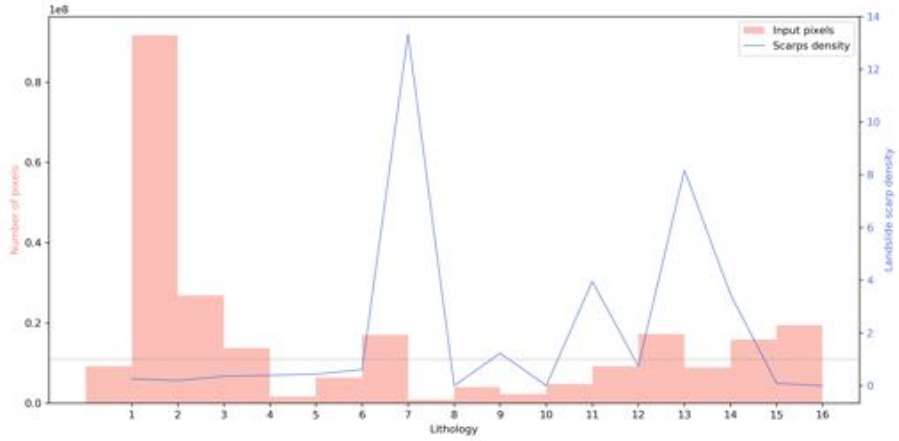


Figure 3.3: *Histogram of the number of pixels per class compared to the landslide density per class. 1. Quaternary, 2. Neogene, 3. Paleogene, 4. Paleogene-Intrusive, 5. Cretaceous, 6. Cretaceous-Jurassic, 7. Jurassic, 8. Jurassic-Triassic, 9. Permian-Igneous, 10. Permian, 11. Carboniferous, 12. Carboniferous-Igneous, 13. Devonian, 14. Silurian, 15. Cambrian/Precambrian, 16. Glacier areas*

The Cretaceous unit (5) is dominated by conglomerates, sandstone, clay, siltstone, and gravel. It is predominant in the Gissar range, some areas of the Main Pamir Trust (MPT) and the south of the Tadjik basin. Similarly, the Cretaceous-Jurassic unit (6), is spatially associated and it is composed by limestones, marls, mudstones, clay, sandstones, gypsum and few conglomerates. All those units are characterized by the presence of low landslide density (figure 3.3).

The highest landslide density is associated with the Jurassic unit (7) where conglomerates, shales and, coal are predominant; however, it is important to recognize the low number of pixels that represent this class. Similarly occurs to the Jurassic-Triassic unit (8), composed by pophyritis, tuffs, tuff breccia, quartz conglomerate, sandstone, shale, coal lenses, gravelitas, allite and bauxita.

The Igneous Permian unit (9) present a slightly high landslide density. It is located near to the Gissar suture as well as in the Pamir. It is composed of granite, diorite, gabbro and, andesite. On the other hand, the Permian sedimentary rocks (10) are predominant in the Pamir, characterize by sandstone, limestones, shales, conglomerates and some tuffs. Another unit associated with high landslide density is the Carboniferous sedimentary (11) unit. It is composed by conglomerates, sandstones, shales, limestones, siltstone, and dolomites located mainly in the Tien Shan, but some stripes are identified in the Pamir. The Igneous unit in the Carboniferous (12) is broadly extended in the south of Tien Shan and the Western Pamir. It is composed of granodiorite, granite, porphyry leucocratic granites, basalts, tuff, and breccias.

A high landslide density characterizes the Devonian (13) and Silurian (14) sedimentary rocks. They are composed by limestones, dolomites, sandstones, and shales, but the Devonian sequence presents conglomerates and siliceous rocks too. They are located in the Northern part of the area in the Tien Shan. The Cambrian and Precambrian units (15) are metamorphic rocks like schist, marble, gneiss, migmatite, and cataclasite located mainly in the Pamir and as a small stripe in the north of Tien Shan and are associated to

a low landslide density.

3.3 Precipitation

The Chair of Climatology - TU Berlin, provides a high-resolution atmospheric dataset, as part of the High Asia Refined analysis project (HAR) (Maussion *et al.*, 2014)). The dataset is generated by dynamical downscaling of global data using the Weather Research and Forecasting (WRF-ARW) model. The results are strongly dependent on the quality of the global input data and the capacity of the WRF model to simulate the atmospheric processes.

The data is delivered as binary NetCDF files with different spatial and temporal resolutions. For the study area, the data is available in a spatial resolution of 30 km. Even that the spatial resolution of the data is coarser than the mapping unit, precipitation information is a crucial factor in the landslide susceptibility understanding. On the other hand, the temporal resolution of the data set is proximately 12 years, from October 2000 to December 2012 and the dataset is available with an extend of month or years.

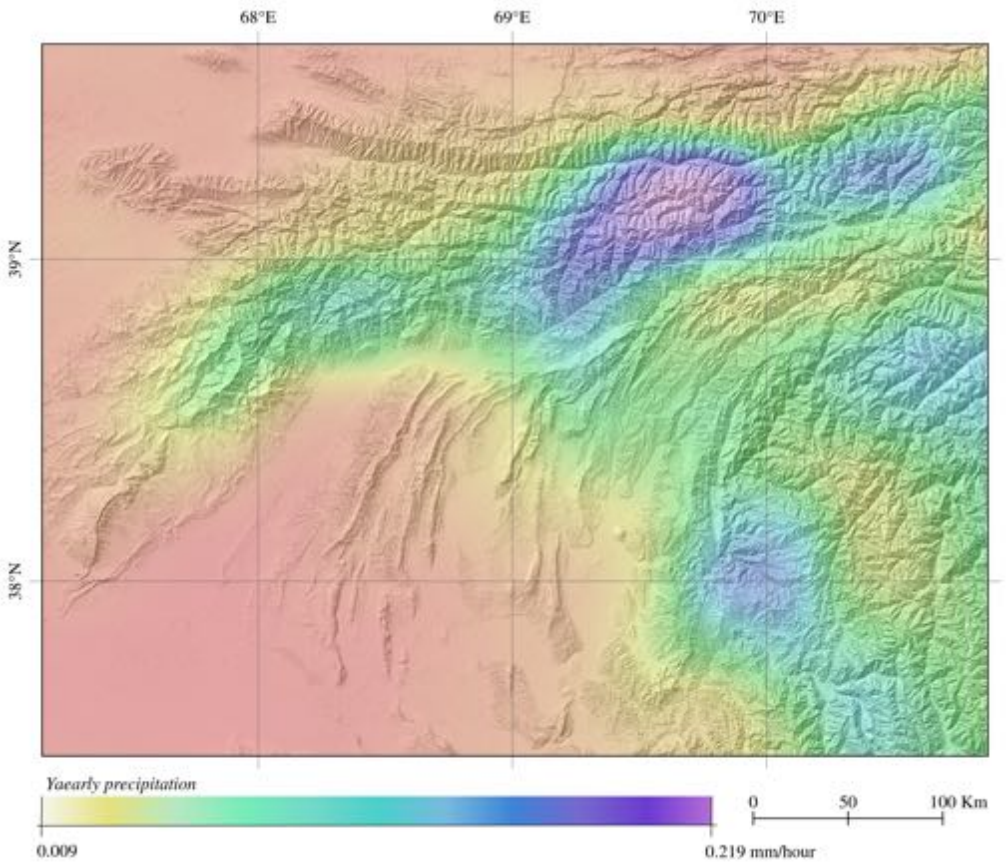


Figure 3.4: *Annual average precipitation map calculated based on the HAR annual precipitation from 2000 to 2014.*

The authors claims that unrealistic precipitation gradients are present in areas with steep topography like the Himalayas, however,Pohl *et al.* (2015) tested different precipitation products for their application in a hydrological modelling approach in the Pamir and found that the HAR provides the most reliable estimates for precipitation compared to the TRMM (Tropical Rainfall Measuring Mission).

The HAR dataset was re-projected from the map projection used by the WRF model to WGS 84. The 12-yearly precipitation information was integrated by the calculation of the mean of the yearly precipitation given in mmh^{-1} and re-sampling by a cubic interpolation method to obtain the precipitation information to be used (figure 3.4).

The precipitation distribution is presented in figure 3.4 from which is possible to conclude that the area that received more precipitation per hour in the last 14 years is located in the eastern area of the Tien Shan mountain range; however, high precipitation values are reported as well in the Pamir. The distribution of the precipitation is coherent with the analysis of Pohl *et al.* (2015) whose reported almost no precipitation during the summertime in the Pamir, a fact that decrease the annual average precipitation to this area compared to the Tien Shan.

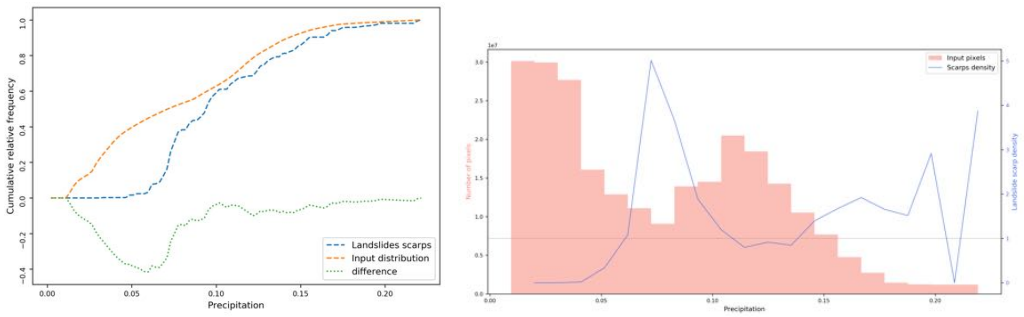


Figure 3.5: *Spatial relation between precipitation and landslides. Left: Cumulative relative frequency. Right: Histogram of the number of pixels per class compared to the landslide density per class.*

Landslides triggered by rainfall are associated either to extreme events (unusual precipitation) or long duration precipitation period. By analysing the difference in the cumulative relative frequency of the landslides and the precipitation variable (figure 3.5); we clearly observe a negative spatial association for areas with low precipitation (< 0.07 mm/h)(figure 3.5).On the other hand, the landslide density per class show a bimodal distribution with higher densities around 0.07 mm/h and above 0.15 mm/h This binomial distribution is a complex to interpret. It is possible to expect association between landslides density and precipitations (Wieczorek & Guzzetti, 1999; Dai & Lee, 2001; Saito *et al.*, 2010) and the higher densities for precipitations above 0.15 mm/h would be reflect his association. However, the peak at 0.07 mm/h is more difficult to interpret and suggest that other factors need to be taken into account.

3.4 Distance to a glacier

The study area is characterized by the presence of a significant number of glacial areas not only in the Pamir but also in the Tien Shan covering the north and east part of the area permanently. Glacial dynamic gives a unique characteristic in regarding geomorphology and surface processes. In order to introduce this factor in the model, The Randolph Glacier Inventory (RGI) was used.

The RGI is a global inventory of glacier outlines as part of the Global Land Ice Measurement from Space initiative (GLIMS). The outlines of glaciers intended to represent how the world's glaciers were, near the beginning of the 21st century. The dataset claims a high priority in the completeness of coverage rather than accuracy in dating, delineation, and georeferencing. However, the 2017 product (used for this project) present a lot of improvements concerning spatial and thematic consistency. The inventory is available for different parts of the world. For Central Asia, the database was created as a compilation of different datasets created by T Bolch (n.d.); Guo *et al.* (2015); Nuimura *et al.* (2015); Shi *et al.* (2009); Raup *et al.* (2000); Kutuzov & Shahgedanova (2009); Kriegel *et al.* (2013), as well as manually mapping and semi-automatic delimitation from ASTER and Landsat images.

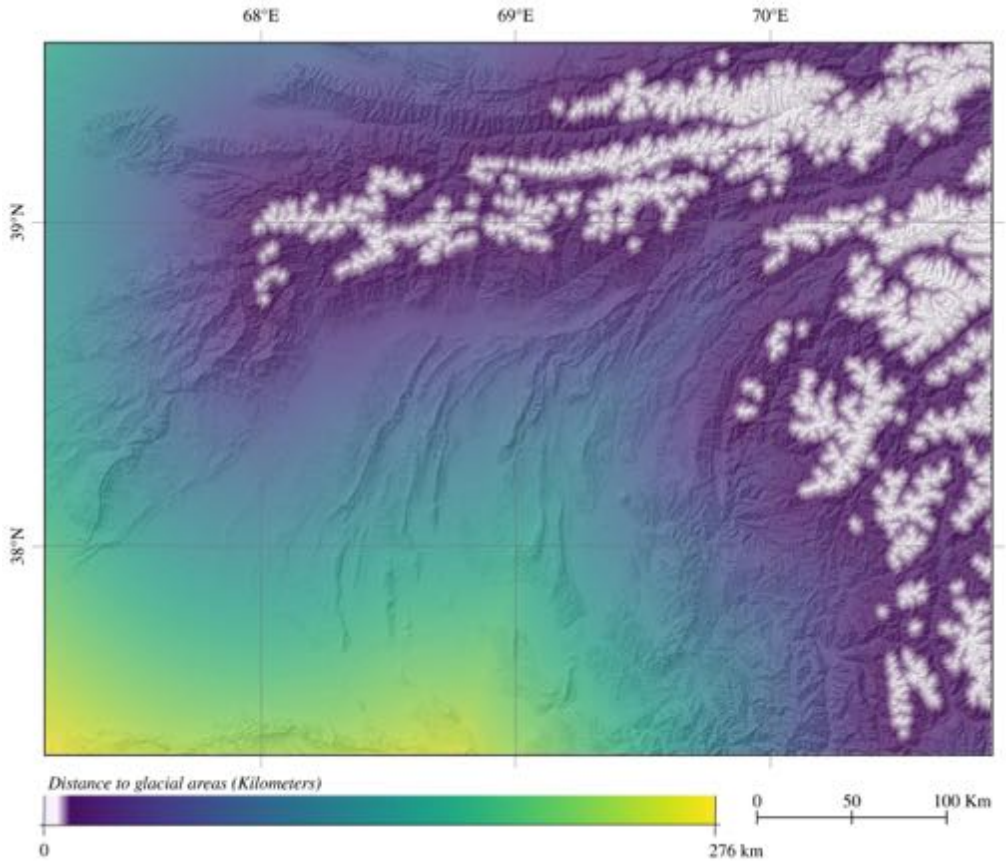


Figure 3.6: *Map of the distance from the the present day glaciers calculated from on the GLIMS inventory .*

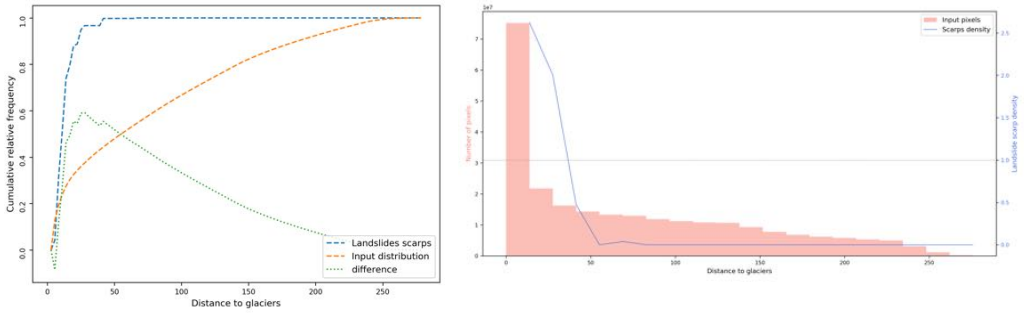


Figure 3.7: *Spatial relation between the glacier areas and landslides. Left: Cumulative relative frequency. Right: Histogram of the number of pixels per class compared to the landslide density per class.*

The glacier contours from the RGI are rasterized with a pixel size of 30 meters. Then, a calculation of the Euclidean distance from every pixel to the glaciates areas is performed to obtain the minimun distance to a glacier (figure 3.6).

Glaciers have a strong influence in the development of the landscape as well as in the erosional and slope processes. They are located in the Pamir and Tien Shan and its influence decrease to the south-west (figure 3.6). Glacier distance as a triggering factor is associated with the increasing of water supply to the soil during the melting season, that leads to mass wasting like landslides or flows because of the saturation of the materials. Another influence is related to the degree of fracturing of the rocks due to gelifraction and the lack of vegetation to give support the outcrops where rock falls or flow can be triggered. Also, moraine deposits are not mapped in the lithological information, but they are unconsolidated materials where big landslide can occur (Korup & Tweed, 2007).

A strong positive spatial association exists between the landslides occurrence and the distance to the glacier areas (figure 3.7), that decreases to a negative spatial association with the increase in distance.

3.5 Distance to channel

The global 1-arc second (30m) SRTM (Shuttle Radar Topography Mission) digital elevation model is used to create the hydrological information as well as the geomorphological parameters.

The Shuttle Radar Topography Mission (SRTM) was a project launched in 2000 to acquire radar data which were used to create a global land elevation dataset. Two sensors collected the information; the first one a SIR-C band Spaceborne Imaging Radar with a Wavelength of 5.6 cm from National Aeronautics and Space Administration (NASA) (2000) and the second a X-Band Synthetic Aperture Radar (X-SAR) from German Aerospace Center (DLR), Italian Space Agency (ASI) (2000) in order to collect interferometric radar data, which compares two radar signals taken at slightly different angle but at the same time to calculate the surface elevation.

The mission was orbiting the earth 16 times each day for 11 days. In total, the mission completed 176 orbits of which 159 were used for operational mapping and also collected rada data successfully, over 80% of the Earth's land surface between 60°north and 56°south latitude. The mission offers three types of data products: SRTM Non-Void filled, SRTM Void Filled and SRTM 1 Arc-Second Global elevation data. The last product is the one used for producing geomorphological data during this research (National Aeronautics and Space Administration (NASA), 2000).

The digital elevation model (DEM) is provided in geographic coordinates with a WGS84 Datum; a horizontal spacing of 1 arcsec that corresponds approximately to 30m of resolution. The vertical elevation is given in meters and WGS84 is used as vertical datum too; this means that ellipsoidal heights are provided. The required horizontal accuracy is 20m (90%). The errors in the horizontal accuracy are related to uncertainties in the antenna position and from the interferometric phase process; but no displacement higher than 20 m are reported (Rabus *et al.*, 2003).

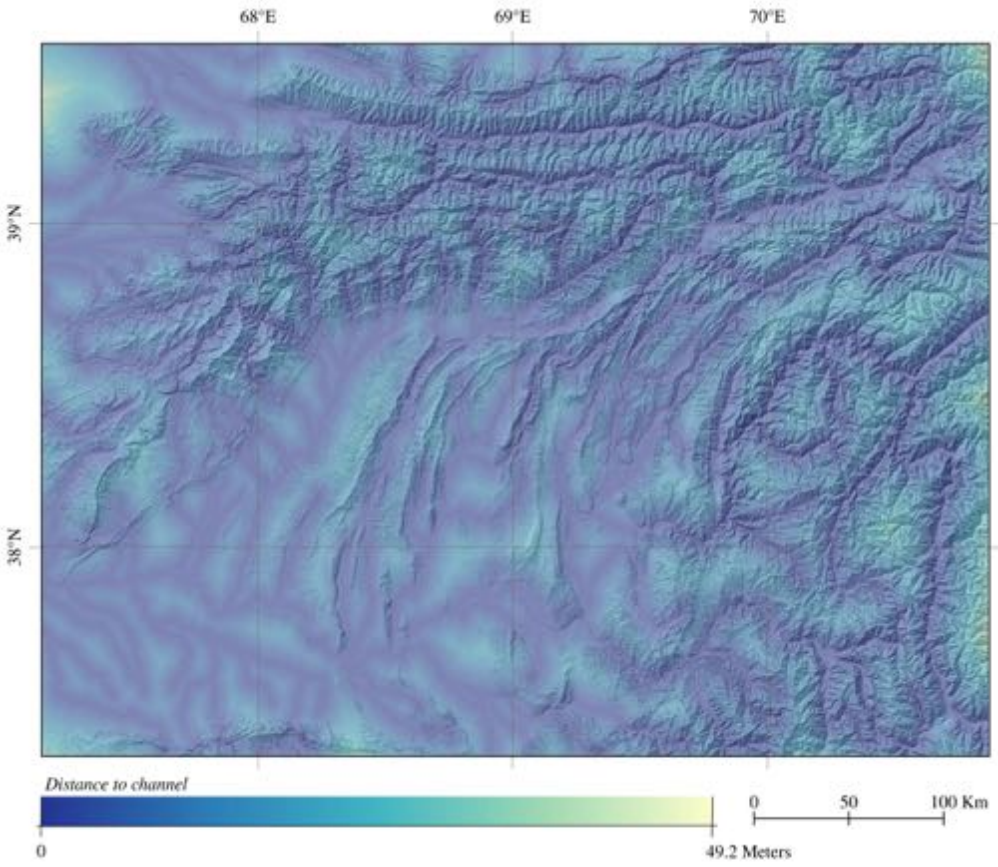


Figure 3.8: *Map of the distance to channel calculated based on the DEM using TecGEMS.*

The absolute vertical accuracy is related to the error budget throughout the entire mission and it is calculated in 16m (90%). Initially, it was expected that only a slow variation of drift of the radar and the start trackers contribute to the vertical error; however, during the calibration phase, it was recognized that those variations affected more than expected. Because these variations are not linear, alternative strategies are currently be-

ing investigated to reduce the error. The vertical relative accuracy is set as 6m (90%) within a 225x225 km area. It is assumed that the user can easily correct the interest area by adding a single corrective height value; however, this requirement can only be met with little margin (Rabus *et al.* , 2003).

The drainage system is extracted from the DEM using TecGEMS 0.4.2 (Andreani *et al.* , 2018). The network extraction process starts with the implementation of a pit filling algorithm as a correction procedure to the DEM to avoid the existence of pixels of group pixels surrounded by cells with larger values. Then, the flow direction and the identification of flat areas is performed by calculating the difference between the high cells surrounding a pixel. To compute the flow direction in the flat areas, a medial axis algorithm is used and then, the medial axes are connected to the outlets to resolve the flow directions and create the flow direction raster. Based on the flow direction raster, the flow accumulation is calculated and it is used to create the flow paths. Build on the number of connecting nodes, the Strahler orders are assigned and the final river network is created.

Streams were identified using a minimum contributing area of $50km^2$ and a minimum length of 50 km, in order to obtain the main rivers and their largest tributaries and avoid high Strahler order information. Euclidean distance is calculated from each main river (figure 3.8).

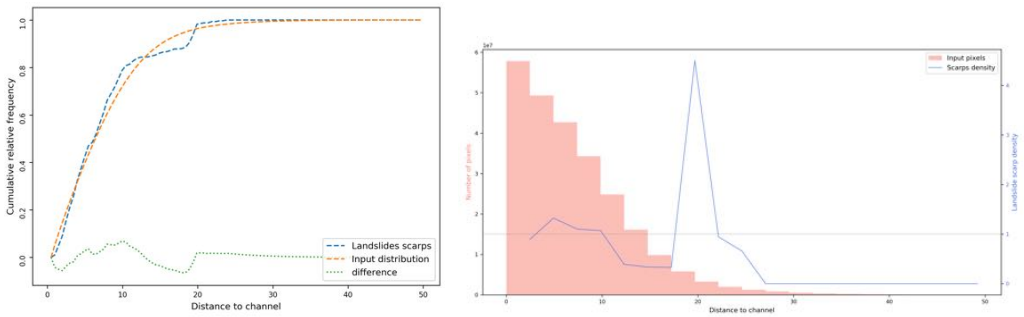


Figure 3.9: *Spatial relation between the distance to river channel and landslides. Left: Cumulative relative frequency. Right: Histogram of the number of pixels per class compared to the landslide density per class.*

The cumulative relative frequency diagram shows a very week positive spatial association to the distance to a river channel and the landslide occurrence (figure 3.9) that start to decrease at 10 km. On the other hand, before 10 km, the landslide density is slightly higher; however, the peak is reached at 20 km where a crucial number of landslides are mapped. After this distance, the influence of the river dynamic is considered as non-significant.

3.6 Elevation Above Channel

The same river network extracted from the DEM using TecGEMS for the calculation of the distance to river is used as a data source.

The elevation above the channel is calculated as the difference in elevation between each pixel an their nearest stream (equation 3.2). The variable is presented as a better approach to relate the river network to the morphology.

$$ElevationAboveChannel = h_{pixel} - h_{nearestStream} \tag{3.2}$$

The elevation above the channel seeks to understand the spatial relation of the slope in terms of elevation. The higher values represent the ridges, while the lowest values are associated to flat areas or the bottom of the valleys. The maximum value in the area is 3983 m; reflecting the maximum depth of some of the major valleys (figure 3.10).

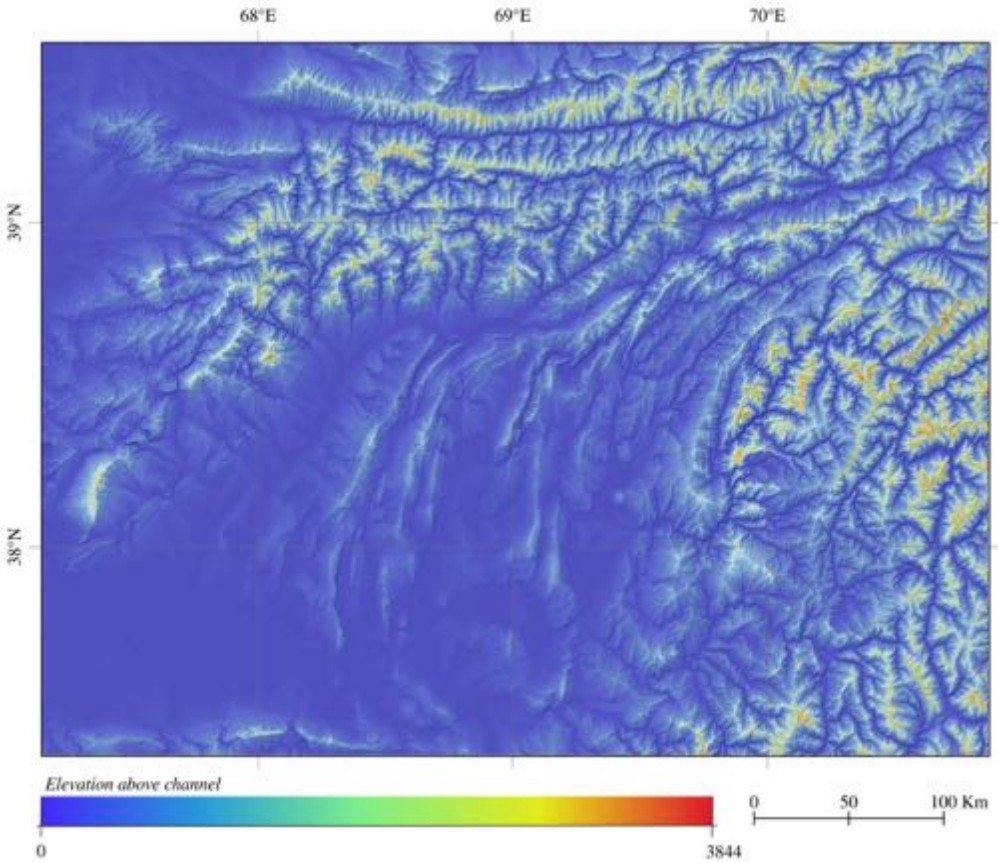


Figure 3.10: *Map of the elevation above the channel calculated based on the river network extracte from the DEM.*

A strong negative spatial association is presented between the elevation above the channel and the landslide occurrence until 700 m (figure 3.11). On the other hand, the landslide density present a peak at 1300 m (figure 3.11) and the landslide density re-

mains high (above 1) until 2500 m. This means that landslides are predominant in the upper part of the valleys above 700m.

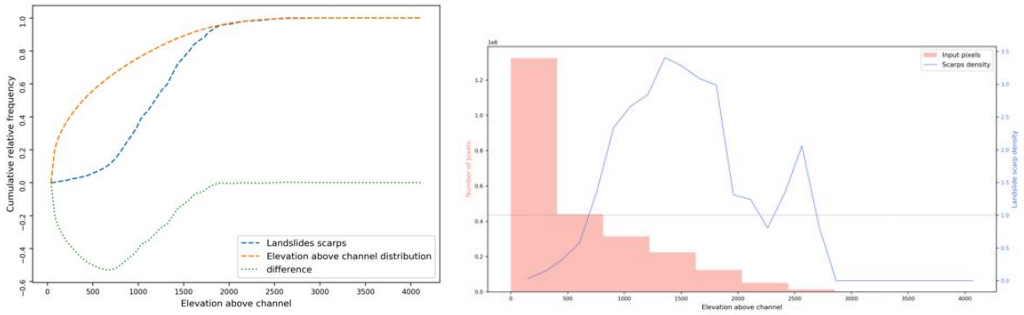


Figure 3.11: *Spatial relation between elevation above the channel and landslides. Left: Cumulative relative frequency. Right: Histogram of the number of pixels per class compared to the landslide density per class.*

3.7 Topographic Wetness Index

A DEM is used as a data source to compute the topographic wetness index. It is calculated as the ratio of the natural log of the specific catchment area (contributing area) to the slope (refeq:TWI) where a is the local upslope area draining through a certain point per unit contour length and $\tan b$ is the local slope calculated from the DEM.

$$TWI = \ln \frac{a}{\tan b} \tag{3.3}$$

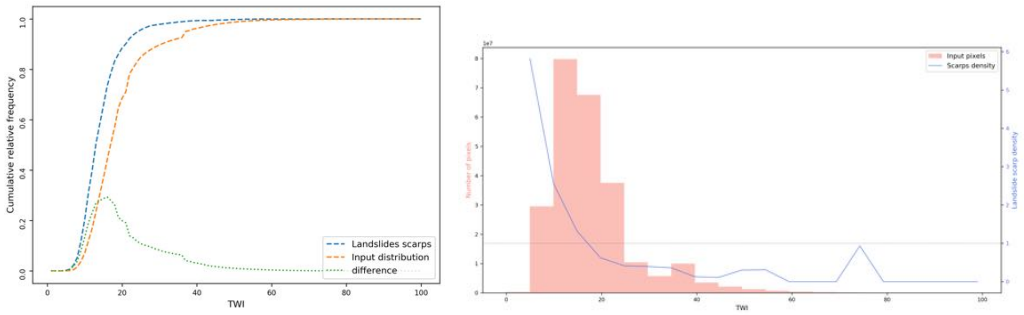


Figure 3.12: *Spatial relation between TWI and landslides. Left: Cumulative relative frequency. Right: Histogram of the number of pixels per class compared to the landslide density per class.*

The topographic wetness index describes the influence of the topography and the river system on a slope. It describes the potential saturation of an area based on the slope characteristics and the river network. High values are located in the Tadjik depression mainly associated with the fluvial plains of the main rivers as well as the lakes and river channels; while low values are located in the mountainous areas (figure 3.13). The intermediate values (around 50) are associated with the rivers and gullies in the mountainous

areas.

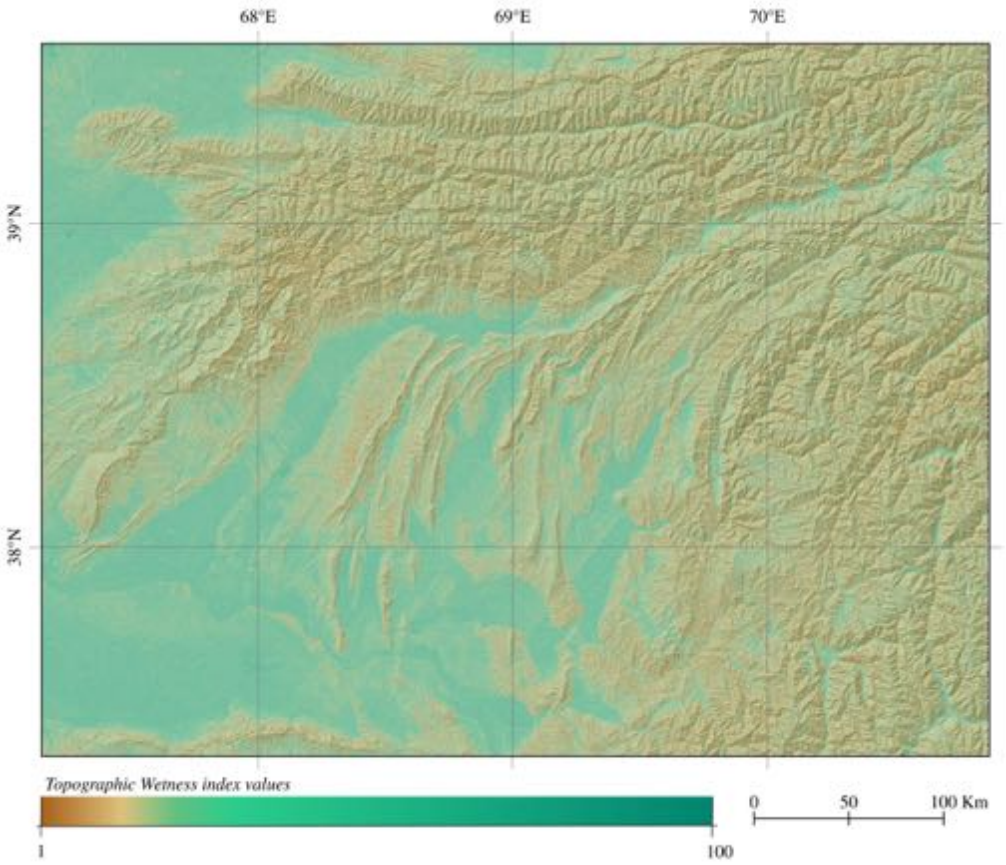


Figure 3.13: *Map of the topographic wetness index calculated based on the river network extracte from the DEM.*

There is a positive spatial association for TWI values below 40 and the landslide occurrence. For values above 20, the landslide density remains low (figure 3.12). The low values of TWI are related to hillslopes and upper channels where the landslides area mainly located.

3.8 Normalized difference vegetation index

The **normalized difference vegetation index** (NDVI) is a transformation of the spectral signature calculated based on band math between the Near Infra-red and the Red band (equation (3.4)). It is obtained from the processing of the selected Landsat images. First, the NDVI is calculated using a Python 3.0 script for every single image, and then the results are merged using histogram matching in the borders to obtain a continuous single image for the area using ENVI.

$$NDVI = \frac{NIR - Red}{NIR + Red} \tag{3.4}$$

The Landsat 8 was launched on February 11, 2013. It is a spatial satellite mission focused on the collection of high resolution multi spectral data of the Earth's surface on a global range. The Landsat 8 carries two instruments: The Operational Land Imager (OLI) and the Thermal Infrared Sensor (TIRS). The Landsat 8 OLI Multispectral bands have a pixel size of 30 meters. It has nine bands with spectral resolution as 433 nm in Band 1 and is described as the Coastal Aerosol band. The second, third and fourth are the corresponding Blue, Green, Red band with the spectral value in the range of 482, 562 and 655 nm respectively. The fifth band is the Near-Infrared (NIR) band (865 nm), while the Short Wavelength infrareds (SWIR1 - SWIR2) are assigned to the sixth and seventh band. The Panchromatic band with the center in 590 nm corresponds to the eight band while the last band corresponds to the Cirrus with center in 1375 nm (U.S. Geological Survey, 2015).

The Landsat 8 OLI/TIRS Level-2 data products - Surface reflectance provides an estimate of the surface spectral reflectance as it would be measured at ground level in the absence of atmospheric scattering or absorption. They are generated at the Earth Resources Observation and Science (EROS) Center with a spatial resolution of 30 meters (U.S. Geological Survey, 2015).

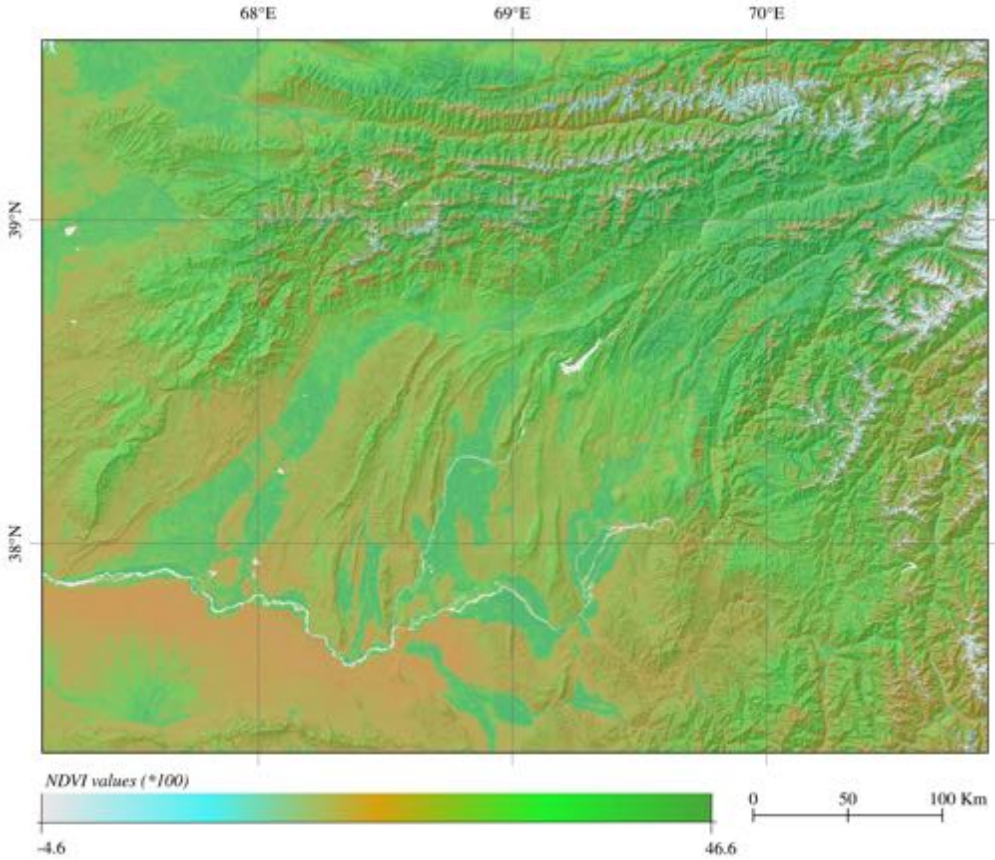


Figure 3.14: *Map of the distribution of the values of NDVI multiply by 100 calculated based on Landsat 8 images.*

The area of study is covered by sixteen images in the paths 152, 153, 154 and 155 and rows from 32 to 35. The months of July and August 2017 are selected based on the low cloud cover in most of the area; however, some areas from Tien Shan are analysed based on images taken in the same period of time but in 2016.

The NDVI values are used to identify areas with vegetation, soil predominance or rock exposure. Based on the spectral signature of the materials, it is possible to discriminate water because it is characterized for a decreasing signature from the red to the NIR band, corresponding to very negative values in the NDVI (lower than -0.25). Snow and ice because they have similar values in the NIR and the Red band, with the NIR being slightly lower, that is why the NDVI for those materials is close to -0.04. On the other hand, the soil reflectance increases with the increase of the wavelength and the values in the NIR and the red band are similar. However, the red band reflectance is slightly slower, resulting in values near to 0.025. Higher values of NDVI represent the vegetation, because of the absorption of the chlorophyll of the leaves and the high reflectance of the vegetation in the NIR.

Values between 2.5 to 20 (0.025 to 0.2 NDVI) can be associated with soil coverage or no vegetation presence. Those values predominate in the South of the Panj river in the Tadjik basin along with the Zarafshon valley. Also, the surrounding of the Gissar range and the ranges inside the Tadjik basin are characterized by a predominance of the positive lower values(figure 3.14). NDVI values between 20 to 30 (0.2 to 0.3) are identified as grassland areas and are mainly located in the mountain ranges, except for some areas in the surrounding of the glaciers or local areas. Also, low density crops are enhance in the Tadjik basin. The higher values of NDVI represent dense vegetation and the abundance is limited in the area. The landslide density increase increases rapidly for the positive values, however an association between the NDVI values and the landslides is very weak (figure 3.15)

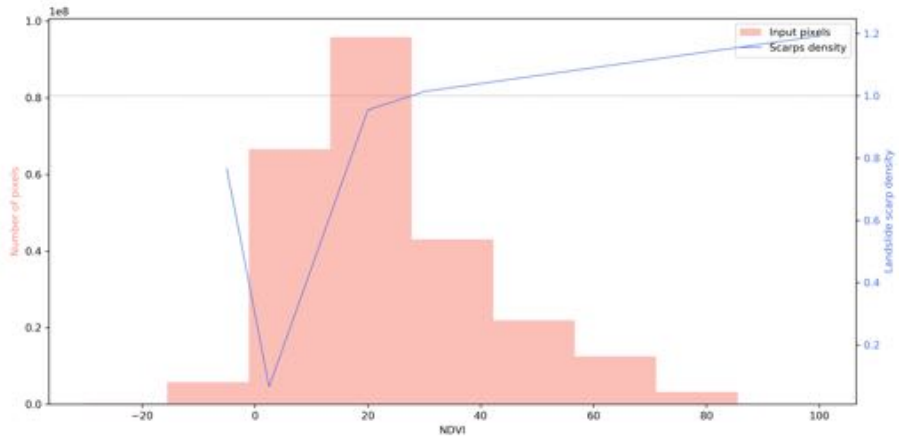


Figure 3.15: *Histogram of the number of pixels in the NDVI classes compared to the landslide density per class.*

3.9 Slope

The most common used derivative from a DEM is the **slope** that is calculated as the maximum change in elevation over the distance between the cell and its eight neighbours based on the equation 3.6.

$$Slope = ATAN\sqrt{[\frac{dz}{dx}]^2 + [\frac{dz}{dy}]^2} \tag{3.5}$$

where,

$$[\frac{dz}{dx}] = \frac{(c + 2f + i) - (a + 2d + g)}{8 * cellsize}, [\frac{dz}{dy}] = \frac{(g + 2h + i) - (a + 2b + c)}{8 * cellsize} \tag{3.6}$$

Being e the pixel of interest and a, b, c, the upper neighbours, d, e the lateral neighbours and g, h, i the down neighbours.

The slope angle and the landslides present a strong negative spatial correlation based on the analysis of the cumulative relative frequency. The lower the slope angle, the fewer the associated landslides. On the other hand, based on the landslide density (figure 3.16), it starts increasing after 40°until it reaches a peak near to 90°. This peak is an artefact created by the landslide catalogue. Because to represent the instability conditions, the depletion zone was selected, the actual slope is higher than the original slope when the landslide occurred. This peak doesn't represent the original conditions of slope.

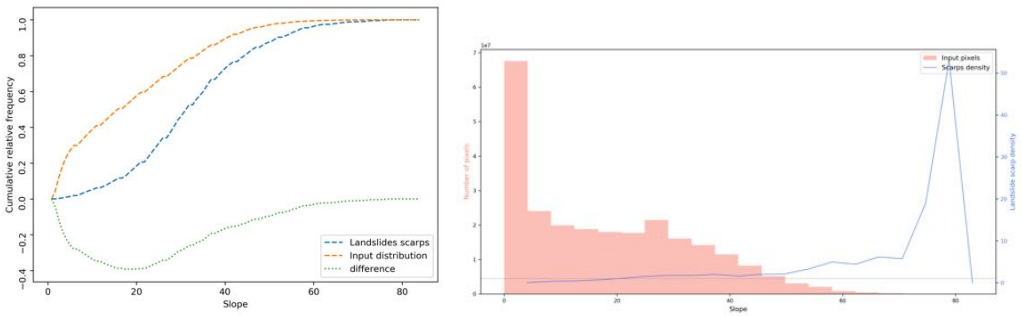


Figure 3.16: *Spatial relation between the slope and landslides. Left: Cumulative relative frequency. Right: Histogram of the number of pixels per class compared to the landslide density per class.*

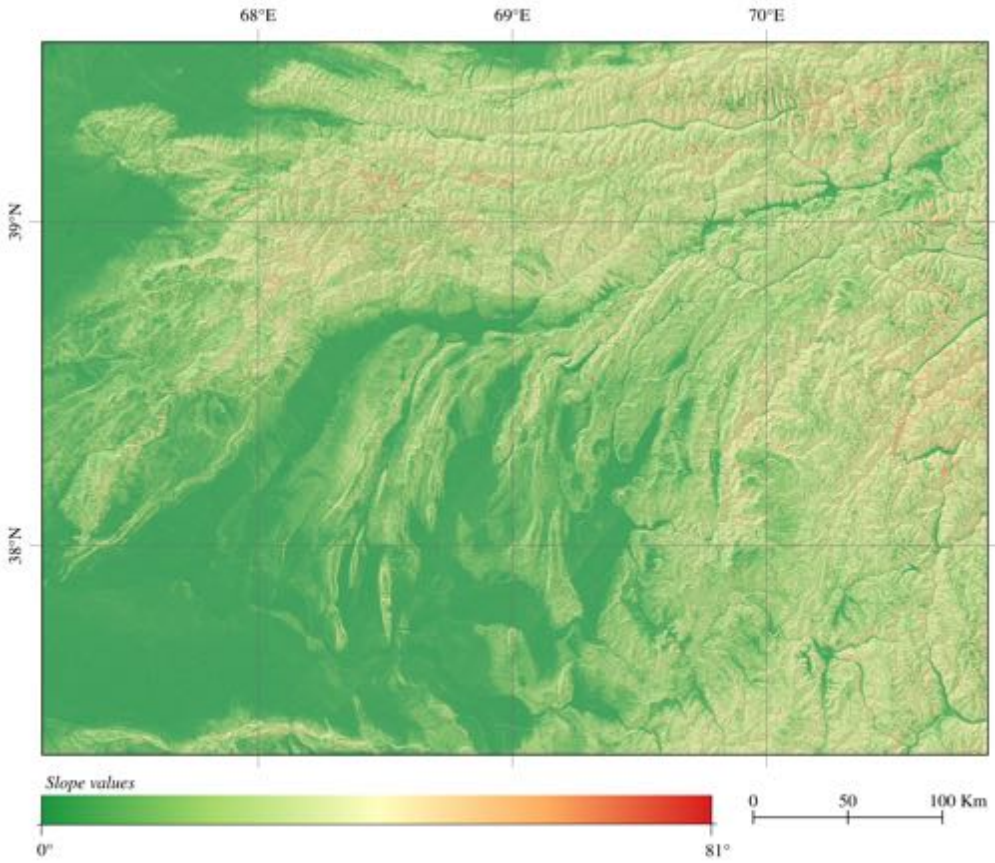


Figure 3.17: *Map of the distribution of the slopes.*

3.10 Aspect

The aspect is defined as the downslope direction of the maximum rate of change in value from each cell to its neighbours. It is usually interpreted as the slope direction and is measured clockwise in degrees from 0 (North) to 360 (again North). It is calculated applying the equation 3.7; where $[\frac{dz}{dx}]$ and $[\frac{dz}{dy}]$ are calculated in the same way as for the Slope.

$$Aspect = \frac{180}{\Pi} * ATAN2([\frac{dz}{dy}], -[\frac{dz}{dx}]) \tag{3.7}$$

The aspect reveals patterns related to the orientation of the slope and common characteristics like sun exposure, wind impact or structural controls where preference planes of erosion are created (figure 3.18). The most common orientation of the slopes is N-NE; followed by orientations S-SW; however, those orientation present a low landslide density; in contrast, orientation S-SE are characterized by a high landslide density as well as those NW-N (figure 3.19). On the other hand, the spatial correlation between the slopes orientation and the landslide occurrence is almost null, from where is possible to conclude that there is not a preference slope orientation where landslides tend to occur more

frequently than others in the study area.

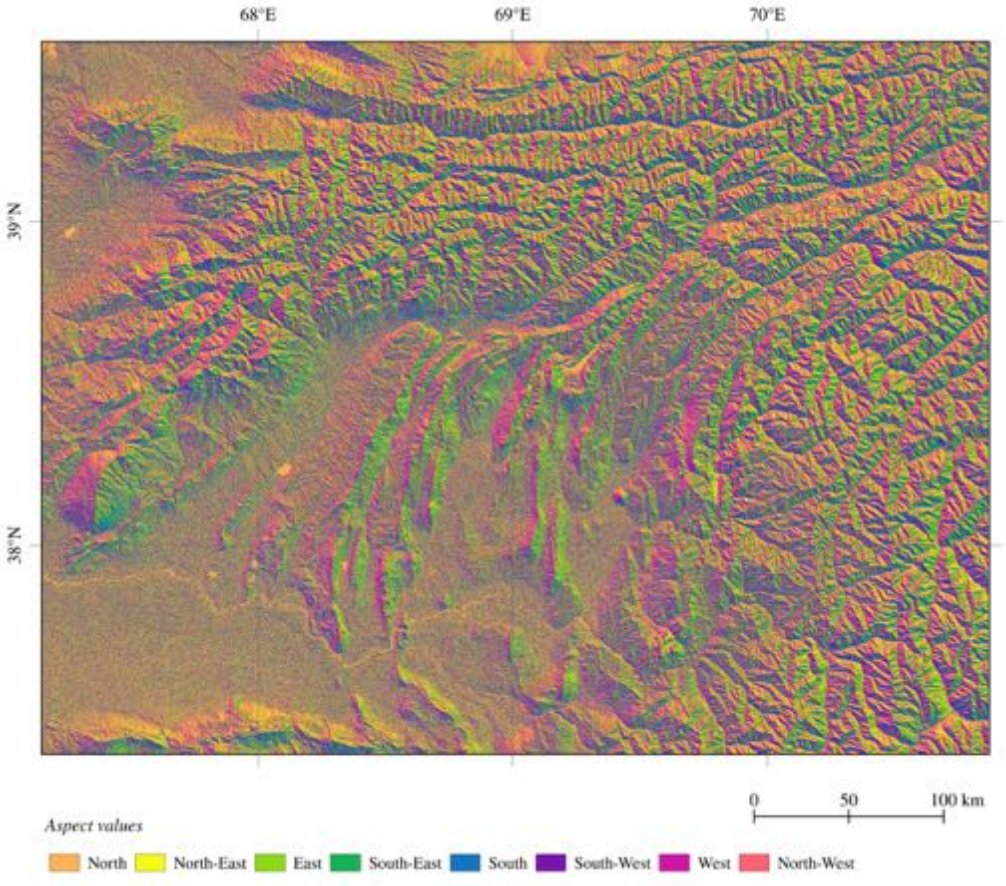


Figure 3.18: *Map of the distribution of the orientation of the slopes (Aspect).*

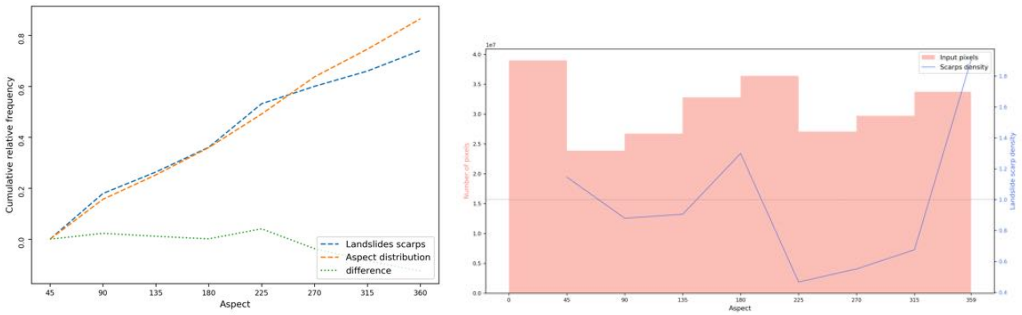


Figure 3.19: *Spatial correlation between the Aspect and landslides. Left: Cumulative relative frequency. Right: Histogram of the number of pixels per class compared to the landslide density per class.*

3.11 Surface Roughness

The surface roughness (Smith, 2014) is expressed as the ratio between the relief (TS = Topographic surface) and a flat surface (FS = Flat surface) as it is given by equation 3.8 (Shahzad & Gloaguen, 2011). SR is commonly used to describe landslide activity because it is related to both landslide mechanics and features (Grohmann *et al.* , 2009). The SR values are close to 1 for flat areas and increase rapidly as the real surface becomes irregular (more dissected by the drainage network). A kernel size of 1000m is used, to have a smaller detail in the information regarding the surface of the slopes.

$$SR = \frac{TS}{FS} \tag{3.8}$$

The surface roughness exhibit a strong negative spatial association for low values as it is shown in the figure 3.21. However, for values higher than 1.2, the landslide density is constantly higher, until reach a peak at 1.9.

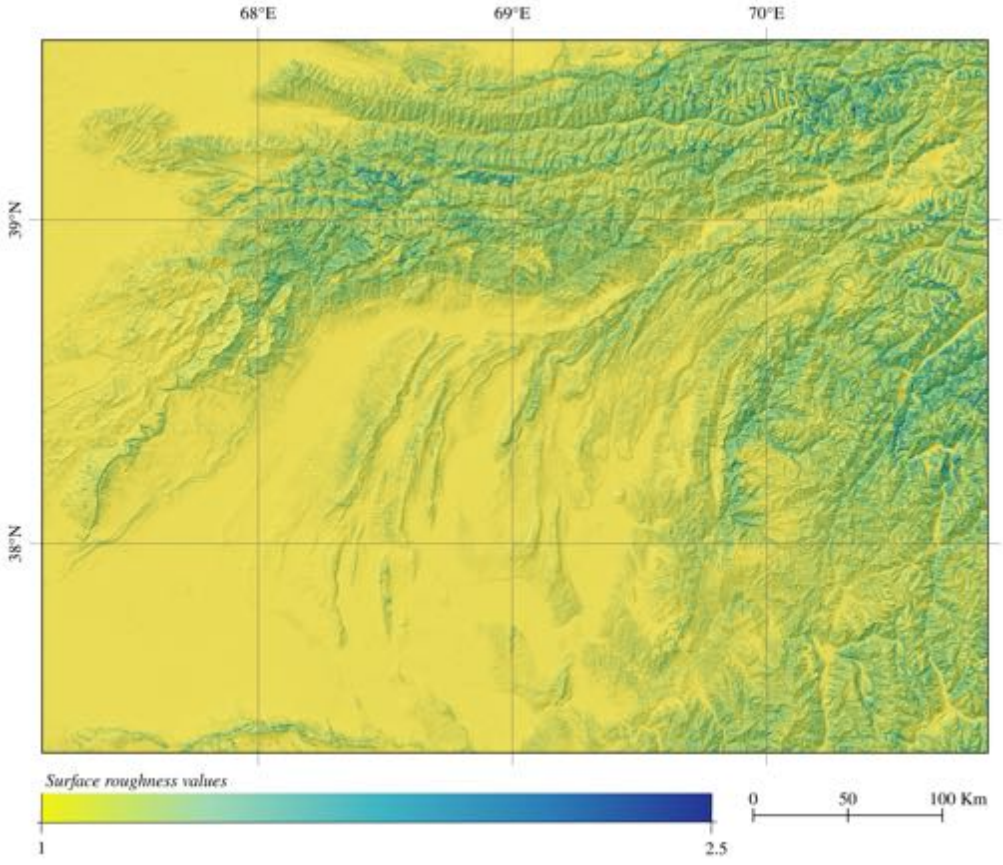


Figure 3.20: *Map of the distribution of the elevation relief ratio.*

The surface roughness enhances the areas that differ from a flat surface. Higher values of SR are associated in small scale to landslides areas because of the deformation of the terrain caused by the material movement, whereas, for large scales, it is related to erosive surfaces. The area is characterized by high surface roughness in the Panj river

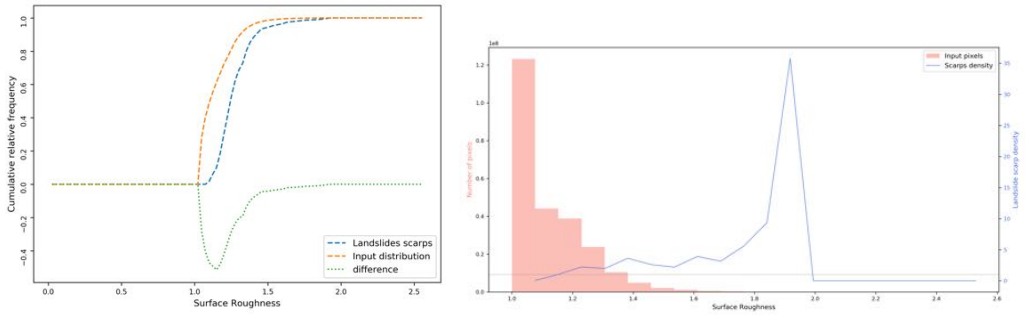


Figure 3.21: *Spatial correlation between the SR and landslides. Left: Cumulative relative frequency. Right: Histogram of the number of pixels per class compared to the landslide density per class.*

valley as well as in the south of the Zaravshan valley in the western zone of the Tien Shan. Also, the source area of the Zaravshan river is enhanced by high values in the SR (figure 3.20).

3.12 Elevation Relief Ratio

The elevation relief ratio (Pike & Wilson, 1971; Strahler, 1952; Schumm, 1956) is considered as an indicator of the cycle of erosion, defined as the total time required for a landscape to reach the base level. It allows discriminating between areas that are considered stable or unstable in terms of landscape evolution. The elevation relief ratio (ERR) express the relative portion of upland to lowland areas, and it is calculated as it is shown in equation 3.9 based on the height values (h). High values of the ERR are possibly related to young active tectonic areas and low values are related to older landscapes that have been more eroded and less impacted by recent active tectonics. A kernel size of 20000m is used to the calculation with the aim of understanding regional tectonic influences.

$$ERR = \frac{\overline{h} - h_{min}}{h_{max} - h_{min}} \tag{3.9}$$

The lower values of ERR enhance the areas that are considered old landscapes where just a few erosional processes are involved. Those areas are located in the Tadjik depression and mainly associated with the flat areas where fluvial deposition is the dominant process (figure 3.22). Commonly, values between 0.3 to 0.6 (Strahler, 1952) are considered as stable landscapes, this means that river incision and the slope are adjusted to each other, so the landscape will not change substantially through time, while values higher than 0.6 are related to young landscape.

The spatial association between the ERR and the landslide occurrence is negative for values below 0.45 (figure 3.23) associated to landscape dominated by V-shaped valley or steep scarps where less landslides occurred. Contrary, values above 0.45 are characterized by a high landslide density and a slightly positive spatial correlation (figure 3.23

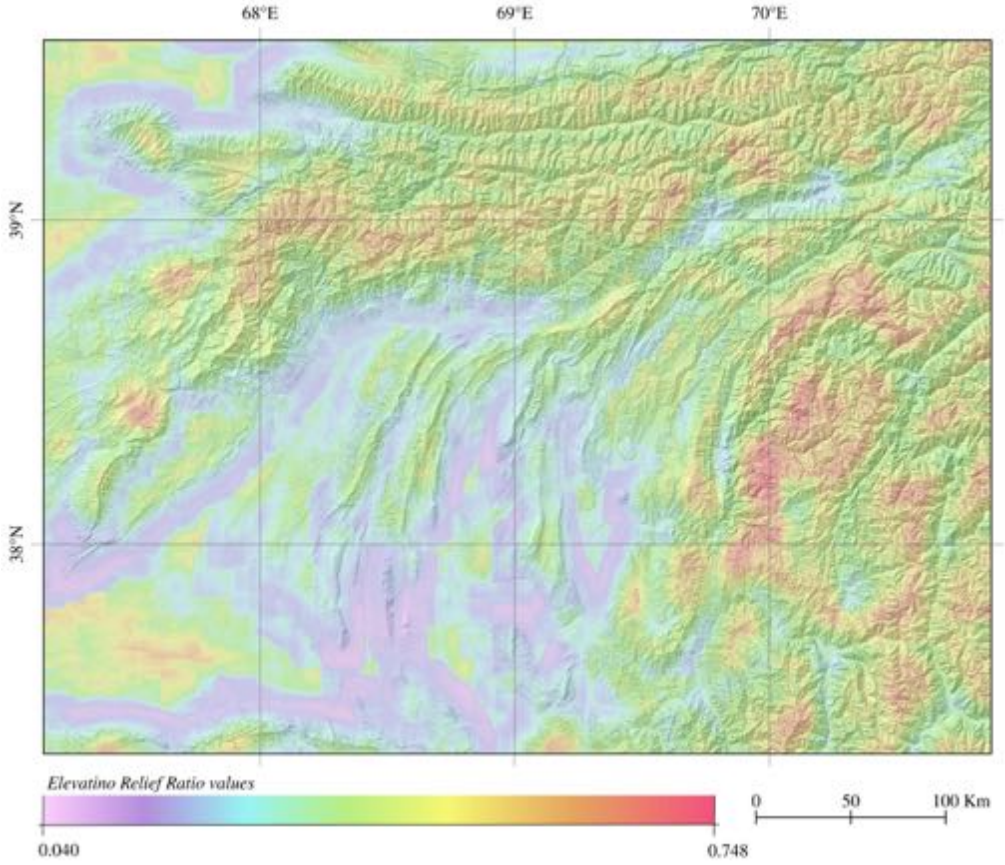


Figure 3.22: *Map of the distribution of the elevation relief ratio.*

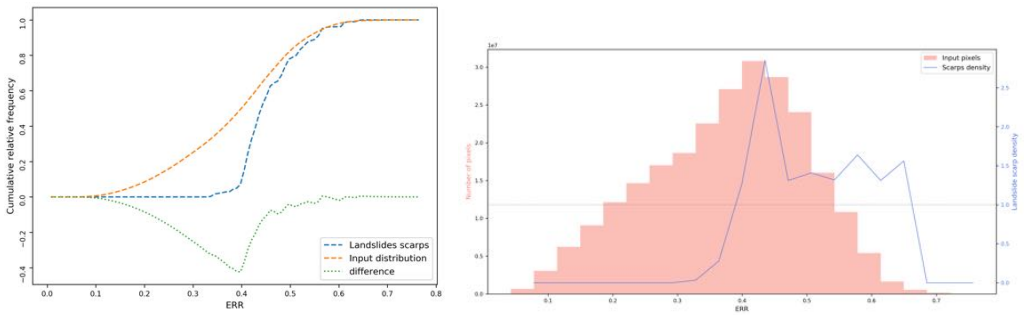


Figure 3.23: *Spatial correlation between the ERR and landslides. Left: Cumulative relative frequency. Right: Histogram of the number of pixels per class compared to the landslide density per class.*

because the drainages are entrenched and strong topographic scarps are finding the equilibrium state. The areas characterize by very high ERR values (> 0.6) are mainly located in the Pamir areas and represent plateau areas incised by depth rivers like the Panj. The high of the plateaus decrease to the west from 4500 m. a.s.l to 3000 m a.s.l. A small residual plateau is identified in the south of the Tien Shan with an average high of 4000 m a.s.l; however, it is hardly dissected and non-continuous. Values below 0.2 represents flat areas or very gentle scarps, describing the Tadjik basin.

3.13 Surface Index

A combination between the SR, ERR and the elevation of the area is proposed by Andreani *et al.* (2014) as surface index. This index map simultaneously preserved and eroded portions of an elevated landscape because, the elevation relief ratio is sensitive to elevated surfaces and poorly eroded scarps, while surface roughness identifies areas with dissection by the drainage network. The computation of the index is presented in equation 3.10. Positive values of SI are mainly associated with poorly incised surfaces (landscape characterize by a high hypsometric integral or elevation relief ratio and low surface roughness), whereas, negative values are associated with dissected landscapes (high surface roughness).

$$SI = \left(\frac{ERR - ERR_{min}}{ERR_{max} - ERR_{min}}\right) + \left(\frac{h - h_{min}}{h_{max} - h_{min}}\right) + \left(\frac{SR - SR_{min}}{SR_{max} - SR_{min}}\right) \tag{3.10}$$

The surface index separates poorly incised surfaces from those characterized by high dissection. The North and West part of the study area is characterized by a dissected landscape represented by negative values of SI. The most negative areas are located in the valley of the Panj river as well as Easter to the source of the Zerafshon river, however, the more abundant areas present values close to 0. The positive values are associated with poor incised surfaces located first in the Tadjik basin but also, in the Pamir as well as in few areas of the Tien Shan (figure 3.25).

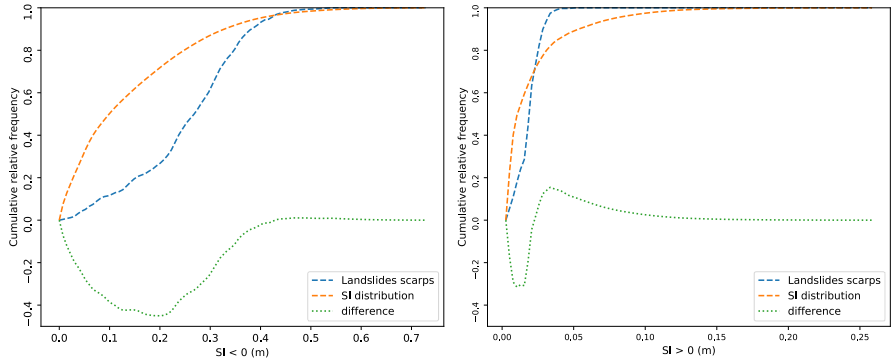


Figure 3.24: *Spatial correlation between the SI and landslides. Cumulative relative frequency*

The negative values close to 0 of SI are strongly negative associated to the landslide occurrence (figure 3.24); however, below -0.4 a high landslide density is observed (figure 3.26). Similarly, the positive values are characterized by a strong negative association until 0.03, where a slightly positive association is presented, nevertheless, the positive values are identified with low landslide density.

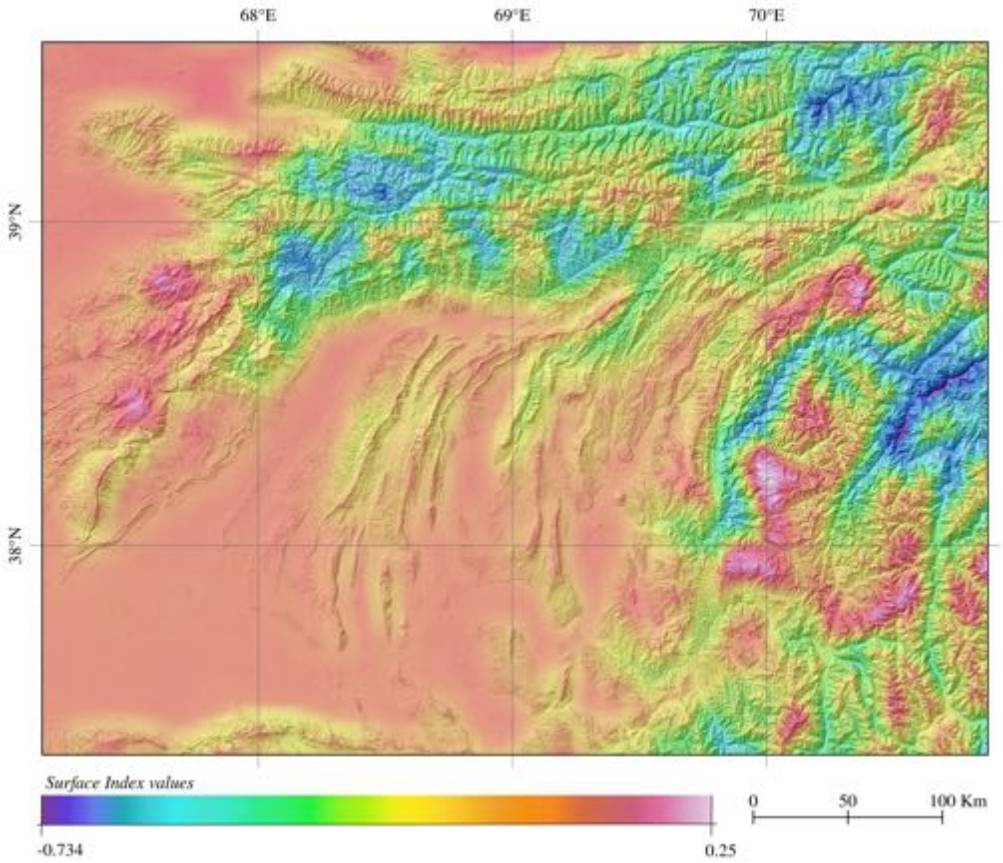


Figure 3.25: *Map of the distribution of the surface index.*

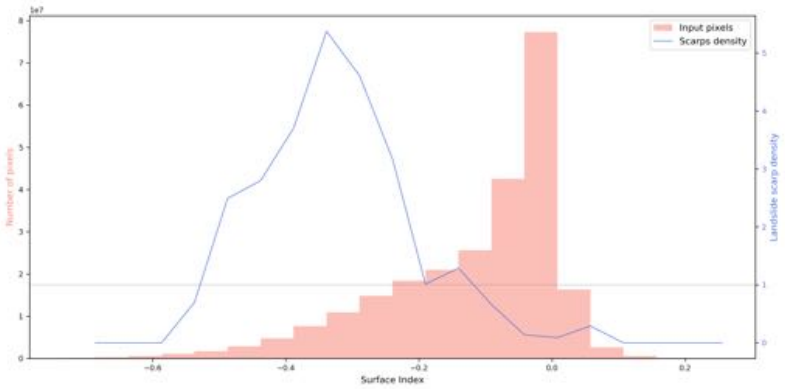


Figure 3.26: *Histogram of the number of pixels per class in the SI compared to the landslide density per class.*

3.14 Local Relief

The local relief (Ahnert, 1984) is the difference between the highest and the lowest elevation in a specific area (equiation 3.11). In general, the local relief is calculated for a watershed in order to detect tectonic influences and/or incision. The watershed also can be defined as kernel areas for large areas. A kernel size of 10000m is used based on the size watershed of the main rivers like the Panj and the Zeravshan.

$$LR = h_{max} - h_{min} \tag{3.11}$$

The changes in elevation are shown in the local relief map (figure 3.27). The Tadjik basin is characterized by a low local relief where small changes in elevation take place. In contrast, the higher values of local relief are located in the mountain ranges.

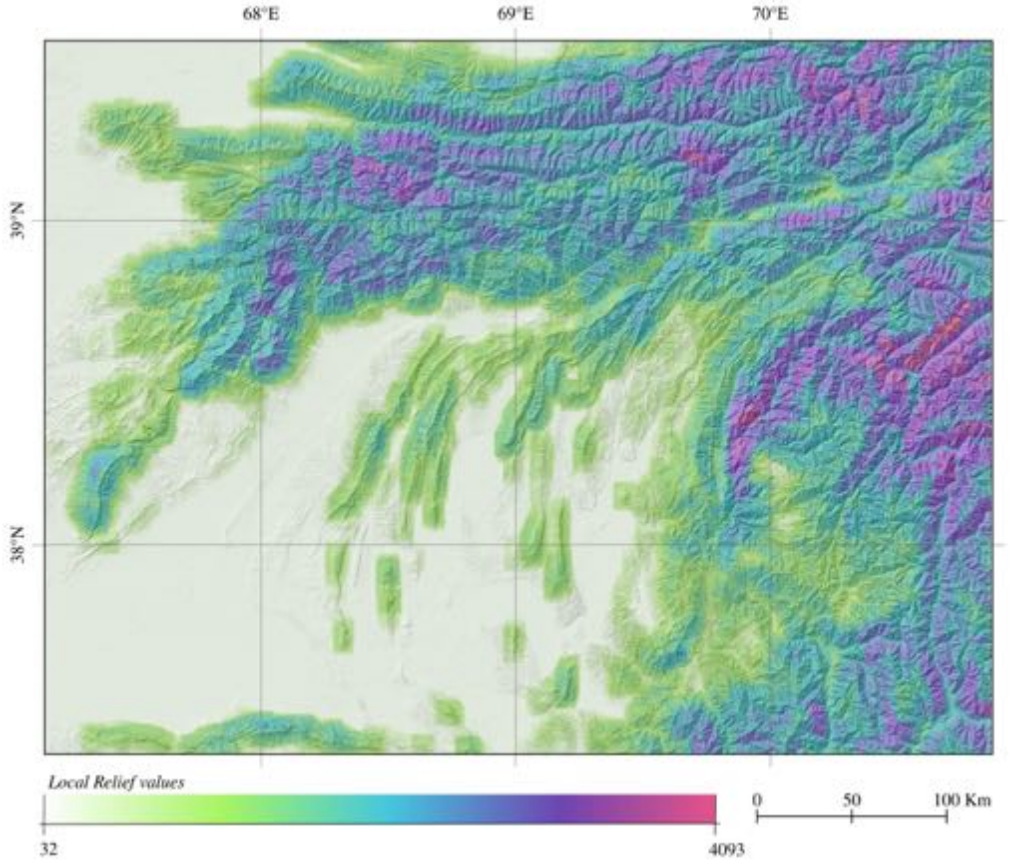


Figure 3.27: *Map of the distribution of the local relief.*

A negative spatial association is presented in the figure 3.28 between low values of local relief and the landslide distribution. This association remains until values of 1800-2200m. However, the landslide density start increasing up to 1500 m , reaching a peak in at 2500 (figure 3.28).

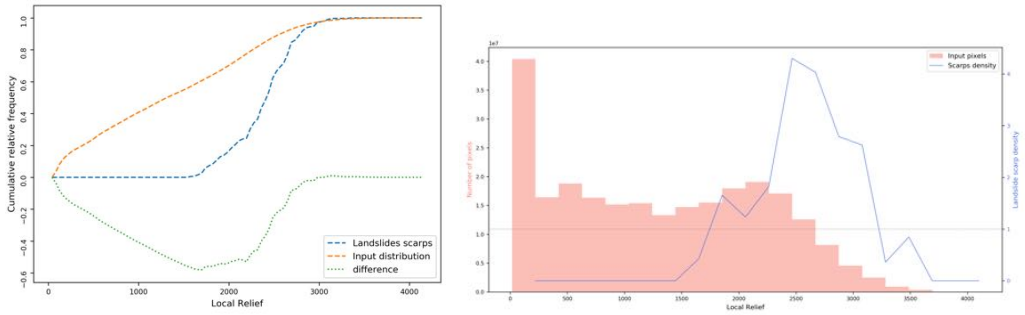


Figure 3.28: *Spatial correlation between the local relief and landslides. Left: Cumulative relative frequency. Right: Histogram of the number of pixels per class compared to the landslide density per class.*

3.15 Topographic Position Index

The topographic position index (De Reu *et al.* , 2013; Trentin & de Souza Robaina, 2018) compares the elevation of each cell in the DEM to the mean elevation of a specified neighbourhood around the cell given a predetermined radius or kernel size (equation 3.12). The range of the values of TPI depends not only on the elevation differences but also on the kernel size. Large kernel sizes will reveal major landscape units, while smaller values highlight smaller features, such as minor valleys and ridges. Positive TPI values indicate that the central point is located higher than its average surroundings, while negative values indicate a lower position than the average. For the study area, a kernel size of 10000m is used in order to obtain information related to the significant landscape units in the area.

$$TPI = h_{centralPoint} - \overline{h}_{surroundings} \tag{3.12}$$

The TPI enhance the bottom of the valleys with negative values and the peaks with positive values (figure 3.29). The depth of the major valleys in the Tien Shan and the Pamir are similar. Also using the TPI is possible to identify areas of discharge when the depth of the valley change. It is evident in the south-east of the area; where three main tributaries create an shallower intra-mountainous basin. The values near 0 dominate the area, where very low landslide density is associated. A bimodal distribution of the landslide density is observed (figure 3.30). Characterized by a negative association for values near to 0 until -300 and 800 (figure 3.31), where the maximum landslide density is reached. This indicate that the landslides tends to occur either in the bottom of the valley or near the top of the ridges.

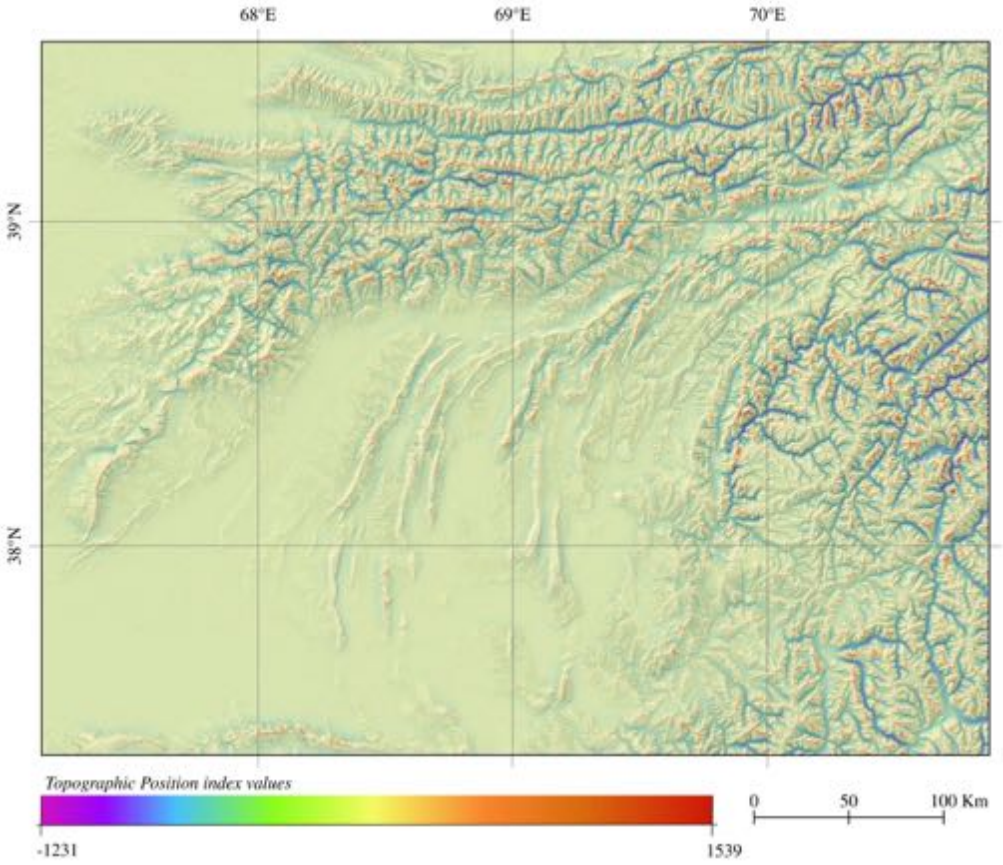


Figure 3.29: *Map of the distribution of the topographic position index values.*

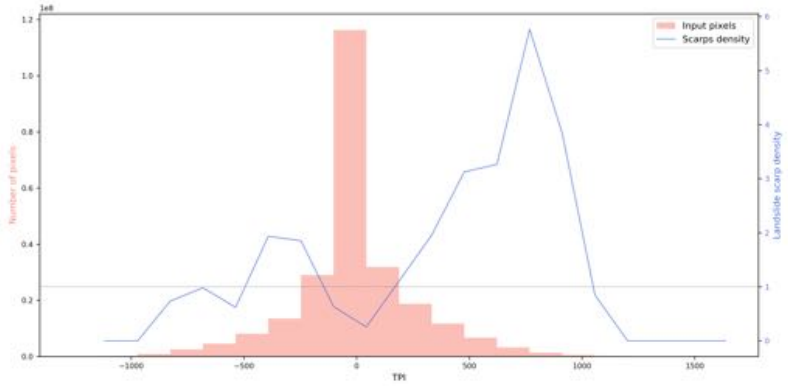


Figure 3.30: *Histogram of the number of pixels per class in the TPI compared to the landslide density per class.*

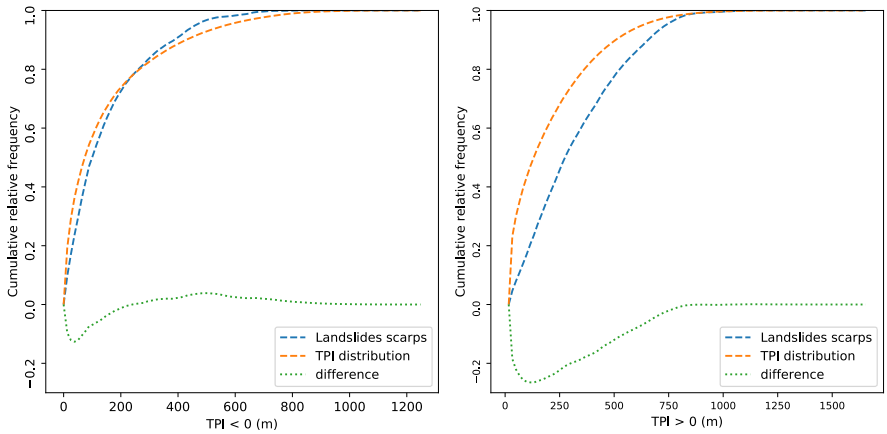


Figure 3.31: *Spatial correlation between the TPI and landslides. Cumulative relative frequency*

3.16 EigenValues

The eigenValues is an approach to analyse the curvature of an image (Frangi *et al.* , 1998) and it has been extrapolated to the landscape. It is calculated based on the Hessian matrix, a 2x2 matrix composed of the second partial derivatives of the elevation values. The calculation is made based on a convolution with derivatives of a Gaussian filter at a scale σ. The result is a matrix with the principal directions in which the curvature of the landscape can be decomposed and its magnitudes represented by the eigenvectors $|\lambda 1| \leq |\lambda 2|$. For the implementation in geomorphology, the $\lambda 2$ is used since the magnitude increase as the local curvature of the feature increase and topographic features without a preferential direction will have low magnitudes.

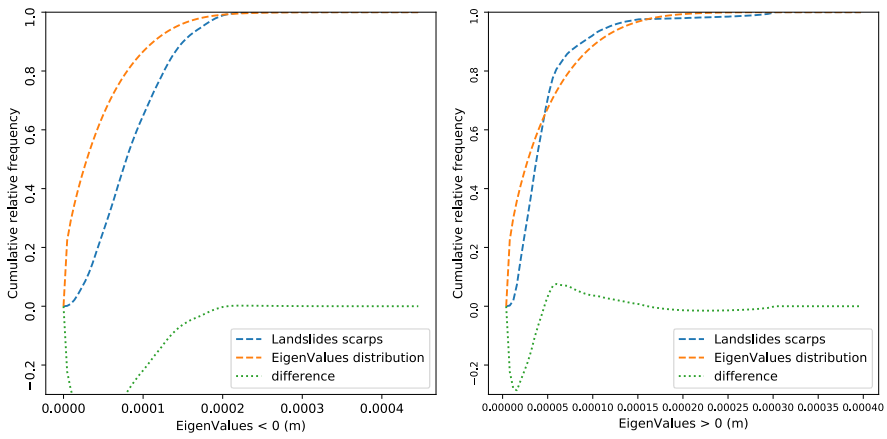


Figure 3.32: *Spatial correlation between the EigenValues and landslides. Cumulative relative frequency*

The Eigenvalues allows to discriminate the ridges in negative values (negative curvature) and the bottom of the valley with positive values (positive curvature) (figure 3.33). The higher the curvature, the more extreme the value will be. Values close to 0 represent

low curvatures.

A strong negative spatial association is observed for the negative values above -0.0002, meaning that wide ridges are less associated to the landslide occurrence (figure 3.32). The positive values are strongly negative spatial associated for values below 0.00005 and a slightly positive association is observed for the value above characterized by a slightly positive landslide density (figure 3.34).

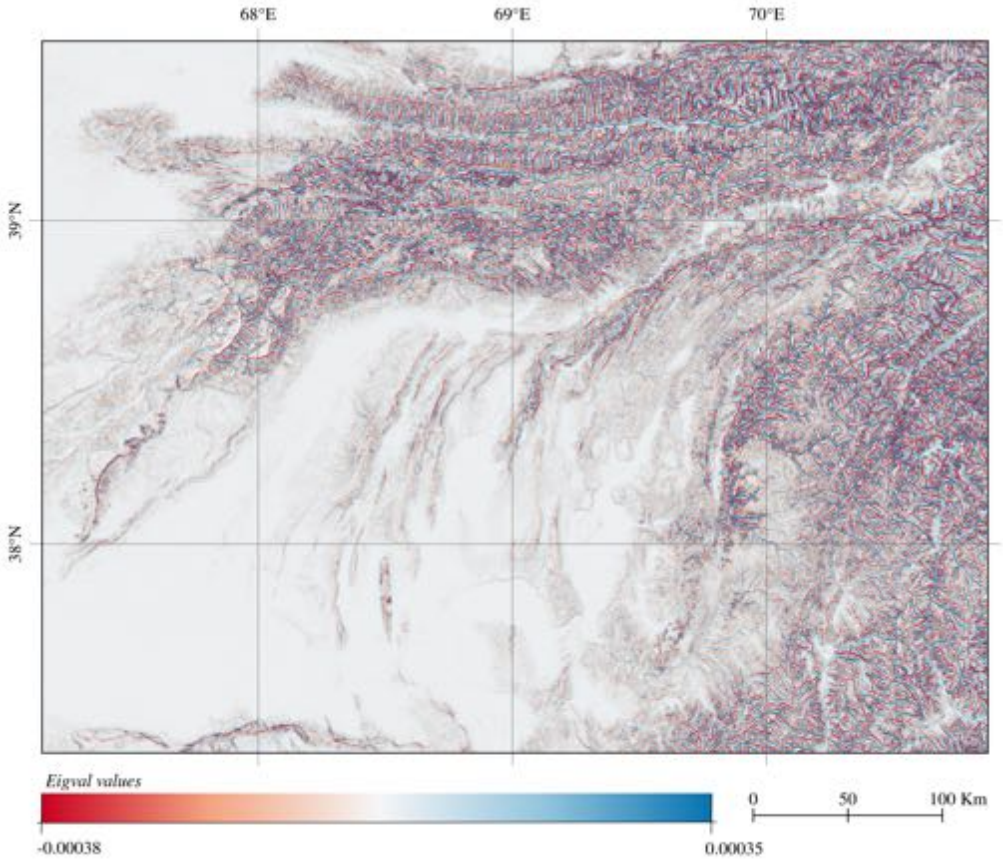


Figure 3.33: *Map of the distribution of the Eigenvalues.*

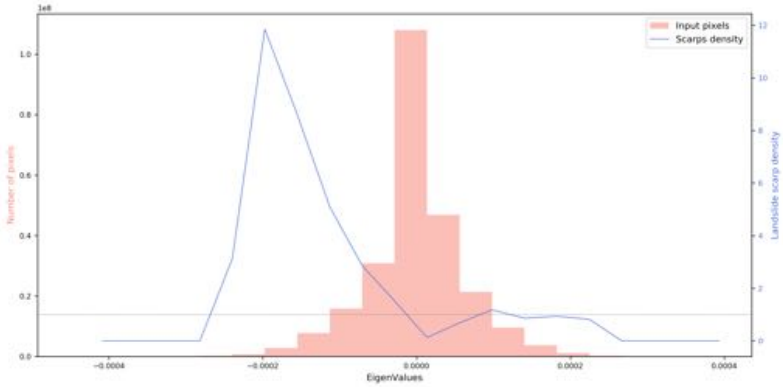


Figure 3.34: *Histogram of the number of pixels per class in the EigenValues compared to the landslide density per class.*

3.17 Distance to fault

A reliable catalogue of active faults was created as part of the Central Asia Fault Database (CAFD) from Eberhard Karls Universität Tübingen and the University of Montana. The catalogue includes in total 1196 faults with detailed information and references.

The source of the information is a collection of previously published literature and databases in the area. One of the primary sources for this database is the HimaTibetMap which is an open-source digital database of active faults located in the Indo-Asia collision zone. It is created based on field observations and interpretations of satellite images and global digital topography.

The CAFD is considered as a database with high thematic accuracy and consistency; based on the fact that the data collection includes fieldwork. Also, the data has a high completeness because first, part of the information was collected by field work and second, each fault contains relevant information like fault name, sense of movement, references and variations in fault name or location if they exist. The authors claim that the fault spatial accuracy depends on the scale of observation used in previous investigations.

The CAFD catalogue is improved using a digitalization of the faults from the paper maps (Federal State Budgetary Institution A.P. Karpinsky Russian Geological Research Institute (FGUP VSEGEI), 2018) provided by Ratschbacher (2018 - personal communication). The faults are rasterized using a 30 meters pixel-based and the Euclidean distance from each fault to the pixels in the raster is calculated.

A crucial concentration of landslides is identified in the valley of the Zerafshon where a large number of active faults are located. On the other hand, small but abundant landslides are located along the Girssar-Kokshal Fault and the Vakhsh Trust System (figure 3.35). Finally, the Darvaz Fault, another of the most essential thrust system in the area is surrounded by small landslides. The remaining thrust systems in the Parmir are not associated with landslide occurrence. Nevertheless this can be interpreted as a limitation of the catalogue rather than the lack of mass wasting.

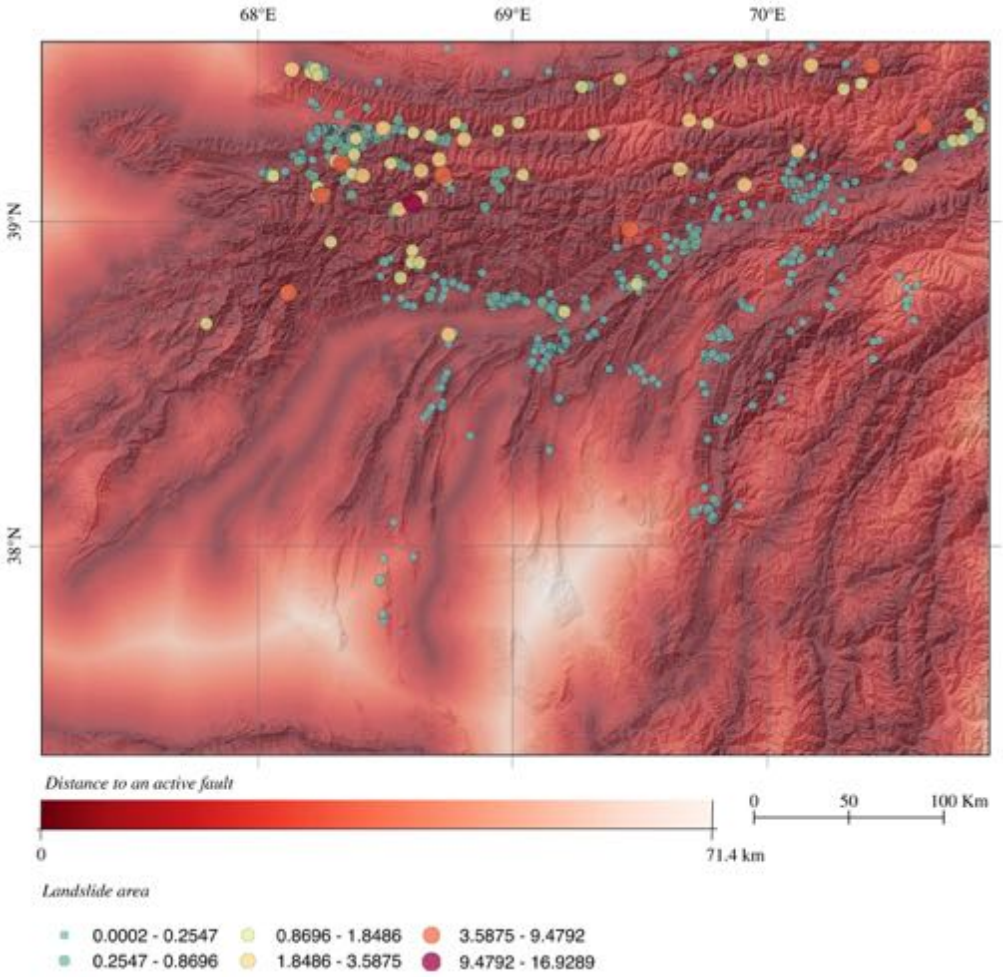


Figure 3.35: *Map of the distribution of the distance to faults in the area and the size of the landslides.*

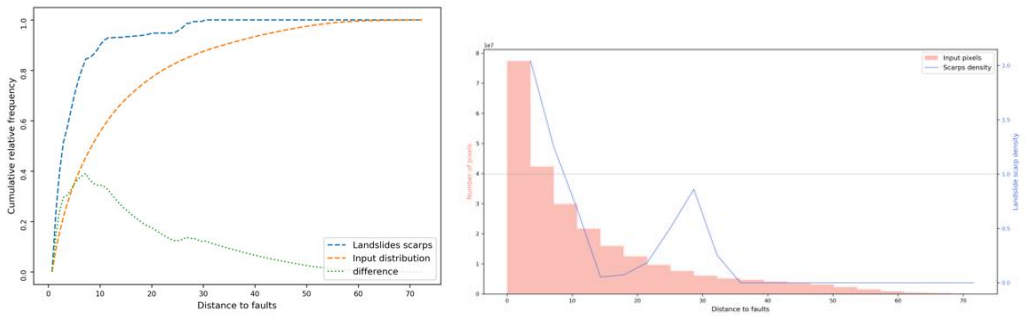


Figure 3.36: *Spatial correlation between the Fault distance and landslides. Left: Cumulative relative frequency. Right: Histogram of the number of pixels per class compared to the landslide density per class.*

There is a strong positive spatial association between the distance to a fault and the landslides (figure 3.36) that start decreasing in less than 10km, but remains slightly posi-

tive until the farthest point. In terms of landslide density, areas located in less than 10km from faults have a high landslide density. Another smaller peak is observed for areas between 25 and 35 km .

3.18 Seimozones

The area of study is considered as one of the most active areas in terms of seismicity in the world; however, it is a lack of instrumentation like accelerometers that allows the calculation of the peak ground acceleration(PGA); the ideal parameter to analyse how shaking during an earthquake affect the materials in the surface; however, probabilistic seismic hazard assessment studies are another source to obtain PGA. The Global Seismic Hazard Assessment Program (GSHAP) was launched in 1992 by the International Lithosphere Program (ILP) with the support of the International Council of Scientific Unions (ICSU) in order to mitigate the risk associated with the earthquakes events. The area of study is included in the Northern Eurasia region and the seismic hazard assessment was performed by Ulomov *et al.* (1999). The expected peak ground acceleration with 10% of exceedance probability in 50 years was computed in terms of expected Medvedev-Sponheuer-Karnik scale (MSK) intensity and then transformed to PGA by an empirical relationship. The results of the GSHAP project are coarse an even though site approach methodology has been implemented in order to improve the results (Bindi *et al.*, 2012) they do not cover with consistency the whole area of study. However, the Earthquake Model Central Asia (EMCA) project, created an update earthquake catalogue used to improve the GSHAP. The earthquake catalogue (Natalya Mikhailova, 2015), as well as the seismozones used to improve the GSHAP (Shahid Ullah, 2015) are available as shapefiles.

The EMCA project computed **seismozones** in the area in order to improve the earthquake susceptibility assessment. The seismozones are characterized for the presence of unique earthquake conditions not only related to the seismo sources but also to the earthquakes associated with them. 14 of those seismozones cover the area; however, most of them are characterized by very low to not landslide density. All the seismo zones are related to active shallow crust tectonic regiment; however, the maximum magnitude expected for the areas 3 (Tien Shan) and 2 (South-east corner) are higher than the expected from the rest of the area. In term of depth of the associated earthquakes, some deep earthquakes are located in the seismozones 2, 5 and 12 in the south-east of the area, as well as the predominance of intermediate earthquakes.

On the other hand, the north of the Pamir and the Tien Shan are associated to shallow earthquakes. The density of earthquakes is also different from each seismozone (figure 3.37). The seismo zone 2, 5 and 12 present the most number of earthquakes, followed by the seismozones 6 and 14 in the MPT area. Lower earthquake densities are located in the seismozones 3, 8 and 13, while just few earthquakes are located in the others seismozones.

The seismozones 3, 4, 6 present high landslide density (figure 3.38), however, the sesimoszones 13 and 14 are also related to landslides, even though they are from a small size than the ones located in the area or 6.

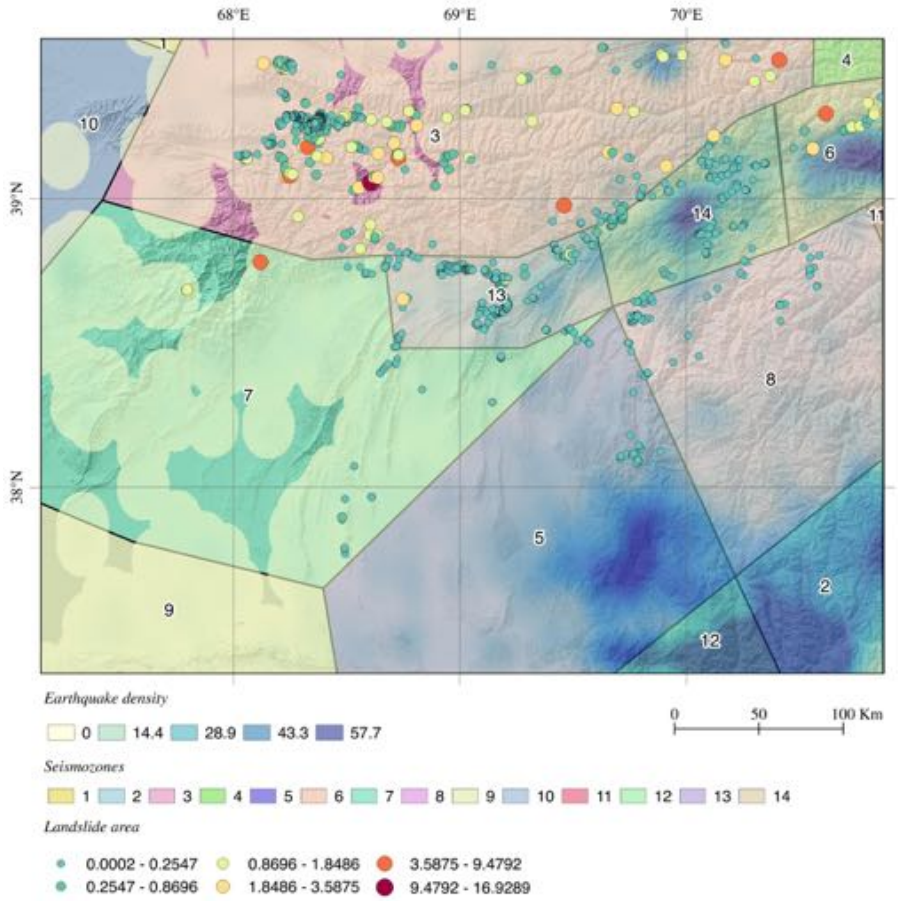


Figure 3.37: *Map of the distribution of the seismozones in the area, landslide distribution by size and earthquake density obtained from the EMCA project.*

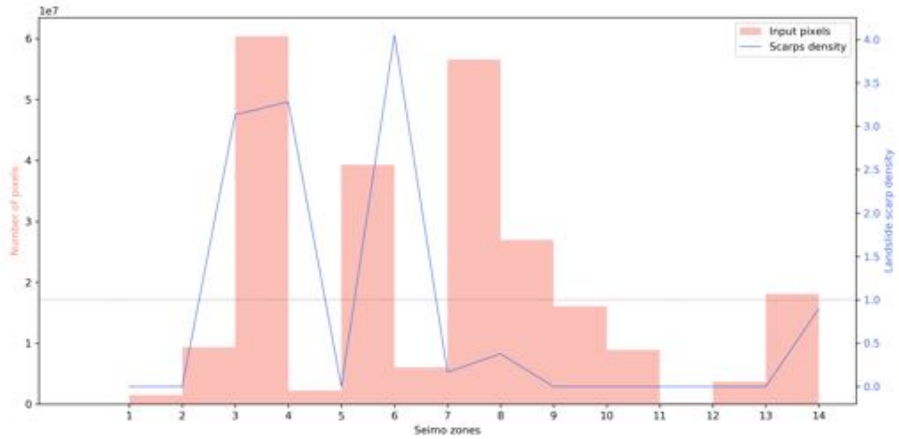


Figure 3.38: *Histogram of the number of pixels per seismozone compared to the landslide density per class.*

Chapter 4

LANDSLIDE SUSCEPTIBILITY MODELS

4.1 Introduction

The golden rule in geomorphology is "the present is the key to the past." The law assumes that the effects of geomorphic processes seen in action today can be used to infer the causes of assumed landscape changes in the past (Huggett, 2016). Similarly, the landslide susceptibility models assumed that the landslides are going to occur under similar conditions to those that occurred in the past.

Landslide susceptibility (LS) is defined as the likelihood of occurrence of a landslide in an area given certain local conditions (Brabb, 1985). It predicts "where" landslides are likely to occur (Guzzetti *et al.*, 2006); however, does not consider "when" or "how frequently" and nor the magnitude of the expected landslide. Mathematically, LS can be defined as the probability of spatial occurrence of slope failure basis on certain conditions (Chung *et al.*, 1999).

Before the selection of the landslide susceptibility model, the choice of an appropriate terrain subdivision or mapping unit is mandatory to obtain a successful result. The mapping unit can be selected based on different criteria. It can be pixels size, slope units or unique conditions(Reichenbach *et al.*, 2018).

Pixels or grid cells are the most common mapping unit reported in the literature because they are simple to process at all geographical scales as well as the availability of GIS software to transform vector data to raster and manipulate them. However, landslide's shapes differ a lot from a square grid cell. Additionally, the different parameters of the landscape that are related to landslide occurrence are representative at different scales, hence, coarser or finer pixel size. On the other hand, potential statistical problems could be associated with the percentage of area covered by landslides vs the area without landslides, as well as, the difficulty in interpreting some results that predict stable pixels surrounded by unstable pixels (Reichenbach *et al.*, 2018).

Slope units are a partition of the territory into hydrological regions bounded by drainage and divided lines (Carrara *et al.*, 1991). Because landslides occur primary on slopes, this mapping unit results suitable to understand the processes that interact to trigger a landslide or a group of landslide in a detailed way. The size of the slope will be determined

by the type and size of landslide, being particularly relevant for the morphometric variables. Thereby, the presence of landslides is described using the overall geometry of a slope and allows the use of very detailed DEM and their derivatives. However, even the slopes are easy to identify in the field as well as in topographic maps, they are difficult to discriminate, particularly for large areas (Reichenbach *et al.*, 2018).

Unique conditions units are obtained by intersecting all the thematic variables considering important for the susceptibility modelling by a simple operation in GIS software. They were adopted primarily by investigators that used Bayesian approaches to model landslide susceptibility based on the simplicity of its calculation; however, a larger number of terrain units reduces the size of the units, and their representativeness for landslide susceptibility, creating some challenges in the overlying process based on the need to avoid many units of very small size. This process is challenging because of the need to classify the continued variables into reasonably numbered classes along with vector data having many small polygons and digitalization errors.

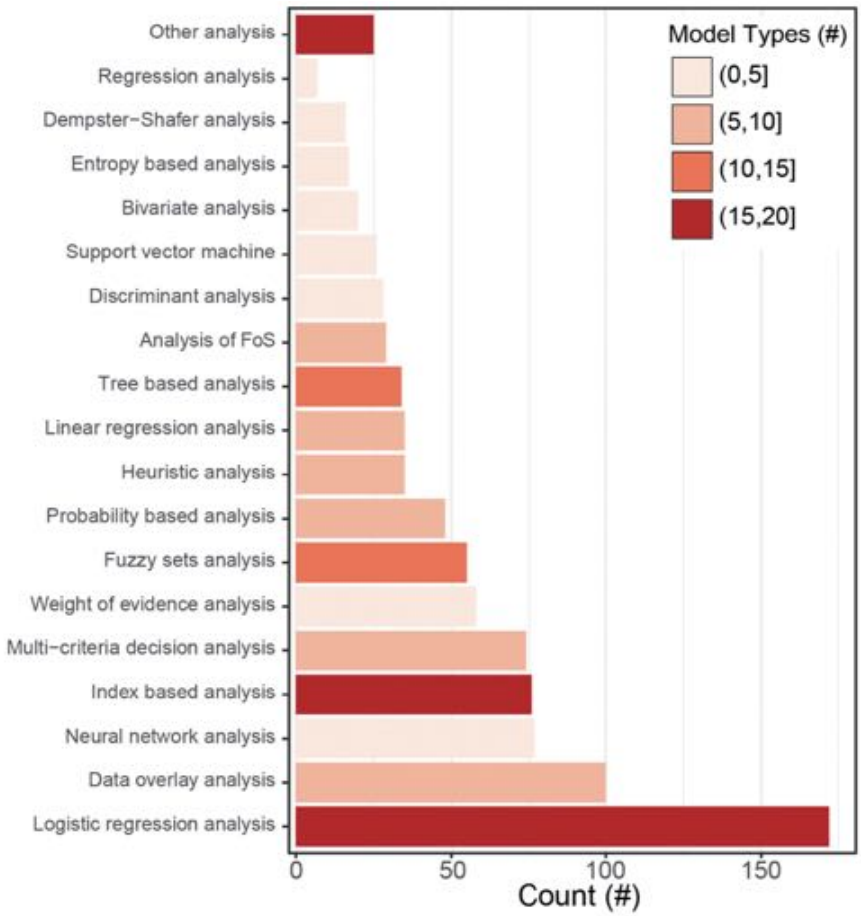


Figure 4.1: *Horizontal bar chart showing the count of 19 model types classes used to group the 163 model names given by the authors in the literature database. Darker colours indicate a large number of single models in the group. source:* Reichenbach et al. (2018).

Several approaches have been proposed for modelling landslide susceptibility and they are independent to the mapping unit selected. Three main groups of methods can be identified, the first is the heuristic method, based on the geomorphological expert deci-

sion and assigning of lever of susceptibility (Barredo *et al.*, 2000); Secondly, Qualitative or semi-qualitative methods that combine thematic layers using decision support tools with approaches like analytical hierarchy process (AHP) (Kayastha *et al.*, 2013; Pourghasemi *et al.*, 2012; Boroumandi *et al.*, 2015; Yalcin *et al.*, 2011) , weighted linear combination model (Akgun *et al.*, 2008) or Multi-criteria decision (MCDA) (Akgun, 2012). The third and bigger group is the deterministic method that is based on mathematical modelling and includes statistical approaches as well as machine learning.

During this study, statistically-based approaches are adopted. The statistical approaches are indirect methods and based partly on the field observations and expert knowledge and partly on statistical computation. The computations determine the importance of a factor based on the observed relationship with landslides as well as assign weight or probabilities of occurrence of a landslide (Regmi *et al.*, 2010). The most common methods for landslide susceptibility modeling include logistic regression, neural network analysis, data-overlay, index-based and weight of evidence analysis. Reichenbach *et al.* (2018) based on a literature review analysis, determined that the most commonly used method is logistic regression with an important increase since 2005.

A general methodology for the landslide susceptibility assessment consist in: 1) the data collection of the landslide catalogue and the thematic variables ,2) data preparation according to the landslide model to be implemented and 3) the implementation of the model.It is essential to define the type of data required (discrete or continuous) and to differentiate between the training data set and the validation data set. After that, the implementation of each landslide susceptibility model is performed based on their individual workflows. Finally, the model is evaluated and discussed (figure 4.2).

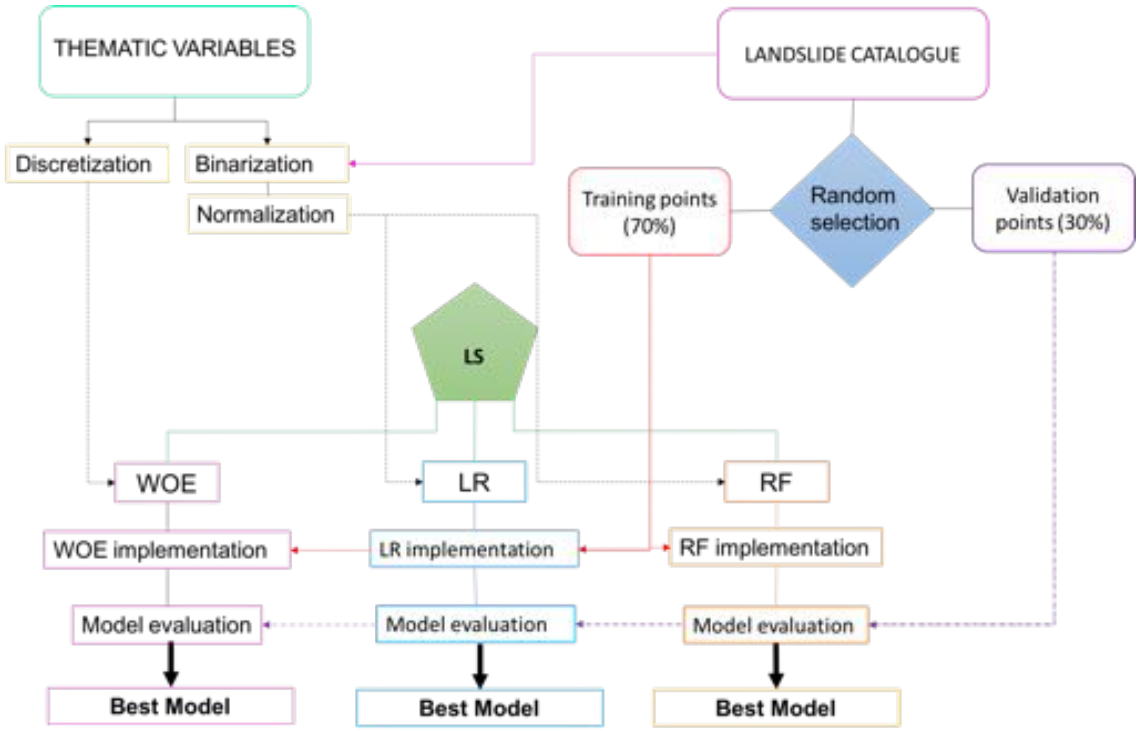


Figure 4.2: *General workflow of the landslide susceptibility assessment.*

4.2 Data Preparation

4.2.1 Preparation of training and validation data set

The training and validation data are obtained from the landslide catalogue. First, the polygons are rasterized using a pixel's size of 30 meters with a unique id per landslide. Second, the upper pixels of each landslide are since the interest of the study is the characteristics of the source area rather than the depletion zone. This method warranty that the training and the validation information correspond to the instability conditions.

The validation of the resulting landslide susceptibility model is made based on an independent dataset from the data used to compute it; however, an independent dataset is not available for the area. To overcome this difficulty, the landslide catalogue is divided into two datasets using a random selection method. 70% of the landslides in the catalogue are used as training points to perform the models; while, the remaining 30% is designated as validation points and is used in the calculation of the receiver operating characteristic curve (ROC) and the area under the curve (AUC). It is important to point, that even though a single landslide can be represented by more than 1 pixel; the random selection is made based on unique labels for each landslide in order to avoid multiple sampling of the same landslides and introduce biases.

Additionally, the logistic regression and the random forest approaches required pixels without landslides to be included in the training dataset. For this purpose, a random selection of a number of pixels equal to the number of pixels with landslides among the unaffected areas is made.

4.2.2 Preparation of thematic variables

In order to prepare the different variables to the requirements of the landslide susceptibility approaches, two main pre-processing must be done. The first one is the discretization of the continuous variables for the implementation of the weight of evidence (WOE) approach; while, for the logistic regression (LR) and random forest approaches (RF), a normalization and binarization of the variables are needed.

4.2.2.1 Discretization

Discretization of continuous variables is a difficult and tricky procedure because of the introduction of bias or overfitting parameters in the model. In most of the literature where the WOE method is implemented, the categorization of continuous values like those derivatives from DEM, are made based on expert knowledge; however, (Regmi *et al.* , 2010) proposed the identification of breaking point based on the variation in weight contrast values (equation 4.12) within the variables. Nevertheless, one of the methodological problems of the WOE is the performance when a very few pixels of landslides are present in a given variable class. This problem was faced by (Regmi *et al.* , 2010), assigning a zero weighted value to the class or combine the class with other classes; a methodology that increases the processing time. In order to avoid the existence of very few pixels or landslides per variable class; an alternative classification based on an equal number of pixels per class is implemented and compare with the results of the last two approaches.

4.2.2.2 Binarization

For LR and RF; the landslide catalogue is used as a dependent variable, taking values between 0 to 1; where 0 is an absence of landslide and 1 represents presence. However, before relating the dependent variable to the independent variables; a normalization of all thematic variables in the range of -1 to 1 is performed.

The categorical data (lithology) is transformed to a continues variable based on the weights obtained from the WOE analyse, and then, standardized as the other continuous variable. This procedure aims to decrease computation power and number of inputs. Otherwise, a single input should be created per lithological category.

4.3 Model evaluation

There are many approaches to evaluate the performance of a model. Success and prediction rate curves, contingency table and receiver operating characteristic curve (ROC) and the area under the curve (AUC) are some examples. In order to use a single approach to evaluate all the methods implemented for the calculation of the landslide susceptibility, the ROC curve, and the AUC are selected.

The ROC curve measures the goodness of the model prediction. It is a plot of the True Positive rate (TPR) and the False Positive rate (FPR) calculated by cross-tabulation; where

$$TPR = \frac{TP}{TP + FN}$$

and corresponds to the proportion of positive data point that are correctly consider as positive, while,

$$FPR = \frac{FP}{FP + TN}$$

corresponds to the proportion of negative data points that are mistakenly considered as positive.

The AUC of the ROC varies from 0.5 (diagonal line) to 1. Values near to 1 indicate a better predictive capability of the model, while, values less than 0.7 indicate poor predictive ability.

In order to understand the variability of the model. 50/100 predictions are implemented using different training and validation dataset. Also, the standard deviation of the results is analyse to detect overfitting in the results.

4.4 Weight of evidence

4.4.1 Method

The Weight of evidence (WOE) is a non-linear statistical technique based on the log-linear form of the Bayesian probability model (Bonham-Carter, 1994). This method has been applied extensively in geology not only for the landslide susceptibility analysis but also to assess mineral potentials areas, prediction of the location of flowing wells and groundwater springs, and to determine the spatial association between mapping cliff instabilities associated with land subsidence (Regmi *et al.*, 2010; Lee *et al.*, 2002; Armaş, 2012; Tseng *et al.*, 2015, e.g.,).

For mapping susceptibility areas related to landslides, the WOE method calculates the weight for each causative factor of landslides based on the presence or absence of landslides within the area. This method assumes that future landslides will occur under similar conditions to those contributing to previous landslides as well as those causative factors remain constant over time.

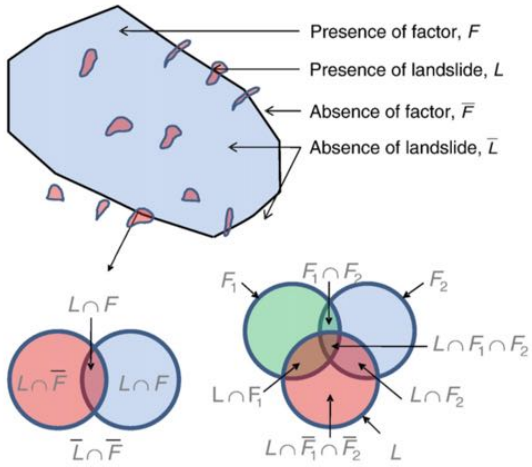


Figure 4.3: *Relation between landslides and factors used in WOE. The Venn diagrams illustrate the presence and absence of a factor(s) in relation to the landslidessource: (Regmi* et al.*, 2010)*

The WOE is based on the calculation of prior (unconditional) probability and posterior (conditional) probability. The *prior probability* is the probability of an event; It is determined by the same type of events that occurred in the past for a given period. Concerning to landslides, this can be determined by taking the ratio of the landslides in the area. The prior probability can be modified using other sources of information or evidence like the thematic variable. This revised probability of past events, based on new evidence, is called *posterior probability*, that is defined as the probability of occurrence of landslides based on a specific factor. The figure 4.3 illustrate the relationships between landslides and factors used to the calculation of the WOE.

The implementation of WOE to calculate landslides susceptibility taking into account multiple variables is explained by (Regmi *et al.*, 2010). First, the conditional probability of occurrence of a landslide given the presence of a certain factor ($PL|F$) is defined by the equation 4.1.

$$P\{L|F\} = \frac{P\{L \cap F\}}{P\{F\}} \tag{4.1}$$

Similarly, the conditional probability of existence of a factor where a landslide occur is express as $(PF|L)$ (equation 4.2)

$$P\{F|L\} = \frac{P\{L \cap F\}}{P\{F\}} \tag{4.2}$$

Because the equations 4.1 and 4.2 are the same, the conditional (posterior) probability of a landslides, given the presence of the factor can be determined as:

$$P\{L|F\} = P\{L\}\frac{P\{F|L\}}{P\{F\}} \tag{4.3}$$

This model can be also expressed in terms of odds $(\frac{P}{1-P})$ as the equation 4.4:

$$O\{L\} = \frac{Probabilitythataneventwilloccur}{Probabilitythataneventwillnotoccur} = \frac{P\{L\}}{1-P\{L\}} = \frac{P\{L\}}{\{\overline{L}\}} \tag{4.4}$$

Likewise,

$$O\{L|F\} = \frac{P\{L|F\}}{1-P\{L|F\}} = \frac{P\{L|F\}}{\{\overline{L}|F\}} \tag{4.5}$$

Dividing both sides of the equation 4.3 by $P\{\overline{L}|F\}$

$$\frac{P\{L|F\}}{P\{\overline{L}|F\}} = \frac{P\{L\}P\{F|L\}}{P\{\overline{L}|F\}P\{F\}} \tag{4.6}$$

Similar to equation 4.1 and equation 4.3, the conditional probability for the absence of landslides given a factor is:

$$P\{\overline{L}|F\} = \frac{P\{\overline{L} \cap F\}}{P\{F\}} = \frac{P\{F|\overline{L}\}P\{\overline{L}\}}{P\{F\}} \tag{4.7}$$

Substituting the value of $P\{\overline{L}|F\}$ (equation 4.7) in the right side of the equation 4.6, produces:

$$\frac{P\{L|F\}}{P\{\overline{L}|F\}} = \frac{P\{L\}P\{F|L\}}{P\{\overline{L}\}P\{F|\overline{L}\}} \tag{4.8}$$

Finally, from equations 4.4, 4.5 and 4.8, the odds of presence of a landslides given the presence of a factor F is:

$$O\{L|F\} = O\{L\}\frac{P\{F|L\}}{P\{F|\overline{L}\}} \tag{4.9}$$

Based on the previous calculations, it is possible to calculate the likelihood ratios LS (sufficiency ratio) and LN (necessity ratio). For the WOE model, the logarithm of these ratios corresponds to the positive and negative weights as it is express in equations 4.10 and 4.11.

$$W^+ = \log_e(LS) = \log_e \frac{P\{F|L\}}{P\{F|\overline{L}\}} \tag{4.10}$$

$$W^- = \log_e(LN) = \log_e \frac{P\{\overline{L}|L\}}{P\{\overline{F}|\overline{L}\}} \tag{4.11}$$

The pattern is positively correlated if the LS value is greater than 1 (W+ = positive) and LN ranges from 0 to 1 (W- = negative), on the other hand, if the patter is negatively correlated, LN would be greater than 1 (W- = positive) and LS range from 0 to 1 (W+ = negative). When LS = LN = 1 (W+ = W- = 0), the patter is uncorrelated with landslides and the posterior probability would be equal to the prior probability, that means the probability of a landslide is unaffected by the presence or absence of the factor.

Additionally, the contrast factor is calculated based on the results of the equations 4.10 and 4.11 for each class in the factor (equation 4.12). The contrast factor quantifies the spatial association between each class of a factor and the occurrence of landslides. Positive values of W_c represent positive associations between the class and the landslides occurrence, while, negative values are associated with a negative association between the class and the landslide pattern. $W_c = 0$ means that the influence of the class in the landslides occurrence is given due to chance.

$$C_w = W^+ - W^- \tag{4.12}$$

In order to include more than one factor in the model, a combination of weights of all the factors is needed. Based on the Bayes' theorem, the combination is plausible if the factors F1 and F2 are conditionally independent and it is given by the equation 4.13.

$$P\{L|F1 \cap F2\} = \frac{P\{F1 \cap F2|L\}P\{L\}}{P\{F1 \cap F2\}} \tag{4.13}$$

If F1 and F2 are conditionally independent, then:

$$P\{F1 \cap F2|L\} = P\{F1|L\}P\{F2|L\} \tag{4.14}$$

Thus, from equations 4.13 and 4.15

$$P\{L|F1 \cap F2\} = P\{L\}\frac{P\{F1|L\}P\{F2|L\}}{P(F1)P(F2)} \tag{4.15}$$

The previous equation can be formulate as well in term off odds (equation **??** and later in term of likelihood ratios (equation 4.16).

$$\begin{aligned} O\{L|F1 \cap F2\} &= O\{L\}\frac{P\{F1 \cap F2|L\}}{P\{F1 \cap F2|\overline{L}\}} \\ &= O\{L\}\frac{P\{F1|L\}P\{F2|L\}}{P\{F1|\overline{L}\}P\{F2|\overline{L}\}} \\ &= O\{L\} * LS_1 * LS_2 \end{aligned}$$

$$Logit\{L|F1 \cap F2\} = Logit\{L\} + W_1^+ + W_2^+ \tag{4.16}$$

Finally, a general expression for combining i = 1,2,3 ..., n maps containing data of factors is

$$Logit\{L|F_1 \cap F_2 \cap F_1 \cap F_2 \cap ...F_n\} = Logit\{L\} + \sum_{i=1}^{n} W^+ \tag{4.17}$$

Since all the inputs of the model are categorical data(multi-class maps), they contain several factors (classes) and the presence of one factor, implies the absence of the other factors of the same input data. Therefore a total weight must be calculated for each factor (equation 4.18 where $W_{MinTotal}$ is the total of all the negative weights in the input.

$$W_{map} = W^+ + W_{MinTotal} - W^- \tag{4.18}$$

4.4.2 Implementation

The implementation of WOE required discrete data as input. The workflow followed to the implementation is presented in the figure 4.4. First, the calculation of the weighted values is computed, then a test of statistical independence is performed and finally, different models are computed and evaluated.

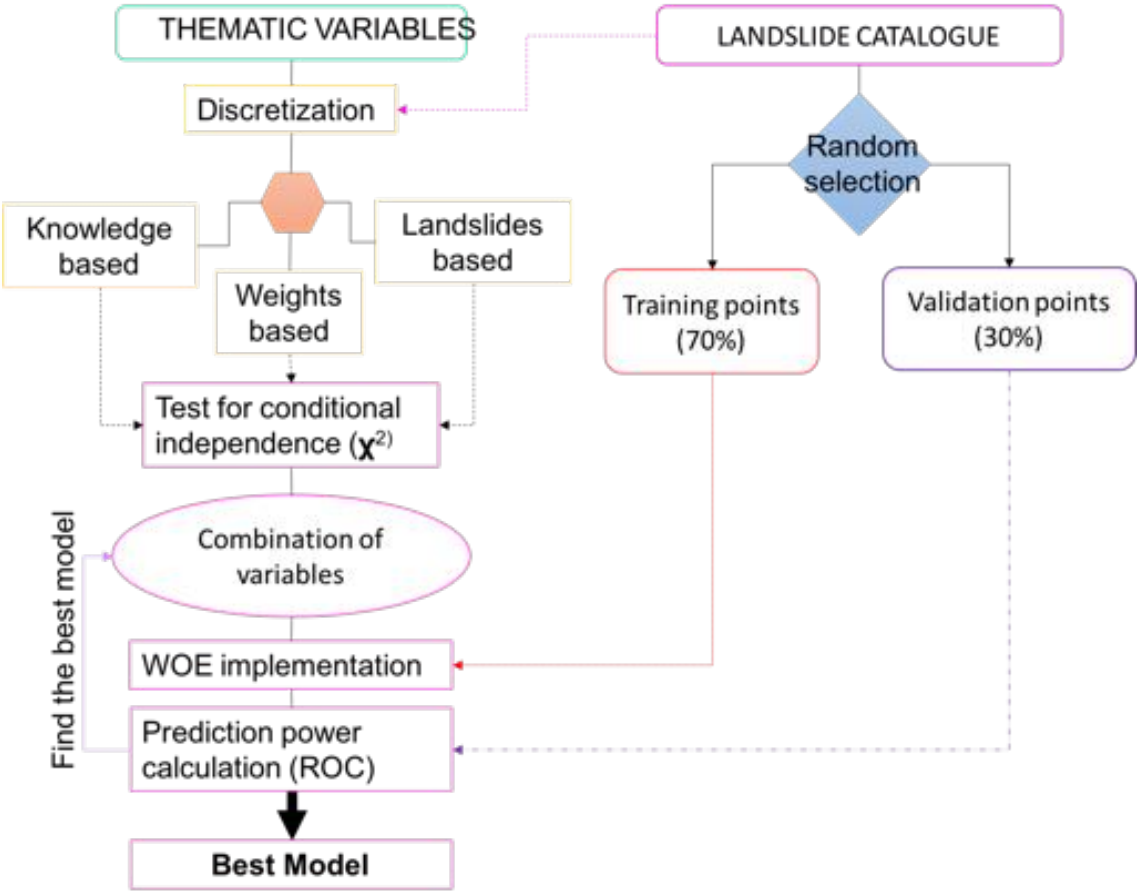


Figure 4.4: *Workflow followed to the calculation of the landslide susceptibility in the area based on the weight of evidence approach*

4.4.2.1 Discrete data

Each of the methodologies results in different break points per variable. The first method is based on the most common breaks points used in the literature, equal interval of values and experience gained by the analysis of the relation between each of the variables and the landslide catalogue. The breakpoints are summarized in the table 4.1.

Table 4.1: *Breakpoints for each of the variables based on the 3 different approaches*

Thematic group	Variable	Method1	Method2	Method3
Climatic and hydrology	Precipitation	0.009	0.009	0.009
		0.05	0.092	0.06
		0.1	0.116	0.08
		0.15	0.157	0.13
		0.219	0.219	0.219
	Distance to glaciar	0	0	0
		25	25	4.3
		50	50	6.4
		275.8	64	9.4
			95	14.6
			275.8	275.8
	Elevation above channel	0	0	0
		200	700	779
		400	1700	1003
		600	1900	1279
		800	2120	1521
		4072	4072	4072
	Distance to channel	0	0	0
		5	5	2.4
		10	15.5	4.26
		20	49.2	6.88
		30		9.7
		40		49.2
		49.2		
	TWI	0	0	0
		10	17	8
		20	44	11
		30	99	13
		40		16
		60		99
		99		
LandCover	NDVI	-25	-25	-25
		10	-5	13
		-5	0	18
		2.5	13	26
		20	37	40
		30	63	100
		100	100	
Geomorphology	Slope	0	0	0
		10	18	20
		20	58	27
		30	83	34
		40		43

Table 4.1: *Breakpoints for each of the variables based on the 3 different approaches*

Thematic group	Variable	Method1	Method2	Method3
		60		83
		83		
	SR	1	1	1
		1.5	1.1	1.15
		2	1.5	1.19
		2.5	1.9	1.24
			2.5	1.33
				2.5
	ERR	0.042	0.042	0.042
		0.53	0.38	0.4
		0.45	0.49	0.42
		0.6	0.55	0.44
		0.75	0.58	0.5
			0.61	0.75
			0.75	
	SI	-0.73	-0.73	-0.73
		-0.5	-0.33	-0.35
		-0.2	-0.13	-0.31
		0.1	-0.025	-0.25
		0.25	0.25	0.25
	Local Relief	16	16	16
		1000	550	2015
		1500	2450	2296
		2000	3100	2423
		2500	4100	2631
		3000		4100
		3500		
		4100		
	TPI	-1261	-1261	-1261
		-1000	-500	-216
		-750	-30	66
		0	120	264
		500	760	459
		1000	900	1637
		1637	1637	
	EigenValues	-0.00045	-0.00045	-0.00045
		-0.0002	-0.0002	-0.000137
		-0.00005	-0.00007	-0.000088
		0	0	-0.000046
		0.00005	0.00014	0.0000427
		0.0002	0.00019	0.00039
		0.00039	0.00039	
Tectonic	Distance to fault	0	0	0
		5	0.5	0.5
		10	5	1.4
		20	10	3.2
		71.64	71.64	5.7
				71.64

The second approach to categorized the variables is based on the methodology followed by (Regmi *et al.* , 2010), whose defined the breakpoints based on the change in the curves obtained from de variation on weight contrast in the values of the class. This approach is based on a graphical interpretation of the implementation of the WOE theory using the whole landslide catalogue. The results are summarize in the figure 4.5.

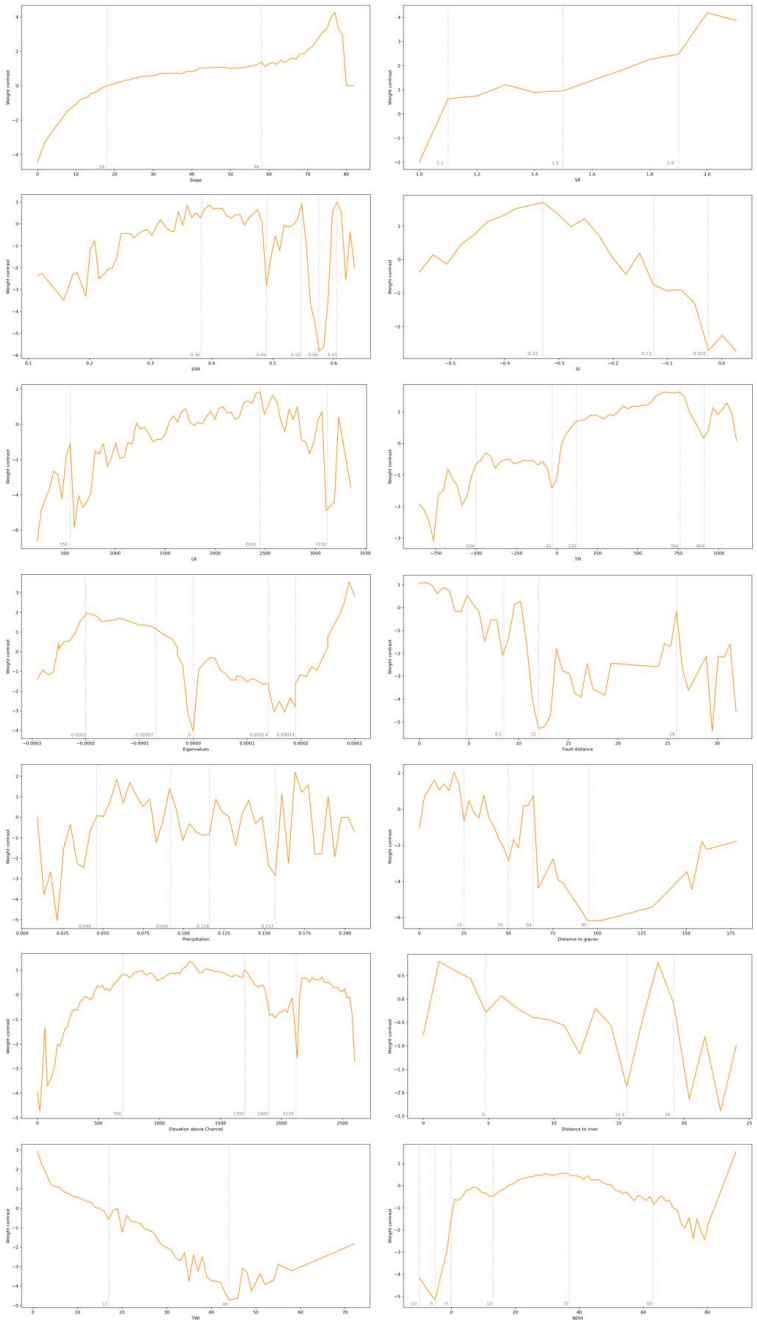


Figure 4.5: *Weight of contrast curve created by the computation of the WOE using the whole catalogue. Changes in the curvature are selected manually and are used as breakpoints for the discretization process.*

The last approach implemented is the selection of the breakspoints based on the number of landslides. It aims to divide the variable in n number of classes with a similar number of landslides. This approach seeks to avoid the methodological problem of having not enough landslides per class. A graphical representation of the break points and the values per class against the weight contrast is used. Each of the variables is divided into 5 classes with an approximately same number of landslides (figure 4.6).

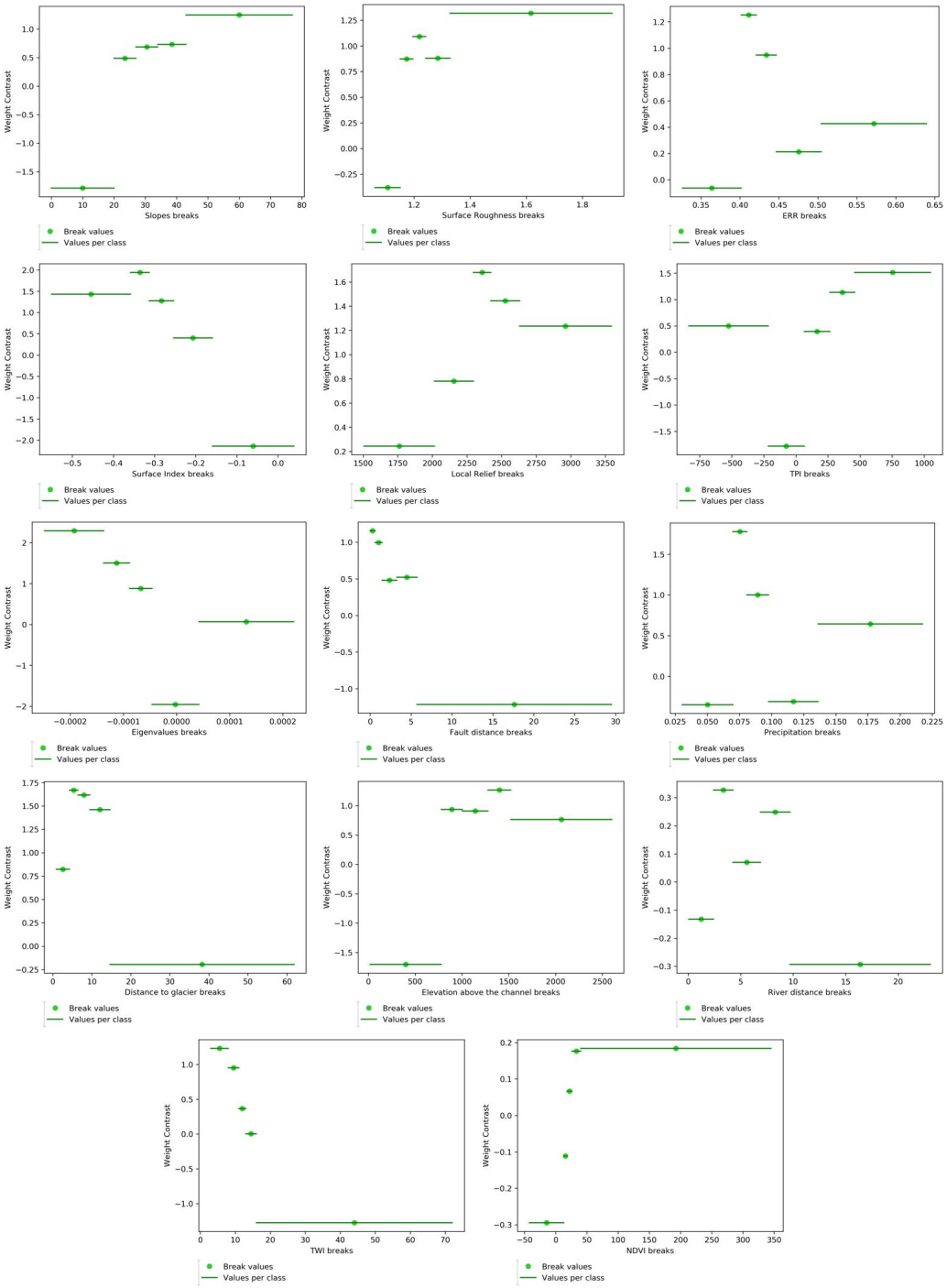


Figure 4.6: *Diagram that represent the break points selected by the method of discretization 3. Green lines represent the range of values per class and its relation to the weight of contrast.*

4.4.2.2 Calculation of weighted values

The calculation of the weighted values for each of the thematic variables is made based on the equations 4.10 and 4.11. They can be translated in term of number of pixels and implemented in Python 3 using the equations 4.19 and 4.20; where $N1$ = Number of landslide pixels present in the variable class ; $N2$= Number landslide pixels not present in the variable class , $N3$ = Number of landslide pixels with no presence in the variable class, and, $N4$ = Number of pixels in which neither landslide nor the variable class is present.

$$W+ = \log_e \frac{\frac{N1}{N1+N2}}{\frac{N3}{N3+N4}} \tag{4.19}$$

$$W- = \log_e \frac{\frac{N2}{N1+N2}}{\frac{N4}{N3+N4}} \tag{4.20}$$

Table 4.2: *Weight calculation for each of the thematic variables using the first method of discretization.*

Dataset	**Class min**	**Class max**	**nPixels**	**nlandslides**	**W+**	**W-**	W_c
Slopes	0	18	118680689	6235	-1.44	0.58	-2.02
	18	58	112535060	43559	0.56	-1.19	1.75
	58	83	1882006	1872	1.50	-0.03	1.53
Aspect	0	45	36542288	7594	-0.01	0.00	-0.01
	45	90	23863553	3163	-0.46	0.04	-0.49
	90	135	26727298	4602	-0.19	0.02	-0.22
	135	180	32813263	9953	0.37	-0.07	0.44
	180	225	36392829	5006	-0.42	0.06	-0.48
	225	270	27082528	3315	-0.54	0.05	-0.59
	270	315	29724119	5662	-0.09	0.01	-0.11
	315	359	33680370	12290	0.56	-0.13	0.68
TPI	-1261	-500	9790698	434	-1.55	0.03	-1.59
	-500	-30	83780890	9590	-0.60	0.21	-0.81
	-30	120	98565770	11768	-0.56	0.25	-0.81
	120	760	52514311	28615	0.96	-0.57	1.52
	760	900	1389201	1030	1.26	-0.01	1.28
	900	1637	719582	273	0.59	-0.00	0.60
Surface Index	-0.73	-0.33	20675907	13002	1.11	-0.20	1.31
	-0.33	-0.13	68374881	30419	0.76	-0.56	1.33
	-0.13	-0.025	81833205	6957	-0.89	0.25	-1.15
	-0.025	0.25	78433853	1425	-2.44	0.35	-2.79
ERR	0.042	0.38	115777021	16436	-0.38	0.24	-0.62
	0.38	0.49	88777268	29464	0.47	-0.40	0.87
	0.49	0.55	28854947	3922	-0.42	0.04	-0.47
	0.55	0.58	8527189	933	-0.64	0.02	-0.66
	0.58	0.61	4611962	644	-0.40	0.01	-0.40
	0.61	0.75	2769741	404	-0.35	0.00	-0.36

Table 4.2: *Weight calculation for each of the thematic variables using the first method of discretization.*

Dataset	**Class min**	**Class max**	**nPixels**	**nlandslides**	**W+**	**W-**	W_c
Surface Roughness	1	1.1	138403914	7529	-1.34	0.65	-1.99
	1.1	1.5	108038913	41888	0.62	-1.09	1.71
	1.5	1.9	2815783	2050	1.25	-0.03	1.28
	1.9	2.5	60430	336	3.29	-0.01	3.30
Local Relief	16	550	68947449	212	-4.21	0.32	-4.53
	550	2450	149801773	37361	0.18	-0.36	0.54
	2450	3100	26752602	14086	0.93	-0.20	1.13
	3100	4100	3567363	144	-1.64	0.01	-1.65
Eigenvalues	-0.00045	-0.00020	973016	525	0.95	-0.01	0.96
	-0.00020	-0.00007	26801251	24120	1.47	-0.51	1.98
	-0.00007	0.00000	87102560	17842	-0.01	0.01	-0.02
	0.00000	0.00014	129287512	9036	-1.09	0.54	-1.63
	0.00014	0.00019	4068238	92	-2.22	0.01	-2.23
	0.00019	0.00039	1086483	188	-0.18	0.00	-0.18
Elevation above channel	0	700	156996968	13282	-0.93	0.76	-1.69
	700	1700	67517344	34226	0.86	-0.76	1.62
	1700	1900	6420286	2333	0.53	-0.02	0.55
	1900	2120	4664213	501	-0.69	0.01	-0.70
	2120	4072	5042217	1280	0.17	-0.00	0.17
Lithology	Quaternary		9228554	6085	0.73	-0.07	0.80
	Neogene		0	0	0.00	0.00	0.00
	Paleogene		26862125	1411	-1.80	0.16	-1.96
	Paleogene Intrusive		13786220	1540	-1.04	0.06	-1.10
	Creataceous		1642547	4	-4.87	0.01	-4.88
	Createceous Jurassic		6330003	2368	0.17	-0.01	0.17
	Jurassic		16998217	4150	-0.26	0.03	-0.29
	Jurassic Triassic		860398	576	0.75	-0.01	0.76
	Permian Igneous		4041794	45	-3.35	0.03	-3.37
	Permian		2239737	288	-0.90	0.01	-0.91
	Carboniferous		4806542	6	-5.54	0.03	-5.57
	Carboniferous igneous		9231340	6179	0.75	-0.07	0.82
	Devonia		17292244	3586	-0.42	0.04	-0.47
	Silurian		8897271	12292	1.47	-0.22	1.70
	Cambrian/ Precambrian		15869855	11347	0.81	-0.15	0.97
	Glacier areas		19484756	26	-5.47	0.13	-5.60
Seimo zones	1		9343	11	1.70	-0.00	1.70
	2		1391793	0	0.00	0.00	0.00
	3		39362394	106	-4.38	0.18	-4.56
	4		6060020	3992	1.12	-0.05	1.17
	5		56884955	1642	-2.01	0.24	-2.25
	6		27115562	322	-2.90	0.11	-3.01
	7		15855625	0	0.00	0.00	0.00

Table 4.2: *Weight calculation for each of the thematic variables using the first method of discretization.*

Dataset	**Class min**	**Class max**	**nPixels**	**nlandslides**	**W+**	**W-**	W_c
	8		0	0	0.00	0.00	0.00
	9		153909	0	0.00	0.00	0.00
	10		3569232	0	0.00	0.00	0.00
	11		8729753	3594	0.65	-0.03	0.68
	12		9528464	1300	-0.46	0.02	-0.47
	13		69693495	40527	0.99	-1.18	2.18
	0	5	96480313	42077	0.74	-1.18	1.93
	5	8.5	35446767	5951	-0.21	0.03	-0.24
Fault distance	8.5	12	26198513	2894	-0.63	0.05	-0.69
	12.	26	53347847	542	-3.02	0.23	-3.25
	26	71.64	37845620	339	-3.14	0.16	-3.30
	0	25	93479531	43945	0.82	-1.42	2.23
	25	50	28944368	4857	-0.21	0.02	-0.24
Distance to glacier	50	64	13763013	1507	-0.64	0.03	-0.67
	64	95	28291777	1341	-1.48	0.09	-1.57
	95	275.8	84840373	153	-4.75	0.41	-5.16
	0.009	0.049	100537168	4769	-1.48	0.42	-1.90
	0.049	0.092	48539161	22563	0.81	-0.36	1.16
Precipitation	0.092	0.116	40163778	8516	0.02	-0.00	0.02
	0.116	0.157	47896341	10750	0.08	-0.02	0.10
	0.157	0.219	12167311	5205	0.72	-0.06	0.78
	0	5.	108483821	31051	0.32	-0.34	0.66
Distance to channel	5	15.5	119930939	18215	-0.31	0.22	-0.54
	15.5	19	10853000	1847	-0.20	0.01	-0.21
	19	49.2	9932299	683	-1.11	0.03	-1.13
	0	17	142138757	42172	0.34	-0.86	1.21
TWI	17	44	95632003	8950	-0.81	0.31	-1.12
	44	99	5475614	18	-4.16	0.02	-4.18
	-25	-5	2002272	52	-2.06	0.01	-2.07
	-5	0	5241782	22	-3.89	0.02	-3.91
NDVI	0	13	55257627	8316	-0.31	0.07	-0.38
	13	37	130692339	30848	0.14	-0.19	0.34
	37	63	41917581	10017	0.16	-0.04	0.19
	63	100	9491830	755	-0.94	0.02	-0.97

Table 4.3: *Weight calculation for each of the thematic variables using the second method of discretization.*

Dataset	**Class min**	**Class max**	**Pixels**	**nlandslides**	**W+**	**W-**	W_c
	0	18	118680689	6235	-1.44	0.58	-2.02
Slopes	18	58	112535060	43559	0.56	-1.19	1.75
	58	83	1882006	1872	1.5	-0.03	1.53
	0	45	36542288	7594	-0.01	0	-0.01
	45	90	23863553	3163	-0.46	0.04	-0.49
	90	135	26727298	4602	-0.19	0.02	-0.22
Aspect	135	180	32813263	9953	0.37	-0.07	0.44

Table 4.3: *Weight calculation for each of the thematic variables using the second method of discretization.*

Dataset	**Class min**	**Class max**	**Pixels**	**nlandslides**	**W+**	**W-**	W_c
	180	225	36392829	5006	-0.42	0.06	-0.48
	225	270	27082528	3315	-0.54	0.05	-0.59
	270	315	29724119	5662	-0.09	0.01	-0.11
	315	359	33680370	12290	0.56	-0.13	0.68
TPI	-1261	-500	9790698	434	-1.55	0.03	-1.59
	-500	-30	83780890	9590	-0.6	0.21	-0.81
	-30	120	98565770	11768	-0.56	0.25	-0.81
	120	760	52514311	28615	0.96	-0.57	1.52
	760	900	1389201	1030	1.26	-0.01	1.28
	900	1637	719582	273	0.59	0	0.6
Surface Index	-0.73	-0.33	20675907	13002	1.11	-0.2	1.31
	-0.33	-0.13	68374881	30419	0.76	-0.56	1.33
	-0.13	-0.025	81833205	6957	-0.89	0.25	-1.15
	-0.025	0.25	78433853	1425	-2.44	0.35	-2.79
ERR	0.042	0.38	115777021	16436	-0.38	0.24	-0.62
	0.38	0.49	88777268	29464	0.47	-0.4	0.87
	0.49	0.55	28854947	3922	-0.42	0.04	-0.47
	0.55	0.58	8527189	933	-0.64	0.02	-0.66
	0.58	0.61	4611962	644	-0.4	0.01	-0.4
	0.61	0.75	2769741	404	-0.35	0	-0.36
Surface Roughness	1	1.1	138403914	7529	-1.34	0.65	-1.99
	1.1	1.5	108038913	41888	0.62	-1.09	1.71
	1.5	1.9	2815783	2050	1.25	-0.03	1.28
	1.9	2.5	60430	336	3.29	-0.01	3.3
Local Relief	16	550	68947449	212	-4.21	0.32	-4.53
	550	2450	149801773	37361	0.18	-0.36	0.54
	2450	3100	26752602	14086	0.93	-0.2	1.13
	3100	4100	3567363	144	-1.64	0.01	-1.65
Eigenvalues	-0.00045	-0.0002	973016	525	0.95	-0.01	0.96
	-0.0002	-0.00007	26801251	24120	1.47	-0.51	1.98
	-0.00007	0	87102560	17842	-0.01	0.01	-0.02
	0	0.00014	129287512	9036	-1.09	0.54	-1.63
	0.00014	0.00019	4068238	92	-2.22	0.01	-2.23
	0.00019	0.00039	1086483	188	-0.18	0	-0.18
Elevation above channel	0	700	156996968	13282	-0.93	0.76	-1.69
	700	1700	67517344	34226	0.86	-0.76	1.62
	1700	1900	6420286	2333	0.53	-0.02	0.55
	1900	2120	4664213	501	-0.69	0.01	-0.7
	2120	4072	5042217	1280	0.17	0	0.17
Lithology	Quaternary		9228554	6085	0.73	-0.07	0.8
	Neogene		26862125	1411	-1.8	0.16	-1.96
	Paleogene		13786220	1540	-1.04	0.06	-1.1
	Paleogene intrusive		1642547	4	-4.87	0.01	-4.88
	Creataceous		6330003	2368	0.17	-0.01	0.17
	Creataceous Jurassic		16998217	4150	-0.26	0.03	-0.29
	Jurassic		860398	576	0.75	-0.01	0.76

Table 4.3: *Weight calculation for each of the thematic variables using the second method of discretization.*

Dataset	**Class min**	**Class max**	**Pixels**	**nlandslides**	**W+**	**W-**	W_c
	Jurassic Triassic		4041794	45	-3.35	0.03	-3.37
	Permian ingenous		2239737	288	-0.9	0.01	-0.91
	Permian		4806542	6	-5.54	0.03	-5.57
	Carboniferous		9231340	6179	0.75	-0.07	0.82
	Carboniferous igneous		17292244	3586	-0.42	0.04	-0.47
	Devonian		8897271	12292	1.47	-0.22	1.7
	Silurian		15869855	11347	0.81	-0.15	0.97
	Cambrian/ Precambrian		16696411	26	-5.32	0.11	-5.43
	Glacier areas		2788345	0	0	0	0
	1		9343	11	1.7	0	1.7
	2		1391793	0	0	0	0
	3		39362394	106	-4.38	0.18	-4.56
	4		6060020	3992	1.12	-0.05	1.17
	5		56884955	1642	-2.01	0.24	-2.25
	6		27115562	322	-2.9	0.11	-3.01
Seimo zones	7		15855625	0	0	0	0
	8		0	0	0	0	0
	9		153909	0	0	0	0
	10		3569232	0	0	0	0
	11		8729753	3594	0.65	-0.03	0.68
	12		9528464	1300	-0.46	0.02	-0.47
	13		69693495	40527	0.99	-1.18	2.18
	0	5	96480313	42077	0.74	-1.18	1.93
	5	8.5	35446767	5951	-0.21	0.03	-0.24
Fault distance	8.5	12	26198513	2894	-0.63	0.05	-0.69
	12	26	53347847	542	-3.02	0.23	-3.25
	26	71.64	37845620	339	-3.14	0.16	-3.3
	0	25	93479531	43945	0.82	-1.42	2.23
	25	50	28944368	4857	-0.21	0.02	-0.24
Distance to glacier	50	64	13763013	1507	-0.64	0.03	-0.67
	64	95	28291777	1341	-1.48	0.09	-1.57
	95	275.8	84840373	153	-4.75	0.41	-5.16
	0.009	0.049	100537168	4769	-1.48	0.42	-1.9
	0.049	0.092	48539161	22563	0.81	-0.36	1.16
Precipitation	0.092	0.116	40163778	8516	0.02	0	0.02
	0.116	0.157	47896341	10750	0.08	-0.02	0.1
	0.157	0.219	12167311	5205	0.72	-0.06	0.78
	0	5	108483821	31051	0.32	-0.34	0.66
River distance	5	15.5	119930939	18215	-0.31	0.22	-0.54
	15.5	19	10853000	1847	-0.2	0.01	-0.21
	19	49.2	9932299	683	-1.11	0.03	-1.13
	0	17	142138757	42172	0.34	-0.86	1.21
TWI	17	44	95632003	8950	-0.81	0.31	-1.12
	44	99	5475614	18	-4.16	0.02	-4.18

Table 4.3: *Weight calculation for each of the thematic variables using the second method of discretization.*

Dataset	**Class min**	**Class max**	**Pixels**	**nlandslides**	**W+**	**W-**	W_c
NDVI	-25	-5	2002272	52	-2.06	0.01	-2.07
	-5	0	5241782	22	-3.89	0.02	-3.91
	0	13	55257627	8316	-0.31	0.07	-0.38
	13	37	130692339	30848	0.14	-0.19	0.34
	37	63	41917581	10017	0.16	-0.04	0.19
	63	100	9491830	755	-0.94	0.02	-0.97

Table 4.4: *Weight calculation for each of the thematic variables using the third method of discretization.*

Dataset	**Class min**	**Class max**	**nPixels**	**nlandslides**	**W+**	**W-**	W_c
Slopes	0	20	127657854	8138	-1.25	0.62	-1.87
	20	27	30814351	8641	0.24	-0.04	0.28
	27	34	28859706	10946	0.54	-0.11	0.64
	34	43	27946252	12642	0.71	-0.15	0.87
	43	83	17819592	11299	1.05	-0.17	1.22
Aspect	0	45	36542288	7594	-0.01	0	-0.01
	45	90	23863553	3163	-0.46	0.04	-0.49
	90	135	26727298	4602	-0.19	0.02	-0.22
	135	180	32813263	9953	0.37	-0.07	0.44
	180	225	36392829	5006	-0.42	0.06	-0.48
	225	270	27082528	3315	-0.54	0.05	-0.59
	270	315	29724119	5662	-0.09	0.01	-0.11
	315	359	33680370	12290	0.56	-0.13	0.68
TPI	-1261	-216	33691792	3310	-0.76	0.08	-0.84
	-216	66	145607451	14354	-0.75	0.57	-1.32
	66	264	35891344	14524	0.66	-0.17	0.83
	264	459	18435118	9613	0.91	-0.13	1.04
	459	1637	13134747	9909	1.28	-0.16	1.44
Surface Index	-0.73	-0.35	17025447	9770	1.02	-0.14	1.15
	-0.35	-0.31	7807615	7438	1.52	-0.12	1.65
	-0.31	-0.25	16565471	11122	1.17	-0.17	1.35
	-0.25	-0.15	38433204	12805	0.47	-0.12	0.59
	-0.15	0.25	169486109	10668	-1.19	0.91	-2.1
ERR	0.042	0.4	131420557	22217	-0.21	0.19	-0.4
	0.4	0.42	17057733	7245	0.72	-0.08	0.8
	0.42	0.44	17353992	5200	0.37	-0.03	0.4
	0.44	0.5	45017550	11345	0.19	-0.05	0.24
	0.5	0.75	38468296	5796	-0.32	0.05	-0.37
Surface Roughness	1	1.15	166757285	15903	-0.78	0.74	-1.52
	1.15	1.19	19072572	7111	0.58	-0.07	0.65
	1.19	1.24	23948944	7904	0.46	-0.06	0.53
	1.24	1.33	23998232	11540	0.84	-0.15	0.99
	1.33	2.5	15542007	9345	1.06	-0.13	1.2
Local Relief	16	2015	179755599	16513	-0.82	0.9	-1.71
	2015	2296	26056114	10337	0.65	-0.11	0.76

Table 4.4: *Weight calculation for each of the thematic variables using the third method of discretization.*

Dataset	**Class min**	**Class max**	**nPixels**	**nlandslides**	**W+**	**W-**	W_c
	2296	2423	10793838	8889	1.38	-0.14	1.52
	2423	2631	13673301	10956	1.35	-0.18	1.53
	2631	4100	18790335	5108	0.27	-0.03	0.29
	-0.00045	-0.00014	6376892	6482	1.59	-0.11	1.7
	-0.00014	-0.00009	13265706	11991	1.47	-0.21	1.68
Eigenvalues	-0.00009	-0.00005	23327530	14635	1.11	-0.23	1.34
	-0.00005	0.00004	159383206	15291	-0.77	0.67	-1.44
	0.00004	0.00039	46965726	3404	-1.05	0.14	-1.19
	0	779	164521217	16667	-0.75	0.76	-1.51
	779	1003	18945428	9161	0.81	-0.11	0.93
Elevation above channel	1003	1279	19526420	10311	0.9	-0.14	1.04
	1279	1521	13707542	7785	0.97	-0.1	1.08
	1521	4072	23940421	7698	0.4	-0.06	0.46
	Quaternary	1	9228554	6085	0.73	-0.07	0.8
	Neogene	3	26862125	1411	-1.8	0.16	-1.96
	Paleogene	4	13786220	1540	-1.04	0.06	-1.1
	Paleogene intrusive	5	1642547	4	-4.87	0.01	-4.88
	Cretaceous	6	6330003	2368	0.17	-0.01	0.17
	Cretaceous Jurassic	7	16998217	4150	-0.26	0.03	-0.29
Lithology	Jurassic	8	860398	576	0.75	-0.01	0.76
	Jurassic Triassic	9	4041794	45	-3.35	0.03	-3.37
	Permian Igneous	10	2239737	288	-0.9	0.01	-0.91
	Permian	11	4806542	6	-5.54	0.03	-5.57
	Carboniferous	12	9231340	6179	0.75	-0.07	0.82
	Carboniferous igneous	13	17292244	3586	-0.42	0.04	-0.47
	Devonian	14	8897271	12292	1.47	-0.22	1.7
	Silurian	15	15869855	11347	0.81	-0.15	0.97
	Cambrian/ Precambrian	16	19484756	26	-5.47	0.13	-5.6
	Glacier areas		2788345	0	0	0	0
	1		9343	11	1.7	0	1.7
	2		1391793	0	0	0	0
	3		39362394	106	-4.38	0.18	-4.56
	4		6060020	3992	1.12	-0.05	1.17
	5		56884955	1642	-2.01	0.24	-2.25
	6		27115562	322	-2.9	0.11	-3.01
Seimo zones	7		15855625	0	0	0	0
	8		0	0	0	0	0
	9		153909	0	0	0	0
	10		3569232	0	0	0	0
	11		8729753	3594	0.65	-0.03	0.68
	12		9528464	1300	-0.46	0.02	-0.47
	13		69693495	40527	0.99	-1.18	2.18

Table 4.4: *Weight calculation for each of the thematic variables using the third method of discretization.*

Dataset	**Class min**	**Class max**	**nPixels**	**nlandslides**	**W+**	**W-**	W_c
Fault distance	0	0.5	15539586	8509	0.97	-0.12	1.08
	0.5	1.4	22609863	12572	0.98	-0.18	1.17
	1.4	3.2	33375410	14919	0.77	-0.2	0.96
	3.2	5.7	33039081	8437	0.21	-0.04	0.24
	5.7	71.64	144755120	7366	-1.41	0.72	-2.12
Distance to glacier	0	4.3	41045944	6101	-0.34	0.05	-0.39
	4.3	6.4	10614858	4916	0.8	-0.06	0.86
	6.4	9.4	11677938	6803	1.03	-0.09	1.12
	9.4	14.6	13602623	10804	1.34	-0.18	1.52
	14.6	257.8	171970469	23179	-0.43	0.58	-1.01
Precipitation	0.009	0.6	134573638	23860	-0.16	0.16	-0.32
	0.6	0.08	0	0	0	0	0
	0.08	0.13	76891929	17887	0.11	-0.05	0.17
	0.13	0.219	37853063	10056	0.25	-0.05	0.3
Distance to channel	0	2.4	56414033	14074	0.18	-0.06	0.24
	2.4	4.26	38098840	12813	0.48	-0.12	0.6
	4.26	6.88	46993939	9790	0	0	0
	6.88	9.7	40693341	6938	-0.2	0.03	-0.23
	9.7	49.2	66999906	8181	-0.53	0.14	-0.67
TWI	0	8	9437735	5007	0.93	-0.06	0.99
	8	11	32704884	12395	0.59	-0.13	0.72
	11	13	33115225	10048	0.37	-0.07	0.44
	13	16	50975824	11847	0.1	-0.03	0.13
	16	99	117012706	11843	-0.73	0.39	-1.12
NDVI	-25	13	62501681	8390	-0.42	0.11	-0.53
	13	18	47114878	7054	-0.31	0.06	-0.37
	18	26	49945048	12463	0.2	-0.06	0.26
	26	40	40898395	13795	0.5	-0.14	0.64
	40	100	44143429	8308	-0.08	0.02	-0.1

4.4.2.3 Test for conditional independence

The theory of WOE assumes that the variables used are conditional independent. Conditional independence means that the probability that one occur does not affect the probability of the other event to occur. To warranty this assumption as true, a conditional independence test is performed based on the integrated W_c results by pairwise comparison using chi-square statistics.

First, the variables are converted into a binary pattern based on the presence or absence of landslides. 2x2 contingency tables for all possible pairs of variables are prepared based on the table 4.5 (Regmi *et al.* , 2010).

Finally, the Chi-square test is performed with 1 degree of freedom following the equation 4.21. The Chi-square values are compared with the value for 1 degree of freedom at the 99% of confidence level (χ^2 = 6.64). A Chi-square value greater than 6.64, suggesting that the pairs are not significantly different, given the occurrence of landslides, so, they

Table 4.5: *2x2 contingency table construction. Observed frequencies (O_i) and expected frequencies (E_i)*

		Binary variable 1 V_1		
	Lanslide	Presence	Absence	Total
Binary variable V_2	Presence	$O_1 = \{V_1 \cap V_2 \cap L\}$ $E_1 = \frac{\{V_2 \cap L\} * \{V_1 \cap L\}}{L}$	$O_3 = \{\overline{V_1} \cap V_2 \cap L\}$ $E_3 = \frac{\{V_2 \cap L\} * \{\overline{V_1} \cap L\}}{L}$	T
	Absence	$O_2 = \{V_1 \cap \overline{V_2} \cap L\}$ $E_2 = \frac{\{\overline{V_2} \cap L\} * \{V_1 \cap L\}}{L}$	$O_4 = \{\overline{V_1} \cap \overline{V_2} \cap L\}$ $E_4 = \frac{\{\overline{V_2} \cap L\} * \{\overline{V_1} \cap L\}}{L}$	T
	Total	$\{V_1 \cap L\}$	$\{\overline{V_1} \cap L\}$	$\{L\}$

are considered statistically dependent and cannot be used together to the implementation of the WOE model.

$$\chi^2 = \sum_{k=1}^{n} \frac{(O_i - E_i)^2}{E_i} \tag{4.21}$$

The results of the Chi-square test for each of the discretization approaches are presented graphically using an association plot, where the dependent variables are highlighted in yellow (figure 4.7)

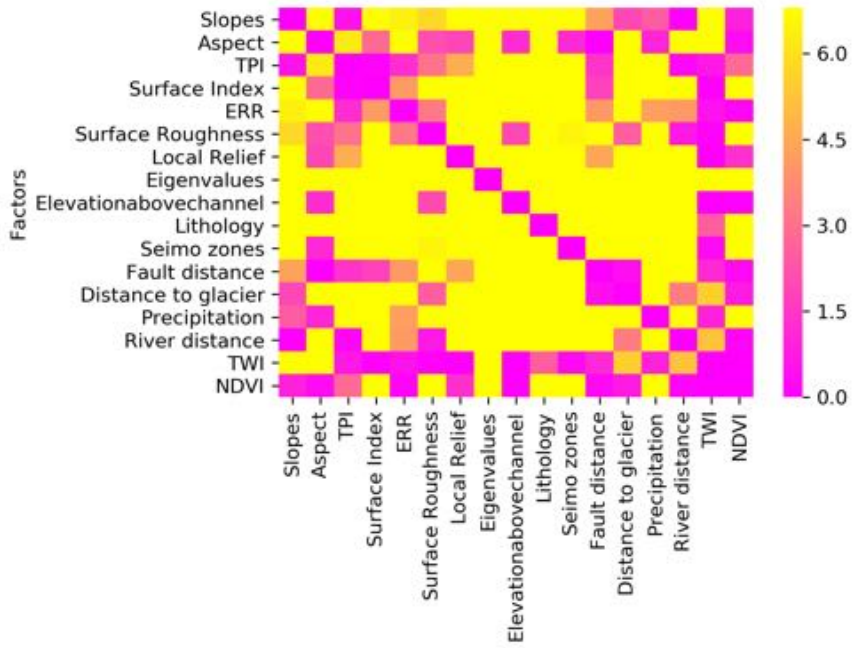


Figure 4.7: *Chi-square test result for 1 degree of freedom and 99% confidence level ($\chi^2 = 6.64$).Method 1*

The conditional dependence of the variables change according to the method implemented to create the different classes; however, for all the methods EigenValues and Lithology are conditional dependent on most of the other variables. For the first method

(figure 4.7) it is possible to use different geomorphological parameters; however, slope and TPI are conditional independents for all the methods. Contrary, the number of dependent variables increase in the method 2 (figure 4.8) and method 3 (figure 4.9).

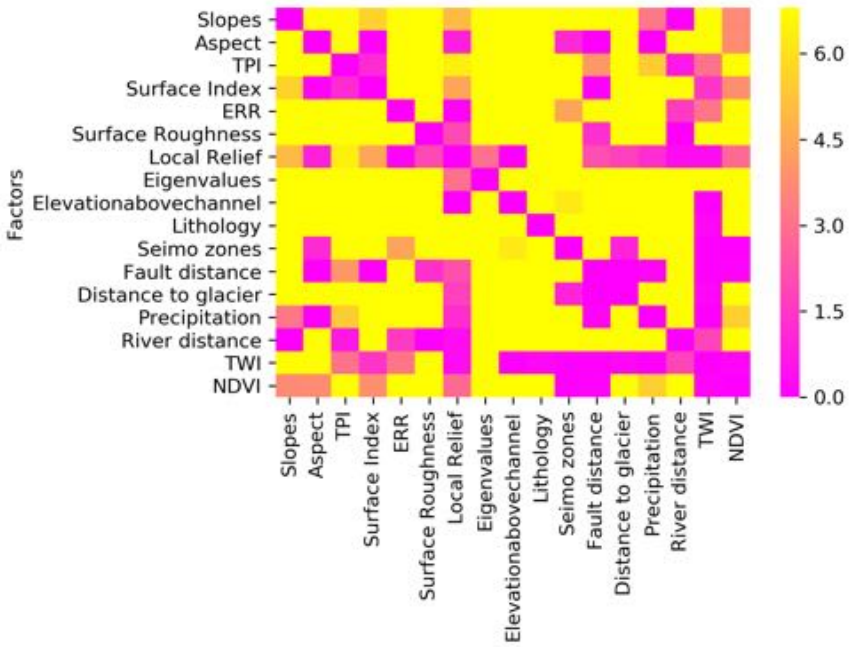


Figure 4.8: *Chi-square test result for 1 degree of freedom and 99% confidence level ($\chi^2 = 6.64$). Method 2*

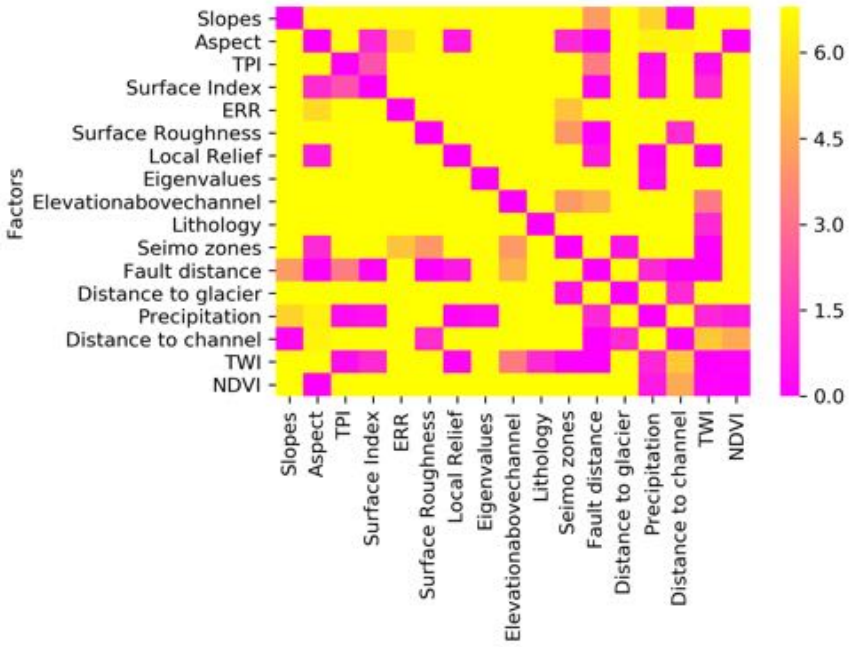


Figure 4.9: *Chi-square test result for 1 degree of freedom and 99% confidence level ($\chi^2 = 6.64$). Method 3*

4.4.2.4 Combination of variables

For each of the categorization approaches, 1 or 2 models are created taking into account the spatial association of the variable to the landslide occurrence as well as the results of the Chi-square test. The models are presented in table 4.6. They are starting model; that means, they are implemented and improved using reduction of variables until finding the variables that produce the best model.

Table 4.6: *Possible combinations of the variables based on the spatial association and the Chi-square statistics to performed the WOE approach.*

		Method1		*Method2*		*Method3*
Thematic group	**Variable**	**Model1-1**	**Model1-2**	**Model2-1**	**Model2-2**	**Model3-1**
Lithology	Lithology	x	x	x	x	x
Climatic and hydrological	Distance to glacier	x			x	
	Precipitation	x	x	x	x	x
	Distance to channel		x	x	x	
	TWI	x	x		x	x
LandCover	NDVI	x	x	x	x	x
Geomorphology	Slope	x		x		
	SR	x	x			
	ERR	x	x			
	SI		x	x	x	x
	Local Relief	x	x	x		
	TPI	x	x	x	x	x
	Elevation above channel				x	
Tectonic	Distance to fault	x	x	x	x	x

4.4.2.5 Model creation and improvement

Each of the five models presented in table 4.6 is implemented. The results are analysed based on the prediction power (ROC) and the quality of the resulting susceptibility map. Improvements are made to each model in order to get the best result.

The Model 1-1 is calculated using 10 predictive variables. Prediction power of 0.81 is obtained; however, the resulting map presents some pixels areas that can be determined as incoherences because of high W_c values are located near to very low values (figure 4.10). This zonation is reduced by the exclusion of the variable distance to glacier (Model 11-2); however, the resulting model doesn't improve the prediction capability (figure 4.11); instead new incoherent areas appear, where higher weight of contrast values are surrounded by lower values (figure 4.10).

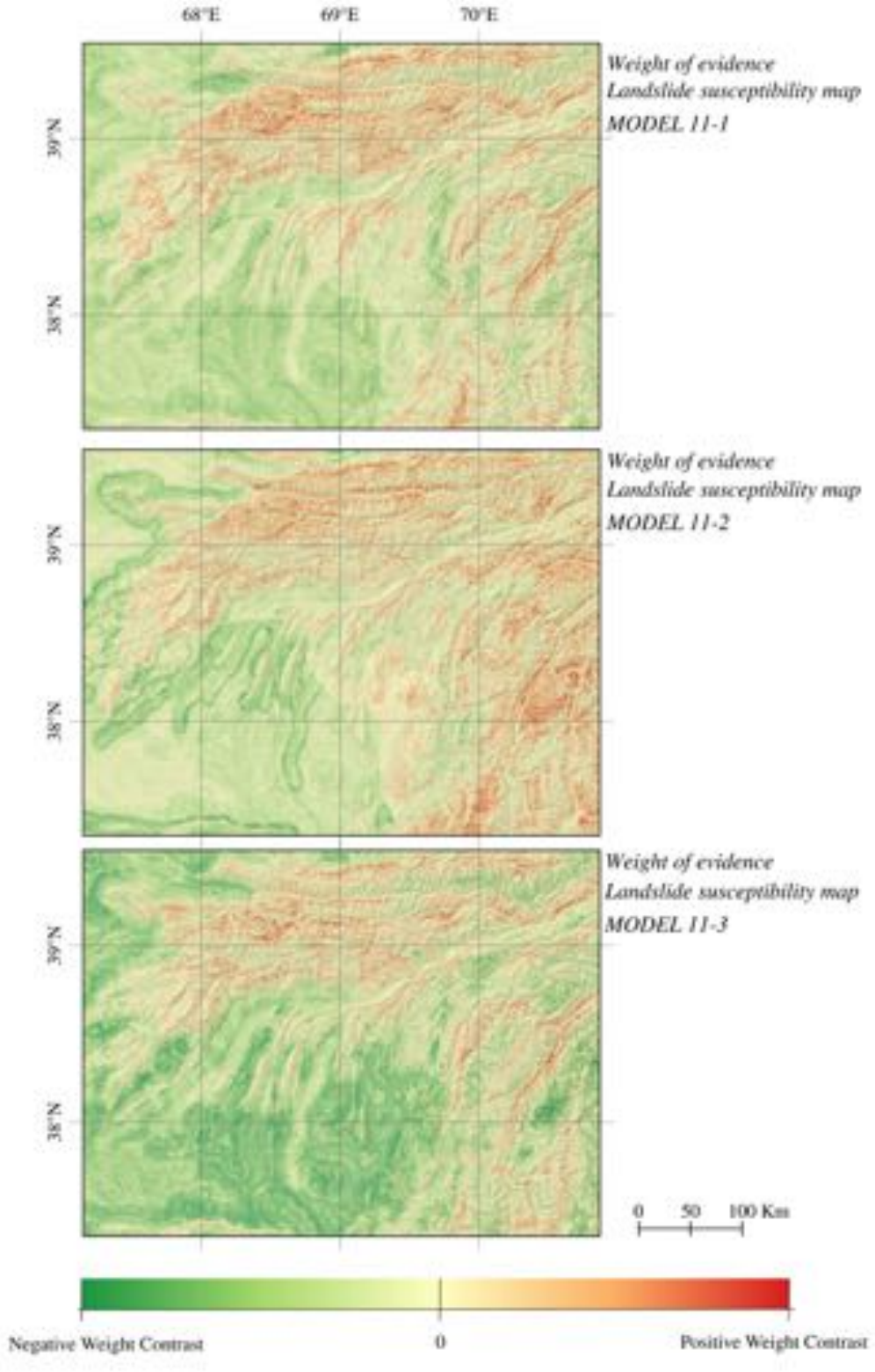


Figure 4.10: *Normalized total weight map resulting for the different variables combination based on the model 1-1 using the WOE approach.*

The model 1-1 is finally improved by the exclusion of the variables distance to glacier and precipitation (model 11-3). Without those variables the model prediction increase to 0.84 and there is a reduction of the incoherent areas; however, they are still present (figure 4.10).

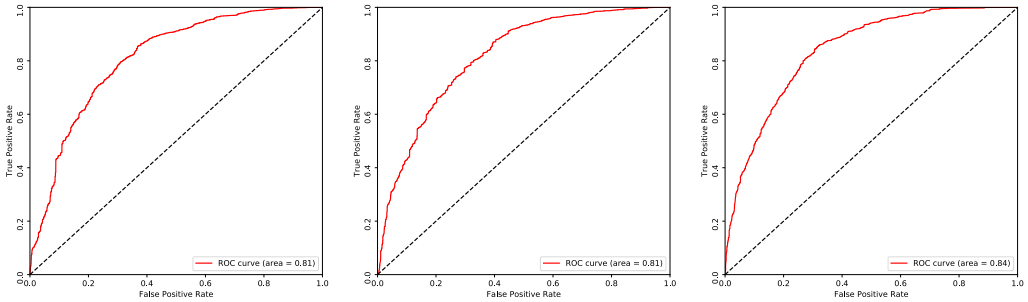


Figure 4.11: *Prediction capability for the model 1-1 and its variations. Left: Model 11-1 implemented with 10 thematic variables. Center: Model 11-2 implemented with 9 thematic variables. Distance to glacier excluded. Right: Model 11-3 implement with 8 thematic variables. Distance to glacier and precipitation excluded.*

A similar process is applied to the model 1-2. The starting model contains 11 thematic variables and has a slightly better predictive capability (AUC = 0.84). However, the weight contrast map presents evident incoherent areas. Those anomalies can be related to the influence of the distance to channel variable or the presence of an active fault. Also, sharp limits are observed in the north of Tien Shan where very high values are in direct contact with very low values. Those sharp limits are probably related to a zonation influenced by the precipitation information (figure 4.13).

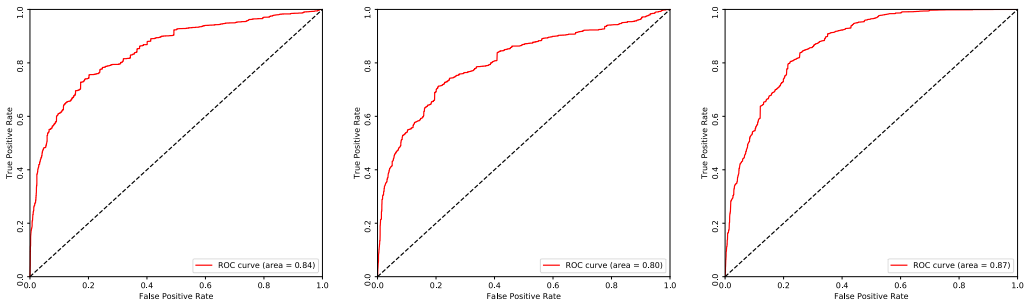


Figure 4.12: *Prediction capability for the model 1-2 and its variations. Left: Model 12-1 implemented with 11 thematic variables. Center: Model 12-2 implemented with 9 thematic variables. Distance to channel excluded. Right: Model 12-3 implement with 8 thematic variables. Distance to channel and precipitation excluded.*

First, the distance to channel variable is excluded from the modelling. An improvement in the number of areas with incoherent values is obtaining; however, the AUC decrease (figure 4.12). Finally, the precipitation information is excluded too. The model 1-3 presents a good predictive capability (AUC = 0.87) as well as a significant improvement in the resulting contrast weight map (figure 4.13).

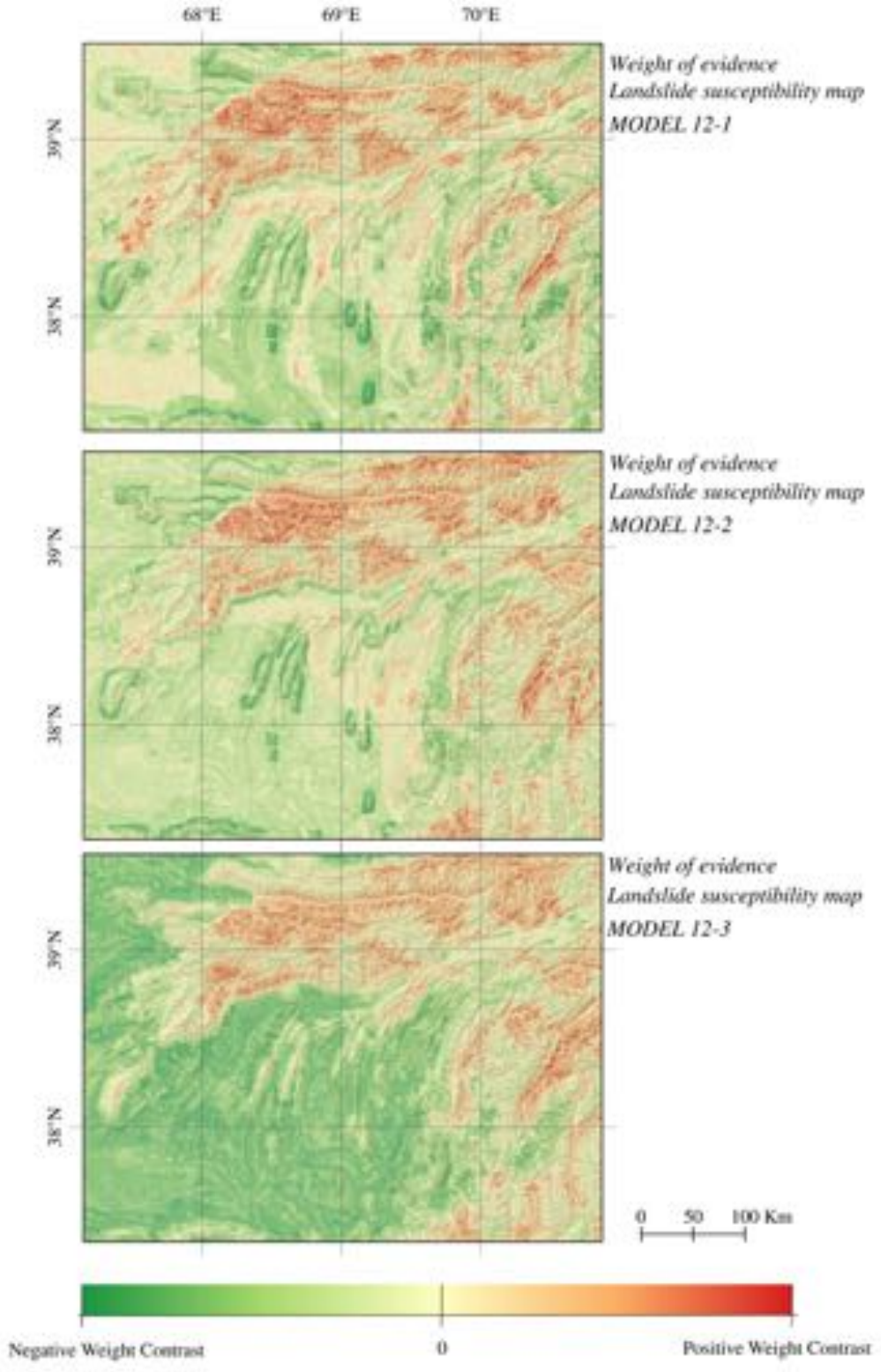


Figure 4.13: *Normalized total weight map resulting for the different variables combination based on the model 1-2 using the WOE approach.*

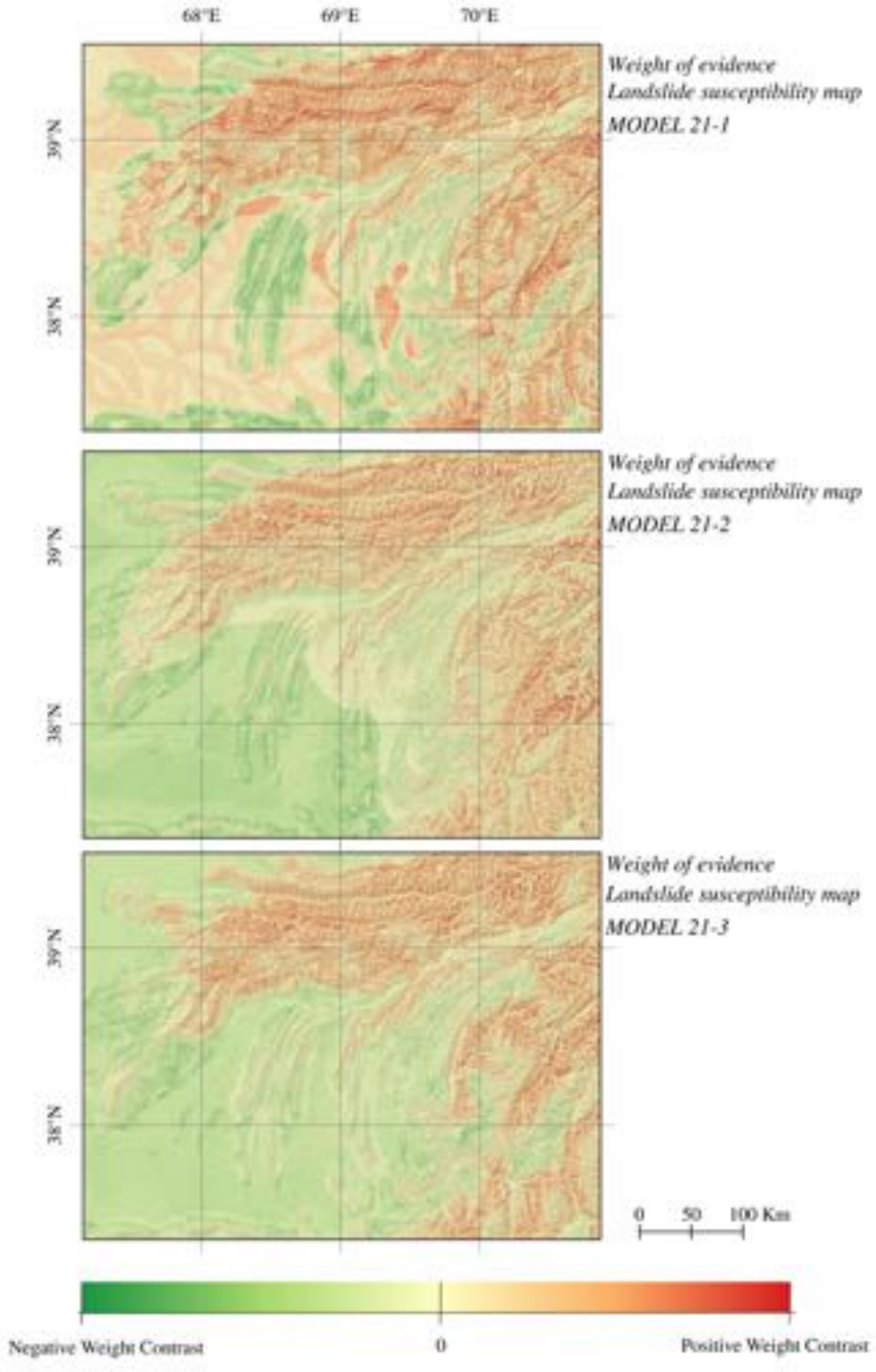


Figure 4.14: *Normalized total weight map resulting for the different variables combination based on the model 2-1 using the WOE approach.*

The model 2-1 is based on the second approach to discretize the continuous variables. The first model is computed based on 8 thematic variables. This model present a bad predictive power (AUC = 0.76) (figure 4.15) that is also supported by extensive areas with incoherent values. For example, for the flat landscape of the Tadjik basin, positive weight contrast values are presented; however, for the small mountain ranges located in the center of the area, high negative values are assigned. Those higher values follow the pattern given by the distance to channel (figure 4.14).

The model 2-1 is improved by the elimination of the distance to channel variable from the model (model 2-2) and the inclusion of TPI to increase the homogeneity of the areas in the Tadjik basin because the TPI present a negative spatial correlation with landslides for values close to 0 (figure 4.14). Even though the predictive capability of the model increase, zonation is still present. In previous models, the zonation is associated with the precipitation information. A final model (model 21-3) is computed without the distance to channel and the precipitation. These changes improve in almost a 10% the predictive capability of the model compared to the initial base model (model 21-1). Also, the resulting weight contrast map is more homogeneous and less incoherent areas are recognized.

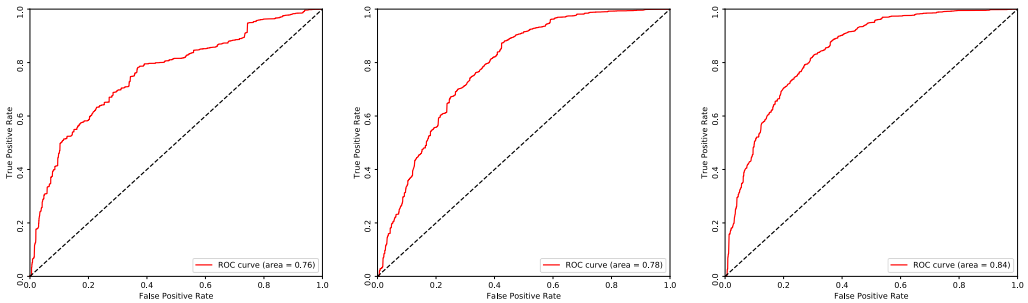


Figure 4.15: *Prediction capability for the model 2-1 and its variations. Left: Model 21-1 implemented with 8 thematic variables. Center: Model 21-2 implemented with 8 thematic variables. Distance to channel excluded and TPI included. Right: Model 21-3 implement with 7 thematic variables. Distance to channel and precipitation excluded, TPI included.*

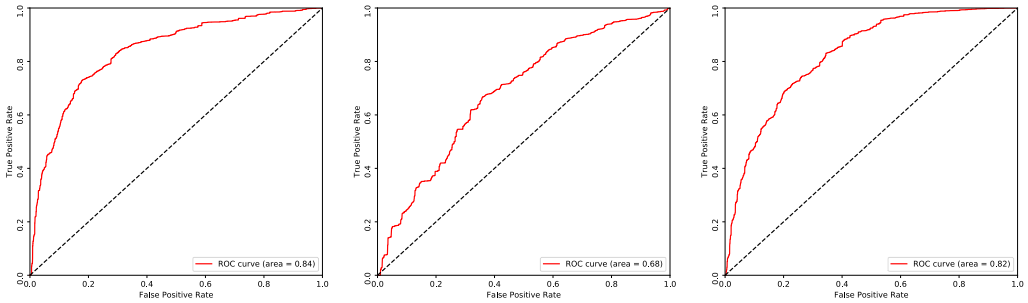


Figure 4.16: *Prediction capability for the model 2-2 and its variations. Left: Model 22-1 implemented with 11 thematic variables. Center: Model 22-2 implemented with 10 thematic variables. Precipitation excluded. Right: Model 22-3 implement with 10 thematic variables. Distance to glacier and precipitation excluded.*

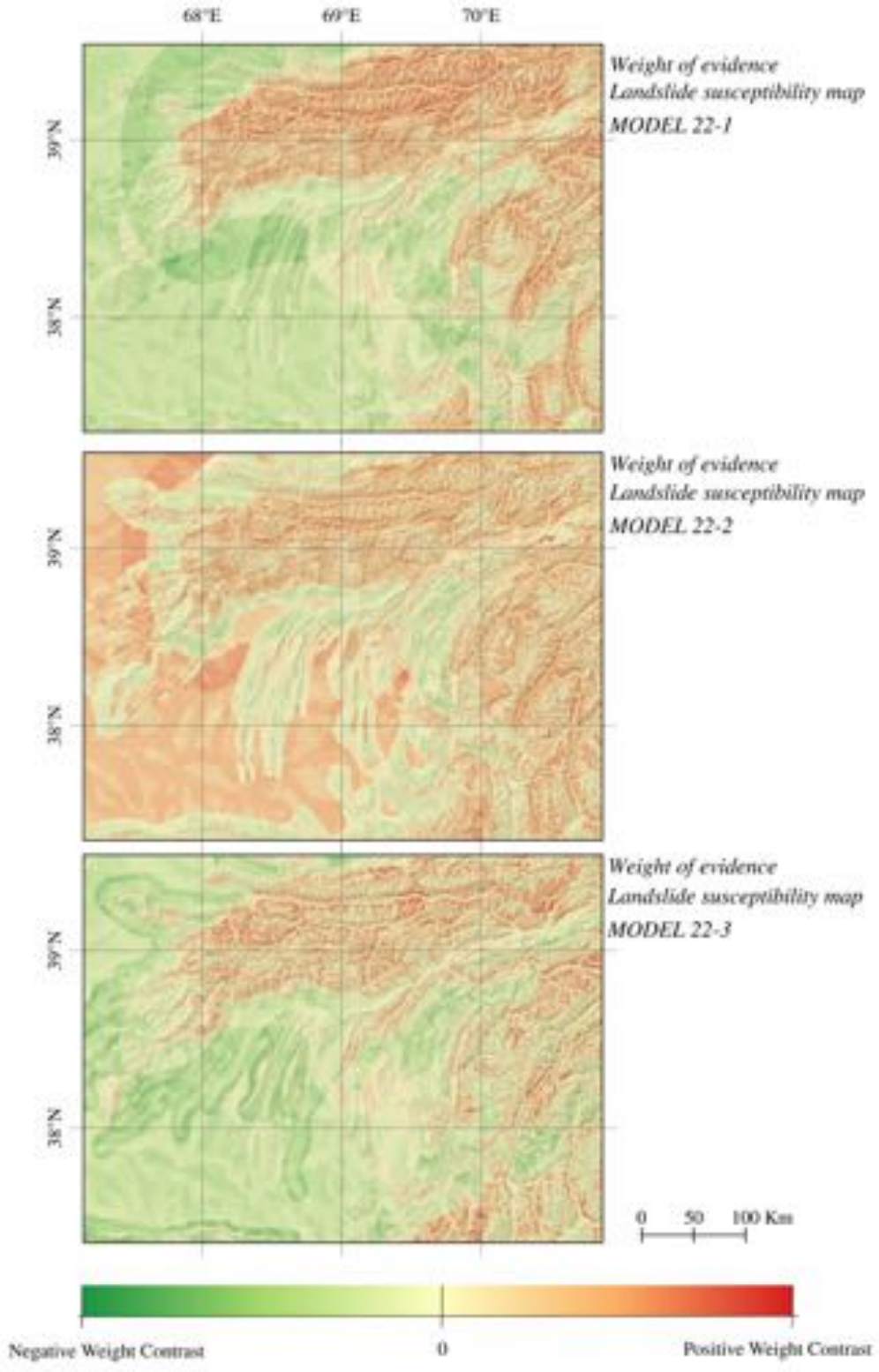
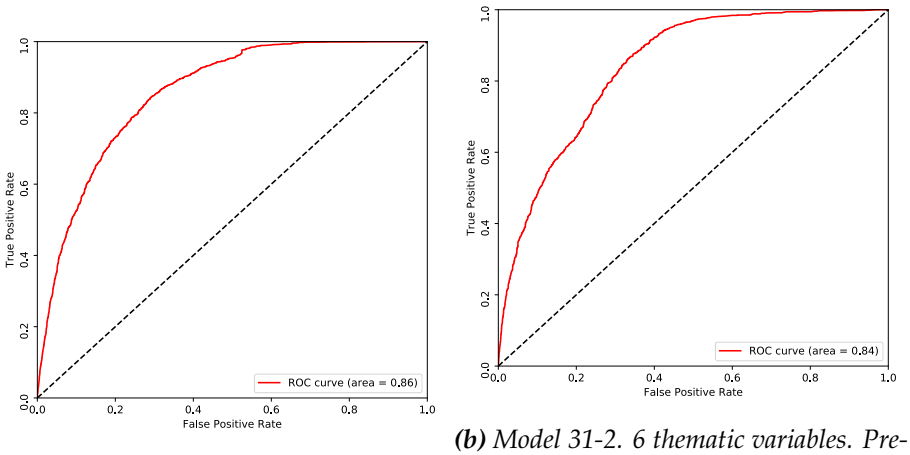


Figure 4.17: *Normalized total weight map resulting for the different variables combination based on the model 2-2 using the WOE approach.*

A second model is also proposed for the second approach of discretization, where, 11 thematic variables are taking into account. The resulting model has a good predictive power (AUC = 0.84) (figure 4.16); however, the weight contrast map exhibit areas with incoherent values associated with a zonation that creates irregularities in the whole area. As it was shown in previous models, those areas correlate to the influence of the precipitation in the model. Thus, the model 22-2 is computed without the precipitation variable. The predictive power of the model decrease dramatically along with the resulting map, which presents erratic values that are not coherent between each other. A third model (model 22-3) is implemented omitting the precipitation, the distance to glacier instead. The results evidence a decreasing in the AUC = 0.82 and in congruence with the resulting weight contrast map(figures 4.17 and 4.16).

For the third method, a single model is proposed because of the high conditional dependence among the variables. Model 3 has a good performance. It has a prediction capability of 86% (AUC = 0.86). The resulting map exhibit some areas with lower negative weight contrast values that follow the distribution of the NDVI. Also, values near 0 are enhanced in the areas associated with active faulting (figure 4.19).

The model 31-2 is computed without the precipitation information. The resulting model decreases the predictive capability (AUC =0.84); however, just a few incoherent areas are identified in the model. Also, the impact of the distance to fault decrease (figure 4.18).



(a) *Model 31-1. 7 thematic variables*

(b) *Model 31-2. 6 thematic variables. Precipitation excluded*

Figure 4.18: *Prediction capability for the model 3-1 and its variations. Left: Model 31-1 implemented with 7 thematic variables. Right: Model 31-3 implement with 6 thematic variables. Precipitation excluded.*

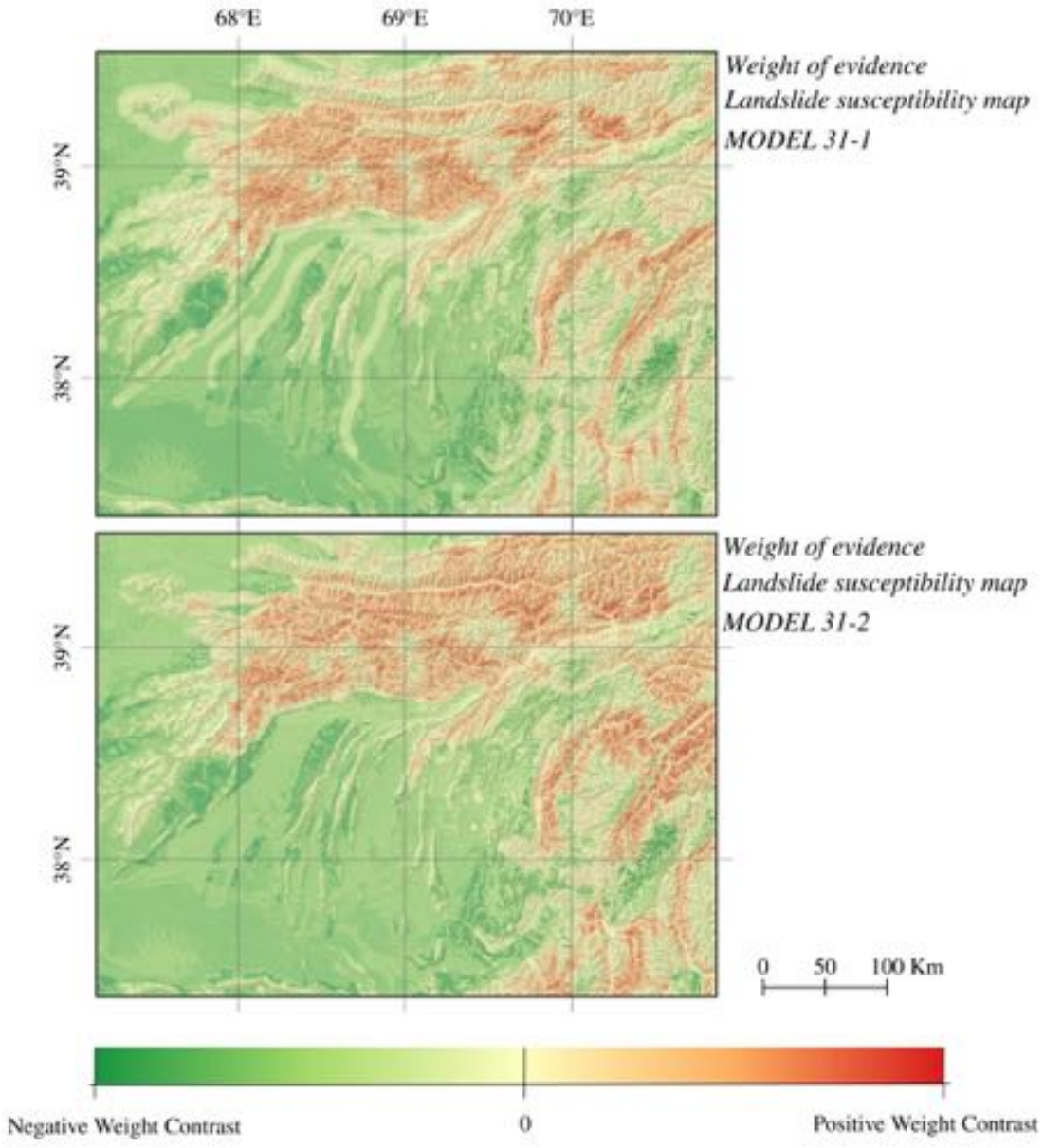
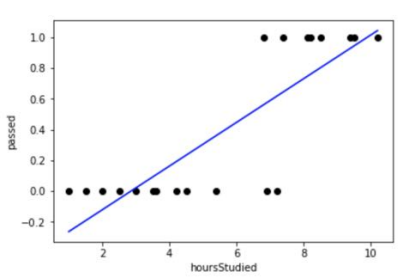


Figure 4.19: *Normalized total weight map resulting for the different variables combination based on the model 3-1 using the WOE approach.*

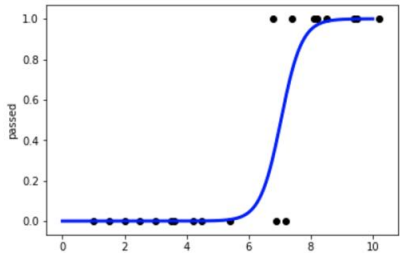
4.5 Logistic Regression

4.5.1 Method

Logistic regression is a statistical approach that aims to find the best fitting model that describes the relationship between a dichotomous characteristic of interest (dependent variable) and a set of independent (predictor or explanatory) variables.



(a) *Linear regression over data points*



(b) *Logistic regression fitting for the data points*

The dependent variable output is either 0 (no) or 1 (yes), so if linear regression is fit; values between 0 to 1 will be generated as well as impossible values - negative values and values higher than one- which have no meaning. That is so; logistic regression will fit better the data. The equation 4.22 gives the logistic function, where L is the curve's maximum value, k is the steepness of the curve and x_0 is the x value of a Sigmoid's mid-point.

$$f(x) = \frac{L}{1 + e^{-k(x-x_0}} \tag{4.22}$$

From equation 4.22, a standard logistic regression function can be express as equation 4.23, where $k = 1, x_0 = 0, L = 1$.

$$S(x) = \frac{1}{1 + e^{-x}} \tag{4.23}$$

The Sigmoid function graphical representation (figure 4.21) is a S-shaped curve with a finite limit of $X = 0$ where the values approaches to a negative infinity and $X = 1$ for a a positive infinity.

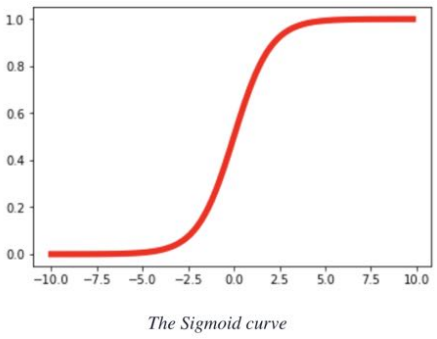


Figure 4.21: *Graphical representation of the Sigmoid's function (Equation 4.23)*

When $x = 0$, the Sigmoid function results is 0.5. Thus, if the output of the equation 4.23 is more than 0.5, it is possible to classify the outcome as 1 (yes), and if it is less than 0.5, the value is assigned as 0 (no); however, the results from the sigmoid function are not able to classify yes or no absolutely, but, the Sigmoid function results can be interpreted in term of probability of yes/no.

4.5.2 *Implementation*

The LR approach is implemented based on the standardization of continuous independent variables and the binary dependent variable that correspond to the landslide catalogue. Because Logistic regression is a modification of linear regression, the first step of the workflow is a multi-colinearity test to exclude the redundant variables. Then, the LR model is applied, and the results are tested in order to find the best model (figure 4.33).

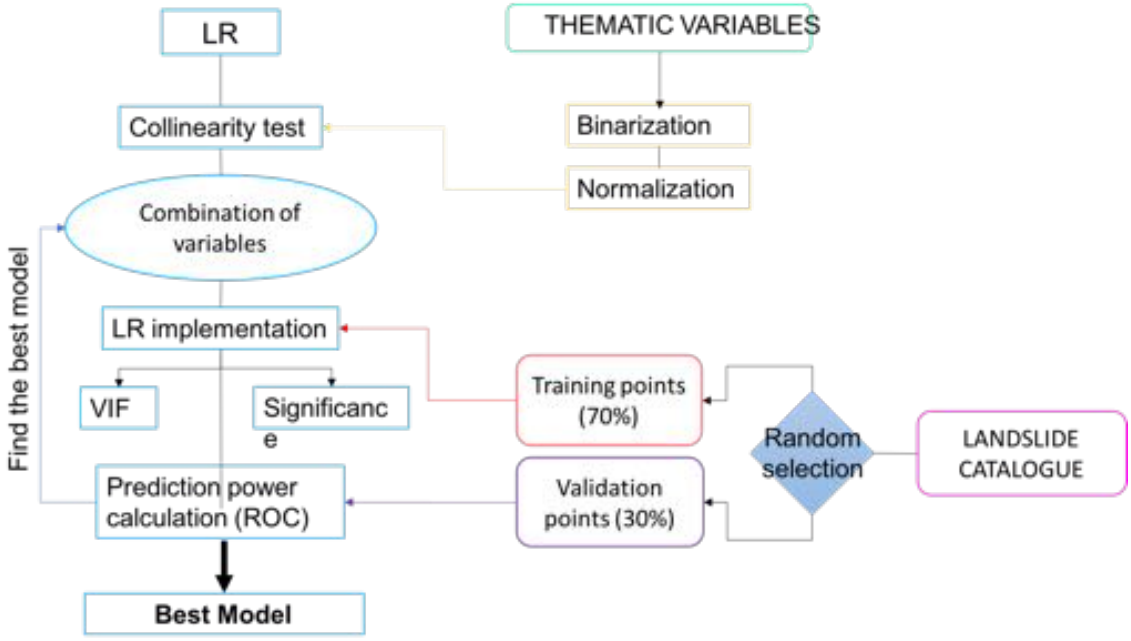


Figure 4.22: *Workflow followed to the calculation of the landslide susceptibility in the area based on the logistic regression approach*

In landslide susceptibility maps, the result of the logistic regression is the probability of belonging to a landslide class (e.g., Lee, 2005; Raja *et al.*, 2017, and references therein).This probability is represented by the general sigmoid curve equation (equation 4.24), where P is the probability of belong to a landslide class ($y = 1$), while z is a feature vector that represents the probability of success for any given observations.

$$P(y = 1) = \frac{1}{1 + e^{-z}} \tag{4.24}$$

The vector z is the logit transform of the log-odds of the probability of success for each of the independent variables (b). It is the simple linear regression model; where the y-intercept (b_0) moves the curve left or right (equation 4.25).

$$z = b_0 + b_1 x_1 + b_2 x_2 + ... + b_n x_n \tag{4.25}$$

4.5.2.1 Multi-collinearity test

Multi-collinearity is defined as the existence of near-linear relationships among variables. If the relationship is perfect ($R^2 = 1$), there will be a problem during the regression calculation because it will result in a division by 0; however, division by minimal quantities still distort the results. Hence, the first step to apply the LR approach is to determine whether multicollinearity exist (Lee *et al.*, 2018).

The correlation matrix among all the variables is obtained using R corr function, which returns the Pearson correlation coefficient between pairs of variables. Coefficient values with more than 0.5 or less than -0.5 suggest moderate to a strong linear relationship and are mutually exclusive for the model creation. Additionally, the variance inflation factors (VIFs) are calculated as support to the correlation matrix. The VIFs is the ratio of variance in a model with multiple terms, divided by the variance of a model with one term alone. It measures how much the variance is increased because of collinearity.

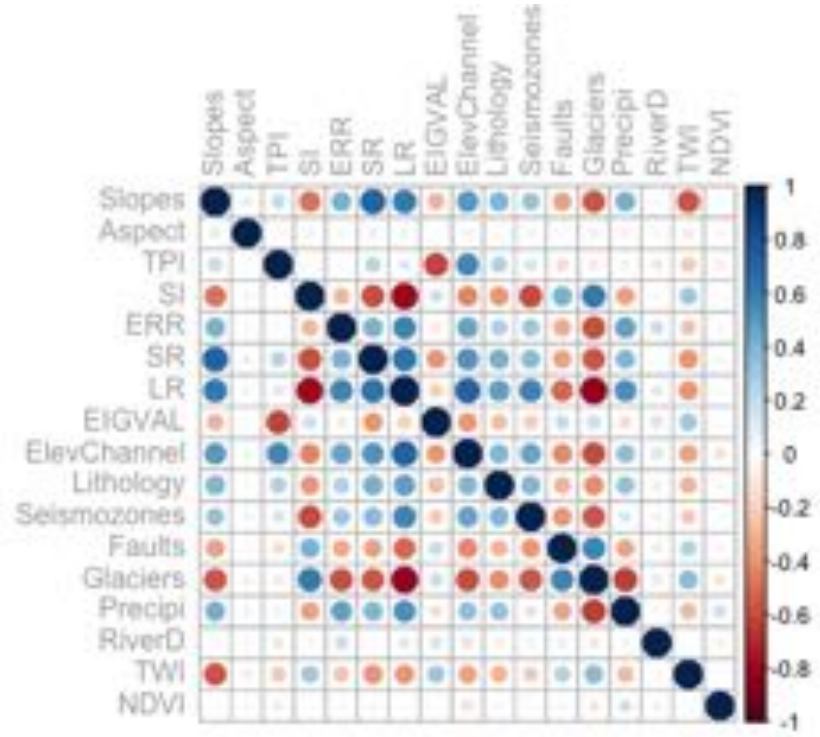


Figure 4.23: *Multicolinearity test results*

The test shows that relations exist between some of the geomorphological variables. A strong negative linear relationship characterizes local relief and surface roughness. Also, elevation relief ratio presents a weak negative linear relationship with SR. Contrary, Surface Index has a strong positive linear relationship with LR, while the EigenValues has a moderate positive linear relationship with TPI (figure 4.23).

Local relief, elevation above the channel and distance to glaciers are highly correlated with other variables. Apart from the geomorphological variables, local relief presents a negative strong linear relationship with elevation above the channel, and an intermediate linear relationship with lithology, seismozones, and precipitation. Contrary, a strong

positive relationship is presented between local relief and glacier distance.

As it was stated, distance to glaciers is characterized by a strong positive linear relationship with local relief as well as precipitation. However, the collinearity relationship that must be avoided is those that are strong and near to 1. In order to get a quantification of the multi-collinearity nature of each of the variables, the VIF is computed (table 4.7). The highest values of VIF characterized local relief, followed by elevation above channel; that means that the introduction of those variables in the model will change the variance by a significant factor.

Table 4.7: *Multi-collinearity test results- VIF*

Variable	VIF	**Variable**	VIF	**Variable**	VIF
Slopes	2.31	TPI	2.97	ERR	2.30
Aspect	1.02	SI	3.20	SR	2.63
Elevation above chanel	4.73	Eigen Values	1.30	LR	5.49
Lithology	1.52	Distance to fault	1.05	NDVI	1.10
Distance to Glacier	2.46	Distance to channel	1.24	TWI	1.33
Precipitation	1.75				

4.5.2.2 Combination of variables

A total of four models are created based on the results of the multicollinearity test as well as the understanding of the behaviour of each variable in relation to the dependent variable (table 4.8). The logistic regression is implemented for each model, and the results are analysed regarding of the significance of the variables to the model, the odds and the ROC.

Table 4.8: *Possible combinations of the variables based on the multicollinearity test. Double dots (:) indicate used of the interaction function of the LR.*

Thematic group	**Variable**	**Model1**	**Model2**	**Model3**	**Model4**
Lithology	Lithology14	x	x	x	x
Climatic and hydrological	Precipitation	x	x	x	:
	Distance to channel	x	x		x
	TWI	x	x		x
	Elevation above channel			x	
LandCover	NDVI	x	x	x	
Geomorphology	Slope	x	x	x	x :
	SR		x		
	ERR	x		x	
	SI			x	
	TPI	x			

Table 4.8: *Possible combinations of the variables based on the multicollinearity test. Double dots (:) indicate used of the interaction function of the LR.*

Thematic group	Variable	Model1	Model2	Model3	Model4
	EigenValues		x	x	x
Tectonic	Distance to fault	x	x	x	x

4.5.2.3 Model creation and improvement

The model 1 is computed using 10 predictive variables. However, the precipitation and the NDVI are characterized by low odds ratios; thus they do not play an essential role in the landslide probability assessment, and they are excluded (figure 4.25). The odds ratios > 1 reflects the association of the slopes, TPI and lithology with landslide occurrence, contrary, TWI, distance to channel, SI, ERR and distance to faults have odds ratios < 1, which indicates a negative association. A low presence of intermediate values characterizes the resulting landslides susceptibility map, and they are limited to certain areas, while very low values and very high values are predominant (figure 4.24).

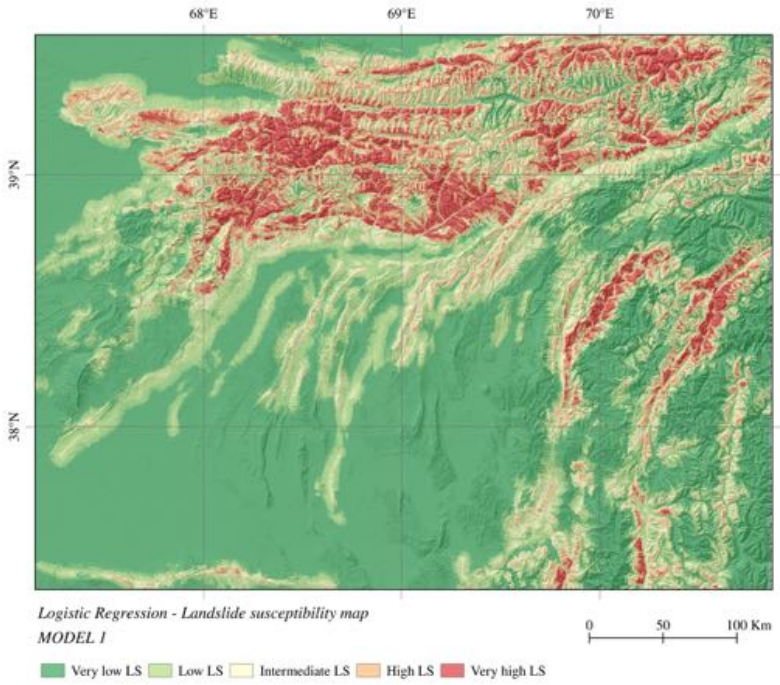
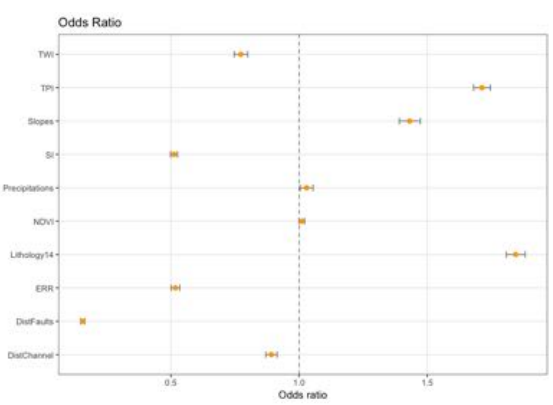


Figure 4.24: *Landslide occurrence probability map resulting for the different variables combination based on the model 1 using the LR approach.*

After the implementation of model 2, a similar result is obtained as the model 1 in terms of the significance of the variables. Precipitation and NDVI are omitted to obtain a final model with 7 variables. Similarly, the slope and lithology plays an important role in the prediction of the landslide occurrence, while TWI, distance to channel, eigenValues, and distance to fault decrease the probability (figure 4.27).




```
Call:
glm(formula = Classes ~ Slopes + TPI + SI + ERR + Lithology14 +
    DistFaults + DistChannel + TWI, family = binomial, data = TrainingData)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-3.1184  -0.6239  -0.0264   0.6868   3.4375

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -1.933315   0.020953  -92.27   <2e-16 ***
Slopes       0.350073   0.014437   24.25   <2e-16 ***
TPI          0.471175   0.009655   48.80   <2e-16 ***
SI          -0.541228   0.011898  -45.49   <2e-16 ***
ERR         -0.483086   0.014628  -33.02   <2e-16 ***
Lithology14  0.644068   0.009944   64.77   <2e-16 ***
DistFaults  -1.838140   0.027182  -67.62   <2e-16 ***
DistChannel -0.130531   0.011635  -11.22   <2e-16 ***
TWI         -0.219520   0.017183  -12.78   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 104560  on 75454  degrees of freedom
Residual deviance:  63743  on 75446  degrees of freedom
AIC: 63761

Number of Fisher Scoring iterations: 6
```

Figure 4.25: *Results of the implementation of the model 1 using logistic regression. Left: Odd ratio of each of the variables in the model 1. Odds ratios near to 1 reflect low significance for the model computation. Right: Model 1 statistical summary for the 8 predictive variables.*

The influence of the eigenValues is evident in the resulting landslide susceptibility map for model 2. Better delimited ridges are associated with a very high landslide susceptibility (figure 4.26) compared to the homogeneous areas represented by model 1 (figure 4.24). Low susceptibility areas associated with the active faults presence and they create some incoherences in the map.

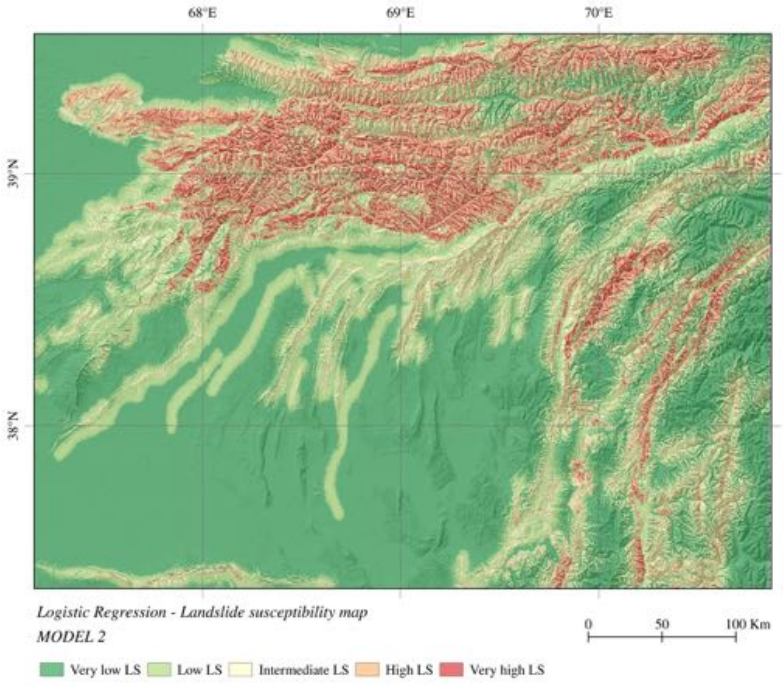
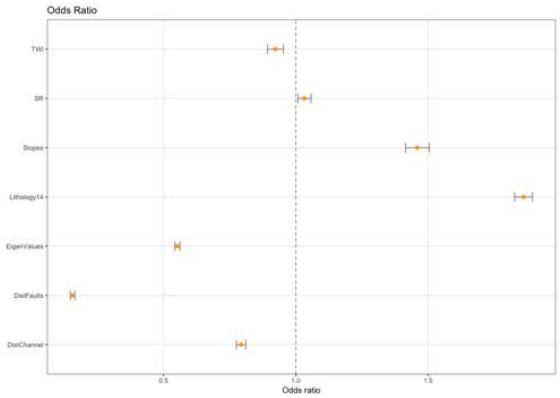


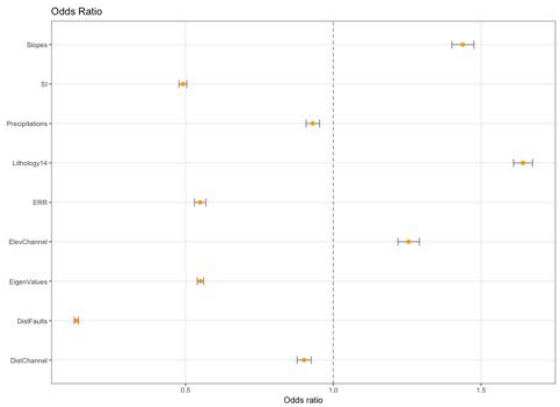
Figure 4.26: *Landslide occurrence probability map resulting for the different variables combination based on the model 2 using the LR approach.*




```
Call:
glm(formula = Classes ~ Slopes + EigenValues + SR + Lithology14 +
    DistFaults + DistChannel + TWI, family = binomial, data = TrainingData)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-3.1719  -0.6615  -0.0300   0.6489   3.4878

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -1.859549   0.020753 -89.605  < 2e-16 ***
Slopes       0.377650   0.015764  23.957  < 2e-16 ***
EigenValues -0.594522   0.009143 -65.024  < 2e-16 ***
SR           0.032012   0.012255   2.612    0.009 **
Lithology14  0.620804   0.009218  67.346  < 2e-16 ***
DistFaults  -1.858540   0.027464 -67.671  < 2e-16 ***
DistChannel -0.232222   0.011492 -20.207  < 2e-16 ***
TWI         -0.080587   0.016657  -4.838 1.31e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 104560  on 75454  degrees of freedom
Residual deviance:  64337  on 75447  degrees of freedom
AIC: 64353

Number of Fisher Scoring iterations: 6
```

Figure 4.27: *Results of the implementation of the model 2 using logistic regression. Left: Odd ratio of each of the variables in the model 2. Odds ratios near to 1 reflect low significance for the model computation. Right: Model 2 statistical summary for the 8 predictive variables.*

The model 3 is the only model where all the variables are significant. However, the odds ratios of precipitation and distance to channel are close to 1, meaning that they could be no significant for some another training dataset (figure 4.28). As it is identified before, slope and lithology influence the most the landslide probability occurrence. Also, elevation above the channel becomes an significant predictor for this combination of variables.




```
Call:
glm(formula = Classes ~ Slopes + SI + ERR + EigenValues + Lithology14 +
    DistFaults + DistChannel + ElevChannel, family = binomial,
    data = TrainingData)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-3.4523  -0.5804   0.1304   0.6239   3.5117

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -2.05733    0.02184 -94.219  < 2e-16 ***
Slopes       0.31324    0.01334  23.477  < 2e-16 ***
SI          -0.67114    0.01320 -50.863  < 2e-16 ***
ERR         -0.64565    0.01703 -37.918  < 2e-16 ***
EigenValues -0.63486    0.01024 -61.980  < 2e-16 ***
Lithology14  0.45045    0.01002  44.957  < 2e-16 ***
DistFaults  -1.93099    0.02814 -68.628  < 2e-16 ***
DistChannel -0.09692    0.01309  -7.403 1.33e-13 ***
ElevChannel  0.21785    0.01504  14.487  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 108067  on 77957  degrees of freedom
Residual deviance:  61931  on 77949  degrees of freedom
AIC: 61949

Number of Fisher Scoring iterations: 6
```

Figure 4.28: *Results of the implementation of the model 3 using logistic regression. Left: Odd ratio of each of the variables in the model 3. Odds ratios near to 1 reflect low significance for the model computation. Right: Model 3 statistical summary for the 8 predictive variables.*

The resulting susceptibility map predicts similar areas as high probability of landslide occurrence; however, by the use of SI and ERR instead SR, those areas are more detailed and better delimited; however, some areas show incoherences too, especially those located along the active faults (figure 4.29).

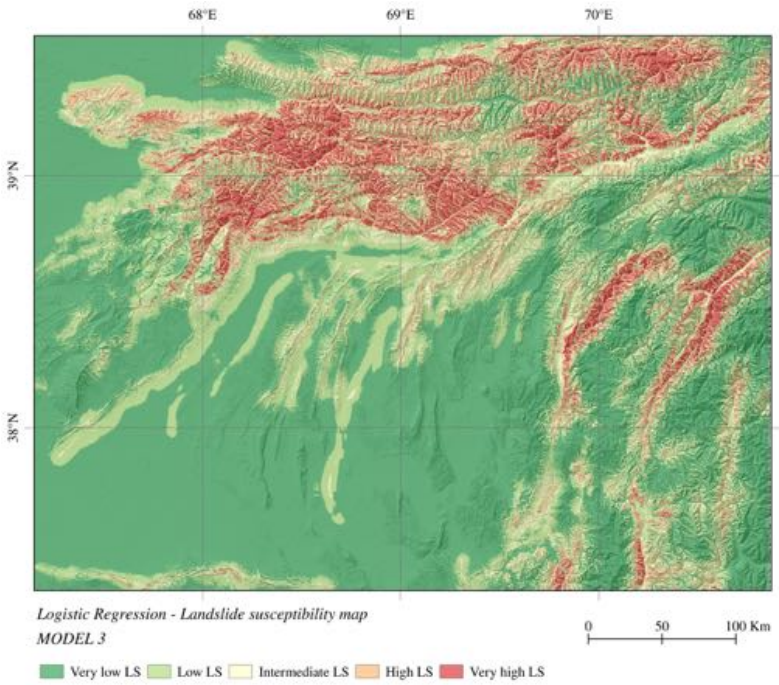


Figure 4.29: *Landslide occurrence probability map resulting for the different variables combination based on the model 3 using the LR approach.*

The model 4 is created by the use of the interaction capability of LR (represented by (:). Table 4.8). The model is computed using 8 variables that include the interaction among slope (positive odd ratios) and precipitation (odds ratios close to 1). This interaction does not present an increase in the landslide probability of occurrence, contrary, decrease it. Nevertheless, the resulting map is characterized by a reduction of the incoherences related to the active fault information, increasing the number of areas characterized by a very low landslides susceptibility (figure 4.31).

Even though the logistic regression approach deprecates the precipitation as a significant predictor variable, the addition or subtraction of it has an essential influence in the resulting map; particular in its interaction with the areas near to active faulting.

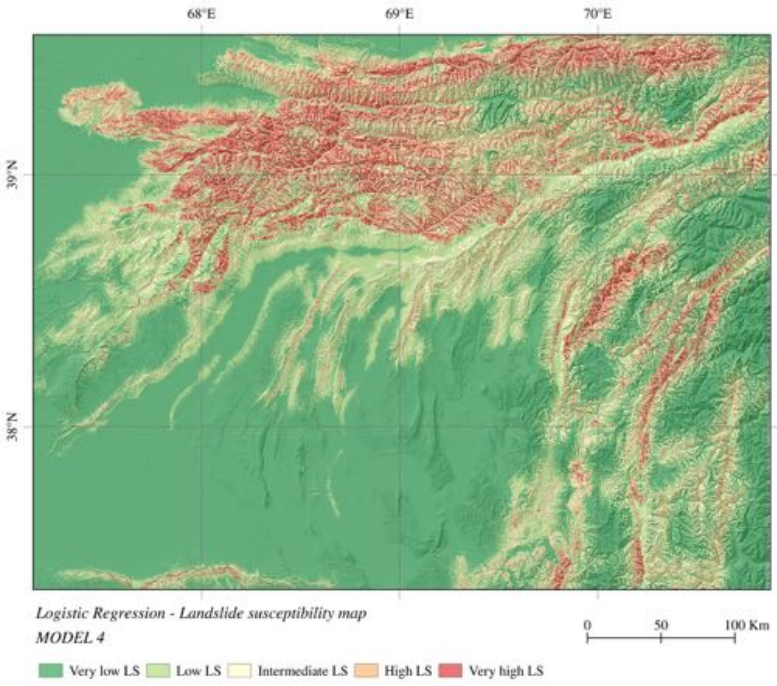
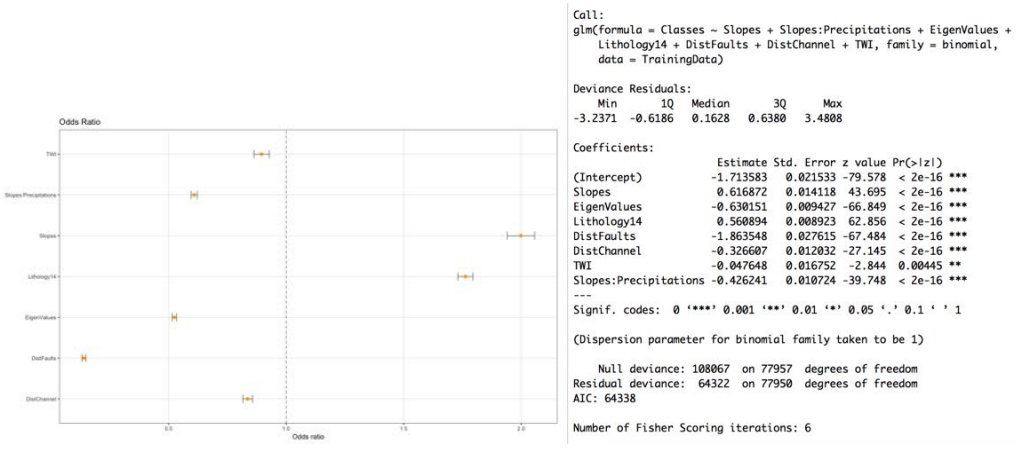


Figure 4.30: *Landslide occurrence probability map resulting for the different variables combination based on the model 4 using the LR approach.*




```
Call:
glm(formula = Classes ~ Slopes + Slopes:Precipitations + EigenValues +
    Lithology14 + DistFaults + DistChannel + TWI, family = binomial,
    data = TrainingData)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-3.2371  -0.6186   0.1628   0.6380   3.4808

Coefficients:
                          Estimate Std. Error z value Pr(>|z|)
(Intercept)           -1.713583   0.021533 -79.578  < 2e-16 ***
Slopes                 0.616872   0.014118  43.695  < 2e-16 ***
EigenValues           -0.630151   0.009427 -66.849  < 2e-16 ***
Lithology14            0.560894   0.008923  62.856  < 2e-16 ***
DistFaults            -1.863548   0.027615 -67.484  < 2e-16 ***
DistChannel           -0.326607   0.012032 -27.145  < 2e-16 ***
TWI                   -0.047648   0.016752  -2.844  0.00445 **
Slopes:Precipitations -0.426241   0.010724 -39.748  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 108067  on 77957  degrees of freedom
Residual deviance:  64322  on 77950  degrees of freedom
AIC: 64338

Number of Fisher Scoring iterations: 6
```

Figure 4.31: *Results of the implementation of the model 4 using logistic regression. Left: Odd ratio of each of the variables in the model 4. Odds ratios near to 1 reflect low significance for the model computation. Right: Model 4 statistical summary for the 7 predictive variables.*

4.6 Random Forest

4.6.1 Method

Random Forest is a multivariate statistical technique that implements the Bayesian tree or binary classification tree and a combination of the idea of bagging (technique that combines the predictions from multiple machine learning algorithms together to make more accurate predictions than any individual model) and random feature selection, to grow a forest of many trees. Random Forest algorithm utilizes bootstrap (method for estimating a quantity from a data sample) and random techniques to select the subsample of data and predictor parameters while growing an ensemble of trees. Besides, each tree is constructed using a different bootstrap sample of the data. Contrary to the traditional decision trees, RF split each node based on the best among of a subset of predictors randomly chosen at that node. This strategy leads to higher performance than other classifiers and makes RF robust against overfitting (Breiman, 2001).

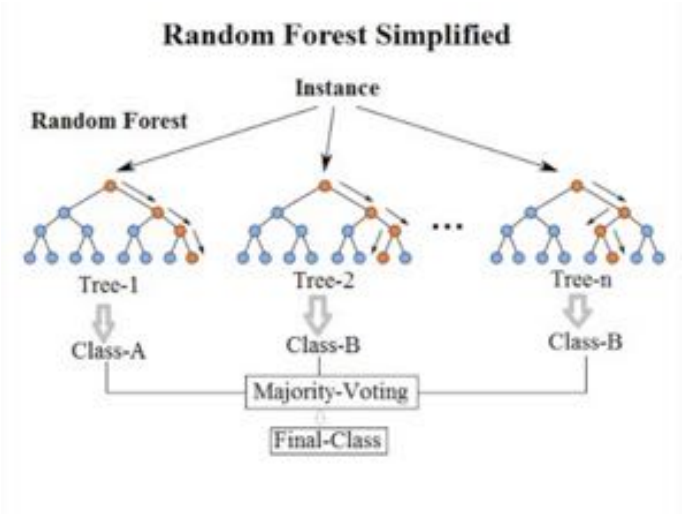


Figure 4.32: *Illustration of the implementation of the random forest algorithm. Source: (Koehrsen, 2007)*

To classify a new object based on the attributes, each tree gives a classification and the forest used the most vote class as a final decision (figure 4.32).

4.6.2 Implementation

The RF approach is implemented using the same input data as for the logistic regression approach (standardize continuous independent variables and binary dependent variable). The first step to implement a random forest is to determine the variables to be used; after that, different models can be created to be tested. For each model, unique model parameters are identified based on cross-correlation methods in order to get the best results from the RF implementation. The models are improved based on the importance of the variable to the creation of the RF results; as well as the ROC values and the model error (figure 4.33).

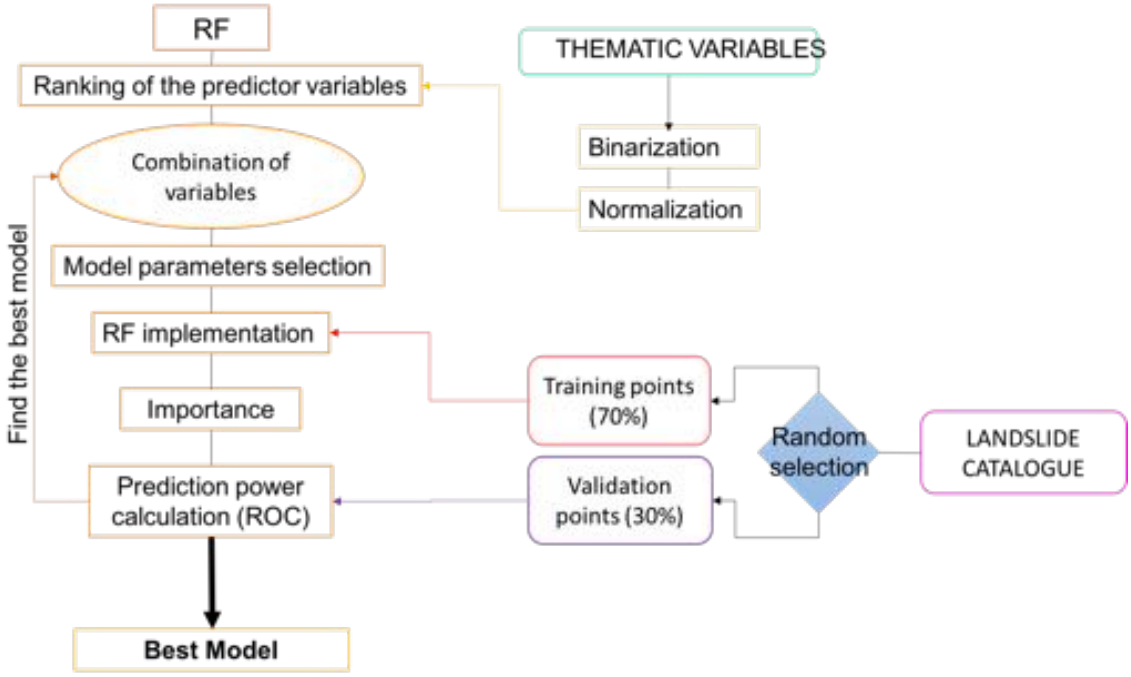


Figure 4.33: *Work flow followed to the calculation of the landslide susceptibility in the area based on the random forest approach*

4.6.2.1 Most important predictor variables

Based on the experience gained from the implementation of others susceptibility models; as well as the data exploration, it is possible to exclude some predictor variables without affect the landslide susceptibility result but improving the computation time. However, in order to test which variables are more important for the RF implementation, a ranking of all the variables is performed using Python 3 for the random forest classifier.

The first methodological step to implement the random forest is the selection of the variables that are more important to prepare an accurate map of landslide susceptibility. In order to decrease the number of possible combination, a ranking of the variables is implemented using generic model parameters (table 4.9). The ranking is created using recursive feature elimination from the Sklearn package in Python 3. The goal of recursive feature elimination is select features by recursively considering smaller and smaller sets of features.

Table 4.9: *Ranking of the variables according to its importance for the RF model.*

Ranking	Variable
1	Lithology
1	Distance to glaciers
2	EigenValue
3	LR
4	Distance to faults
5	SI
6	Elevation above channel

Table 4.9: *Ranking of the variables according to its importance for the RF model.*

Ranking	Variable
7	Precipitations
8	SR
9	TPI
10	ERR
11	Slopes
12	Distance to channel
13	NDVI
14	TWI
15	Aspect

This ranking is supported by the importance results obtained from the RF implementation using all the variable (figure 4.34). The importance is one of the outcomes from the Random Forest Classifier and are calculated based on the out-of-bag error (prediction error of the model) for each data variable during the fitting processing.

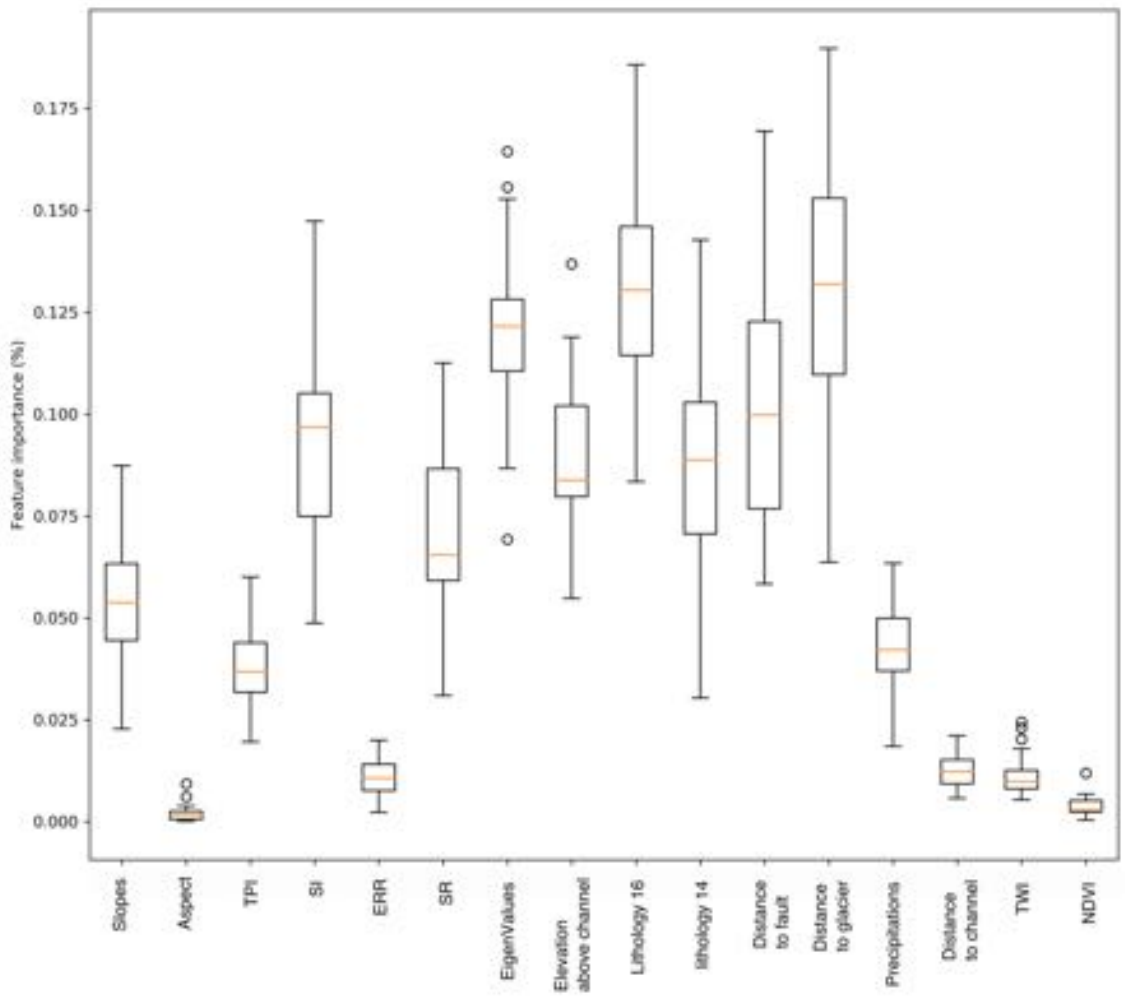


Figure 4.34: *Percentage of importance for each of the variables.*

4.6.2.2 Combination of variables

A total of five models are created based on the results of the ranking of the variables and the importance of each variable (table 4.9, figure 4.34). The first model corresponds to all the first 15 more important variables. Then, less important variables are excluded until finding the best model. From the ranking results presented in the table 4.9 and its comparison to the importance percentage for a generic RF model (figure 4.34) is possible to omit variables like aspect, NDVI, TWI and distance to channel. On the other hand, distance to glaciers is selected as a significant variable; however, the resulting susceptibility map tend to present inconsistent areas based on the buffers associated with this feature, so this variable is not taken into account.

Table 4.10: *Possible combinations of the variables used for the implementation of the random forest approach.*

Thematic group	Variable	Model1	Model2	Model3	Model4	Model5
Lithology	Lithology16	x	x	x	x	x
Climatic and hydrological	Precipitation	x	x		x	x
	Elevation above channel	x	x		x	
LandCover	NDVI					
Geomorphology	Slope	x	x		x	x
	SR				x	x
	SI	x	x	x	x	x
	Local Relief		x			
	TPI	x			x	x
	EigenValues	x	x	x	x	
Tectonic	Distance to fault	x	x	x	x	x

4.6.2.3 Selection of the model parameters

The RF approach is implemented in Python 3 using the library sklearn. This library allows the manipulation of the model parameter in order to get the best of the RF classifier. For each of the models, parameters like criterion (gini, entropy), max_ features (2, 3), min_ sample_ split (0.005, 0.01, 0.025, 0.050) and min_ sample_ leaf (0.005, 0.01, 0.025, 0.050) are selected based on cross-validation.

The criterion is the splitting decision method follow to create a tree. Gini index and the Entropy measure the node purity used to decide where to split the parent node among a child. The max_ features are the number of features considered at each split. It has a significant impact on the behaviour of the RF. For the classification problem, the sqrt(n_ features) is considered as a good approximation. The min_ sample_ split is the percentage of the data required to split the three in an internal node, while min_ sample_ leaf is the minimum of samples required to be at a leaf node.

Other parameters like n_ estimators or max_ depth are fixed. The number of trees should be as large as possible; however, it is limited by the computation time needed for a large forest. Thus, we used 100 based on previous studies (figure **??**) that suggest that the OOBE stabilize after 100 trees Trigila *et al.* (2015); Paudel *et al.* (2016); **?**. The depth of the trees is defined as the maximum possible depth.

For each selected variable combination the cross-validation is implemented obtaining the best parameters per model. The (table 4.11) summarized the criterion and the values of maximum features, the min samples split and the min sample leaf defined per model.

Table 4.11: *Parameters for each model based on the best estimator results.*

Parameters	Model1	Model2	Model3	Model4	Model5
Criterion	entropy	entropy	entropy	entropy	gini
max_features	2	2	2	3	2
min_samples_leaf	0.025	0.05	0.025	0.025	0.025
min_samples_split	0.005	0.01	0.005	0.025	0.025

4.6.2.4 Model creation and improvement

The implementation of the RF is made using the module RandomForestClassifier from Sklearn in Python 3.0, and for each model, the parameters are listed in table 4.11.

The model 1 is computed using 9 predictive variables. Slopes, TPI and precipitations are reported as the less important variables for the model, while Lithology, eigenValues, and local relief are considered as the most important ones. The ranking influence is observed in the resulting landslide susceptibility map (figure 4.35) where a very high landslide susceptibility is associated with the ridges. The influence of the geology is appreciated in the area of Pamir, where straps of very high, high and intermediate LS are presented.

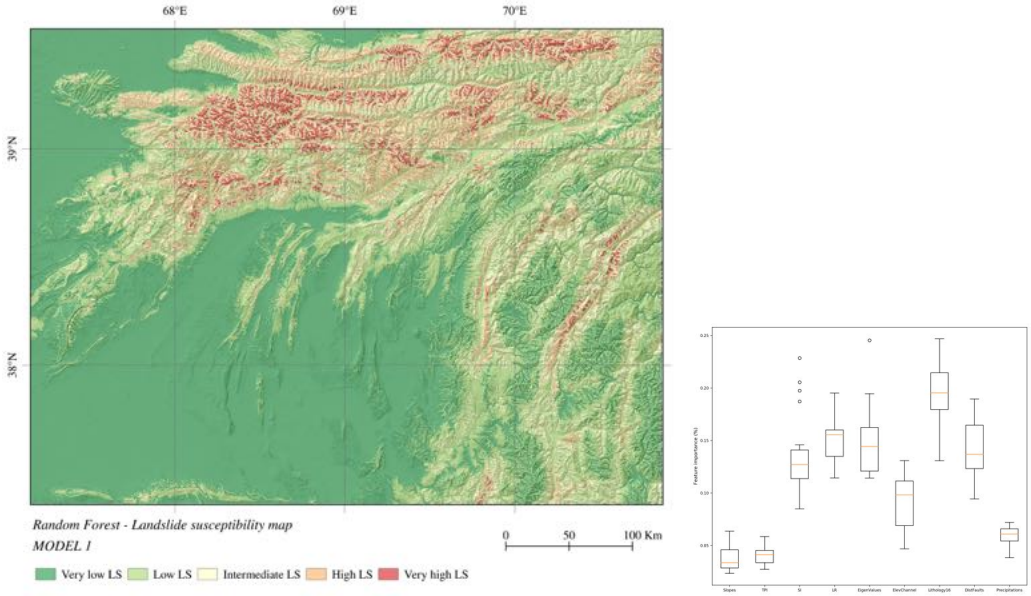


Figure 4.35: *Results of the implementation of the model 1 using the RF approach. Left: Landslide susceptibility map Right: Importances of each thematic variables used for the model 1.*

For the model 2 thematic variables are selected by the elimination of the TPI that is taken as a no relevant variable for the computation. The exclusion of this variable does not affect the importance of the others. However, the resulting map exhibit some changes. First, the areas characterized as high landslide susceptibility increase, giving a more broad view of the areas represented by a significant landslide density (figure 4.36).

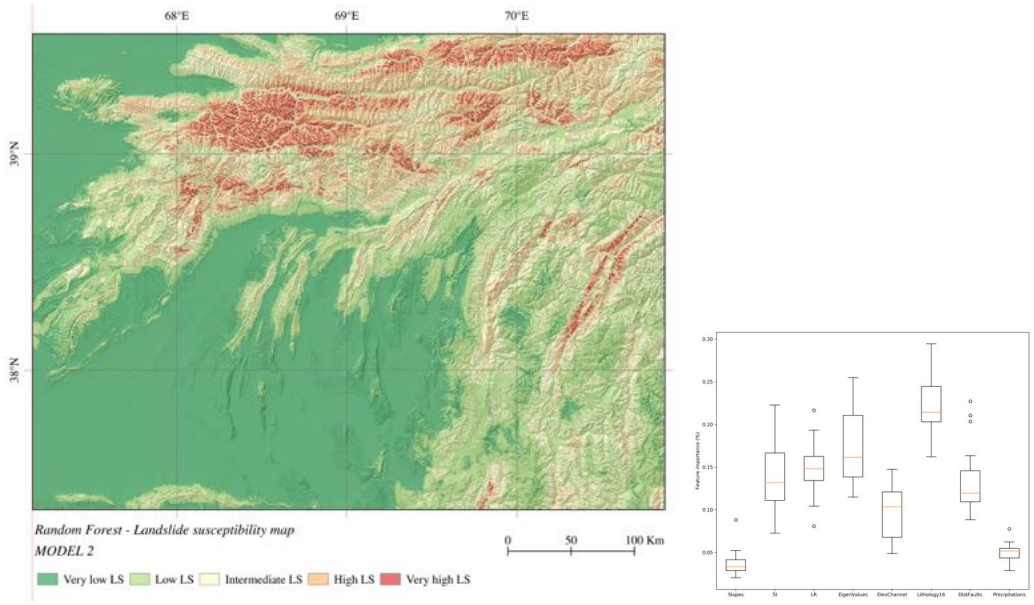


Figure 4.36: *Results of the implementation of the model 2 using the RF approach. Left: Landslide susceptibility map Right: Importances of each thematic variables used for the model 2.*

The model 3 is computed using the 4 more important variables, SI, eigenValues, lithology, and distance to fault. The predictive power of the model is good; however, the resulting landslide susceptibility map is characterized by the presence of areas with strong changes between landslide susceptibility class, probably associated with the lithological information. The increase in the LS is related to the presence of higher values of SI (figure 4.37).

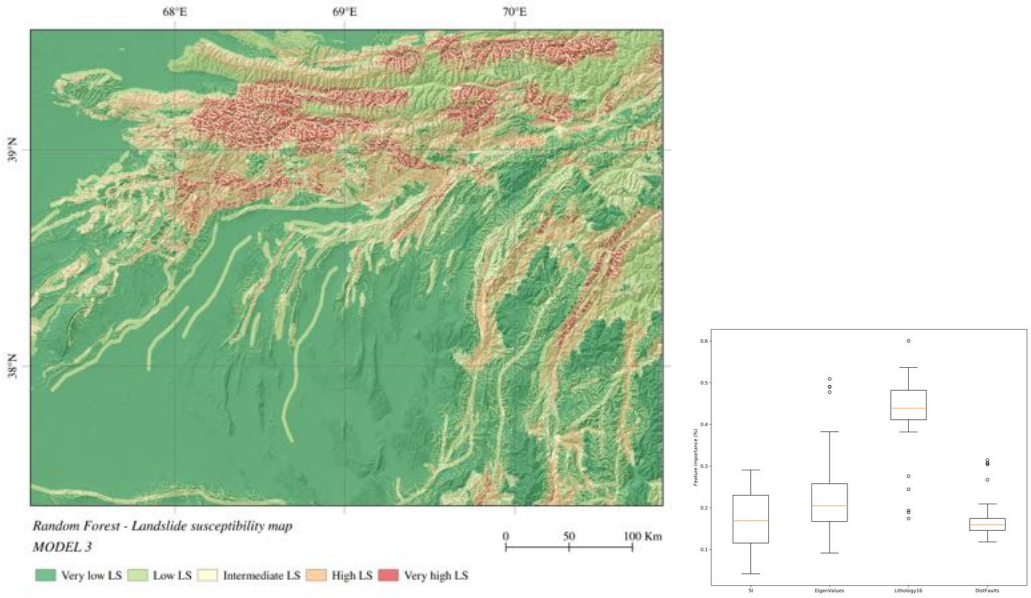


Figure 4.37: *Results of the implementation of the model 3 using the RF approach. Left: Landslide susceptibility map Right: Importances of each thematic variables used for the model 3.*

Model 4 explore the influence of the SR in the results. After the introduction of this variable instead of SR, LR or ERR the percentage of importance maintain unchangeably. Thus, the lithological informational and the eigenValues remains as the most important parameters (figure 4.38). The resulting landslide susceptibility maps are similar to the results obtained for the model 2.

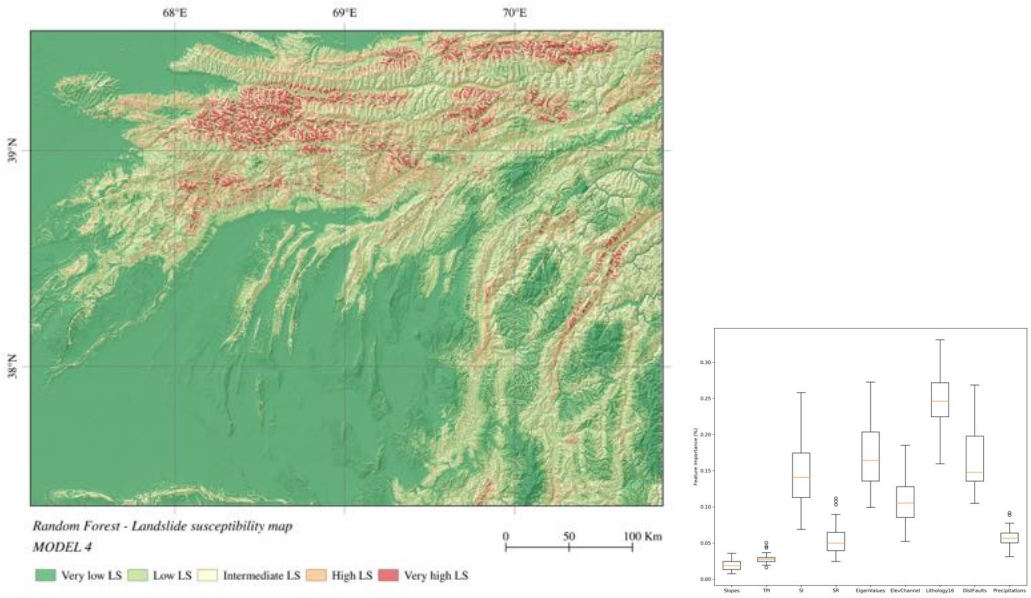


Figure 4.38: *Results of the implementation of the model 4 using the RF approach. Left: Landslide susceptibility map Right: Importances of each thematic variables used for the model 4.*

Finally, model 5 seeks for the improvement of the resulting map regarding homogeneity. For that, the elevation above channel and the EigenValues are excluded from

examining how other parameters increase its importance. As a result, no changes in the importance of the parameters is observed. They present a similar importance percentage as if the other parameters be included.

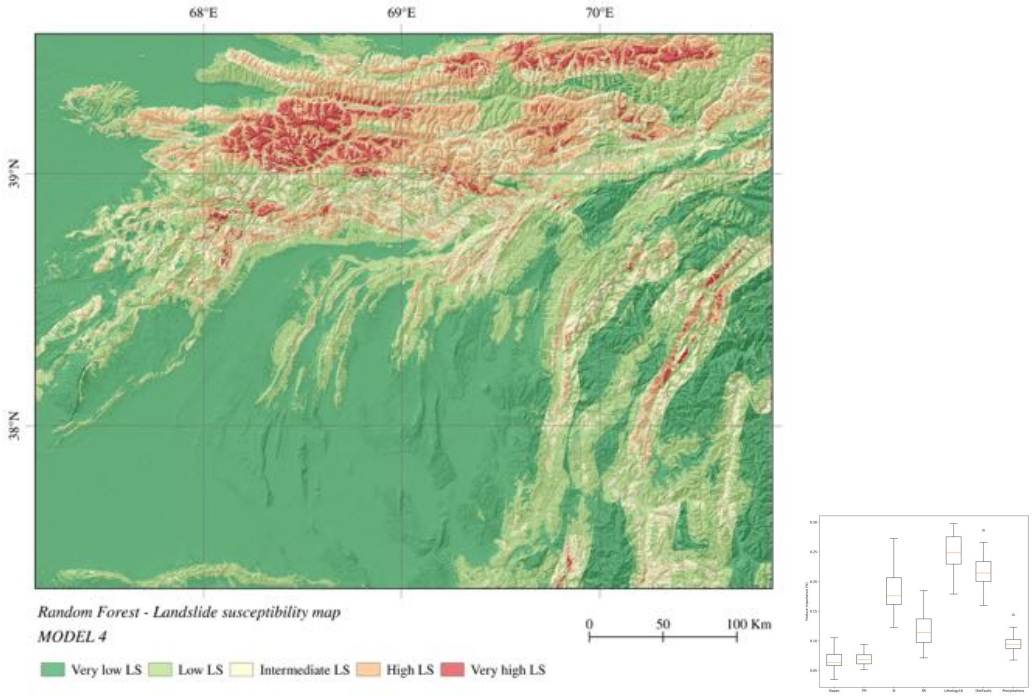


Figure 4.39: *Results of the implementation of the model 5 using the RF approach. Left: Landslide susceptibility map Right: Importances of each thematic variables used for the model 5.*

However, the resulting map is more generalized than others; however, the distribution of the landslide susceptibility classes seems coherent, and just a few areas can be considered as great incoherence..

Chapter 5

DISCUSSION

In this chapter, methodological steps and results are discussed. First, the biases of the landslide catalogue used for the work are presented. Secondly, the data preparation process is examined. Third, the importance of the thematic variables is discussed concerning the preference in the modelling implementation. Forth, a summary of the methodological performance of each of the statistical models implemented point out the limitations, advantages, and disadvantages. Fifth, the best model for each of the methodological approaches is selected and posterior they are compared between others. Finally, a comparison between the actual results and other landslide susceptibility assessments is presented.

5.1 Landslide catalogue and thematic variables

The landslide catalogue used for the landslide susceptibility modelling covers 0.02% of the area of study. It is considered a limited representation of the surface processes in the area. On the other hand, the distribution of the landslides is not homogeneous, a fact that introduces biases in the analysis. As it is possible to perceive in the figure 2.20, there is high landslide density in the Tien Shan, where many studies have been performed, leading to a complete landslide catalogue of the region. On the other hand, the distribution of the size of the landslide is not homogeneous either. Big landslides are reported in the Tien Shan, but just small landslides are mapped in part of Pamir. Some big landslides are reported in Pamir; however, they are not located inside the area.

The influence of the distribution of landslides in the area is recognized in the landslide susceptibility result. The Tien Shan is represented by a more significant amount of areas classified as higher susceptibility in comparison to the areas in the Pamir. Those areas follow the spatial distribution of the landslides. Contrary, the areas characterized by high landslides susceptibility in the Pamir are more limited to specific geomorphological-lithological characteristics rather than the landslides distribution. Thus, the results for the Pamir are less confident for interpretations and decision making than the results obtained for the Tien Shan because of the bias introduced by the landslide catalogue.

5.2 Data preparation

5.2.1 Discretization

Three different approaches were implemented to discretized the continuous variables. The breaking points resulting from the different approaches are similar or close to, for precipitation, SR and ERR. Significant changes in the breaking point are presented in the rest of the variables.

The first method based on the expert knowledge and literature review results in acceptable predictive capability (ROC = 0.84 - 0.87), however, the resulting landslide susceptibility map presented a lot of incoherent areas probably associated with the low number of landslides per class, that created difficulties in the model implementation.

The second method based on the resulting weight of contrast (W_c) exhibit an improvement in the resulting landslides susceptibility map; however, the predictive capability of the resulting models does not improve. This methodology requires an important computation time in order to create the representation of the weight of contrast against the variable values to select the breakpoints; because not all of the values in a continuous variable presents landslides. This creates curves with variations that goes to 0; as well as some erratic curves that decrease and increase strongly related to the values associated with low landslide density. After the elimination of the values with 0 landslides and the use of not single buy smaller classes for the continuous variables, an acceptable curve is created (4.5). After the creation of a smooth curve, changes in the curvature are selected as breakpoints. This methodology leads to bias in the discretization based on the expertise of the person that select the breakpoints as well as the quality of the curve.

The third method based on the selection of classes with a similar or equal number of landslides lead to the models with the higher predictive capability (ROC = 0.86). The resulting landslide susceptibility is coherent and the computation time is short. Also, because the method warranty that each of the classes contains a significant number or landslides, the problems related to the implementation of the WOE model are reduced.

5.2.2 Binarization

The logistic regression approach can use as input continuous and discrete variables. The only discrete variable used for the implementation for the method is the lithological information, categorized in 16 different classes. However, those classes represent the age of the rock, not the importance for the landslides susceptibility, as the logistic regression model understands it. There are two ways to introduce the lithological information. The first one to create a binary input per class, that result in a total of 16 rasters, or to use of the resulting weight of contrast W_c from the WOE approach.

The second method is selected, and the results are consistent. It can be seen in a comparison between the result of the same model using the lithological information discriminate in 14 classes, where the Mesozoic and the Paleozoic sedimentary units are grouped and the one with 16 classes where the sedimentary units are discriminated by periods. From the figure 5.1 it is possible to recognize in purple color the shape of the Paleozoic units as a whole, indicating that the geology14 influence the prediction in the Tien Shan as well as in the Pamir in a positive way. The contrary is observed for the Mesozoic sedi-

mentary units, where the Cretaceous unit that is separated from the Cretaceous-Jurassic sequence is associated with higher landslides susceptibility values. The changes related to yellow and blue colors (smaller changes between the models) are related to the influence of other variables.

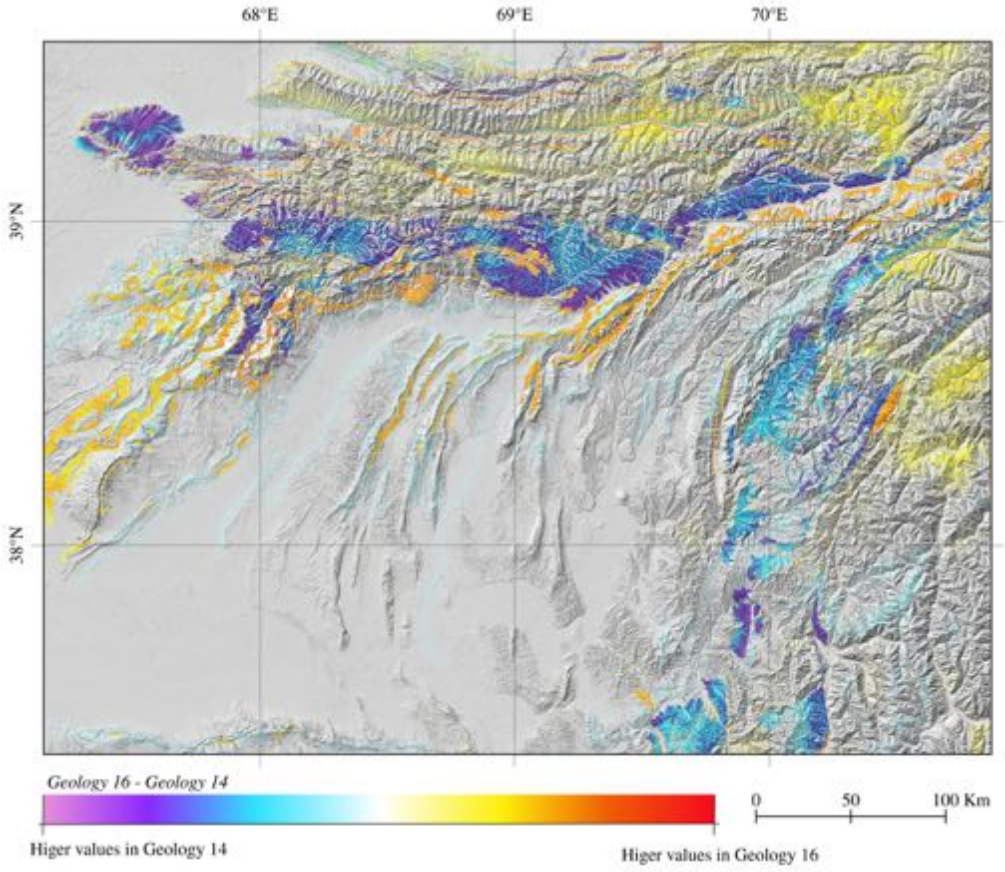


Figure 5.1: *Map of the differences between the geology 16 (Mezosoic and Paleozoic sedimentary units discriminated) to the geology 14 (Mezosoic and Paleozoic sedimentary units grouped)*

5.3 Important variables

Each of the statistical approaches implemented used different variables to obtain the best landslide susceptibility map; however, some of the less important variables are common for two or the three models. The first variable without relation to the landslide occurrence is the Aspect. It is broadly used as a predictor for LS; however, its importance is based on a structurally related origin for the landslides, like the slopes in a synclinal-anticline landscape or in areas where the slope orientation controls the amount of sun, wind or precipitation received. Aspect can be a significant predictive factor at smaller scales than the one used for the study.

The NDVI is used to represent the land cover in the area; however, its implication in the modelling of the LS is not decisive. It is because of the homogeneity in the vegetation type in the area, where grasslands are abundant are distributed all around the mountain-

ous areas (figure 5.2).

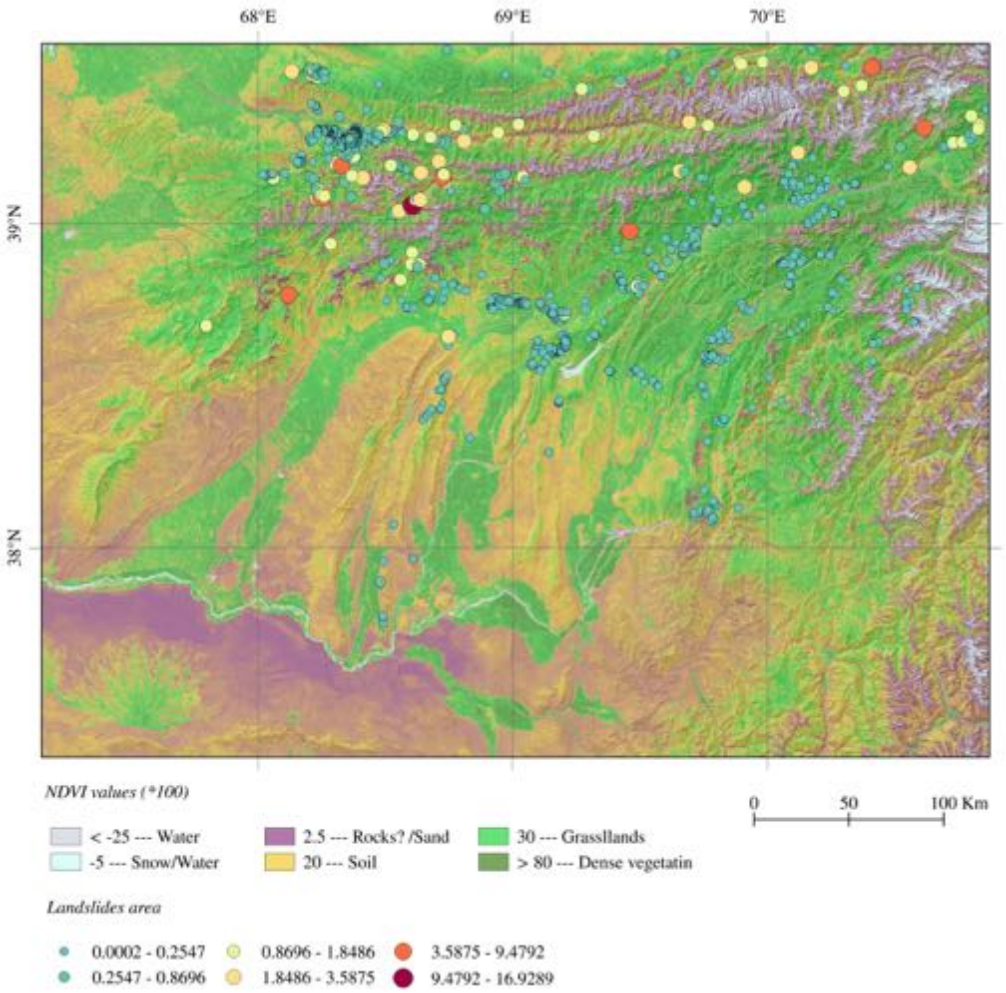


Figure 5.2: *NDVI values classified as water, snow/water, Rock/sand, Soil, Grasslands and Dense vegetation based on the most common thresholds. Landslides area superimposed.*

The TWI is used to discriminate between areas based on the saturation. However, TPI values below 20 characterized the area creating a similar pattern that decreases the possibility of associate specific values with the presence or absence of landslides. Distance to channel is another commonly used predictor variable; however, other hydrological relations can represent better the landslides distribution. Parameters like elevation above the channel follow better the relation between the streams and the landscape, then a Euclidean distance.

The distance to glaciers is identified as an important variable because it is a specific condition of the area; however, the use of the variable in the models implementation frequently created incoherent areas and zoning that decrease the predictive capability of the model and the credibility of the resulting landslide susceptibility map. Probably a similar approach which use elevation instead Euclidean distance can be explored in future works.

The lithological information is an essential input because it is directly related to the materials that can create instabilities. For the study area, Paleozoic sedimentary units are associated with the highest landslide density (figure **??**), especially, the Devonian unit, followed by the Carboniferous and finally the Permian. On the other hand, from the Mesozoic sedimentary units, the Cretaceous and Jurassic units present a high landslide density. All those units are associated with marine sedimentary deposits where siltstone, dolomites, limestones, shales, coal lenses, sandstones, and conglomerates are predominant.

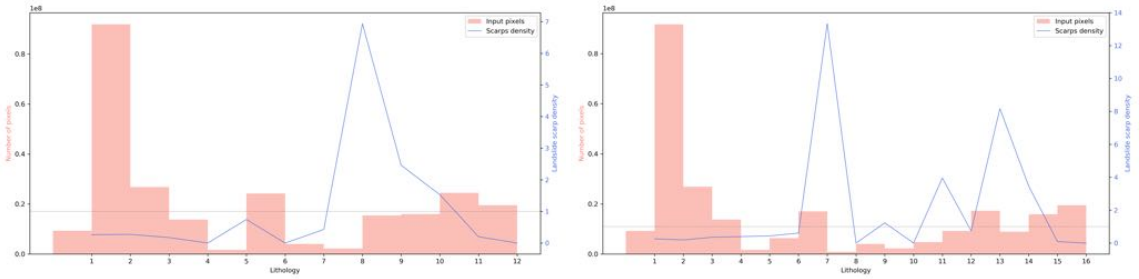


Figure 5.3: *Histogram of the geological units in the area and its associated landslide density. Left: 1. Quaternary, 2. Neogene, 3. Paleogene, 4. Paleogene/Intrusive, 5. Cretaceous, 6. Triassic/Jurassic, 7. Permian/Igenous, 8. Carboniferous, 9. Carboniferous/Igneous, 10. Permian-Silurian-Devonian, 11. Cambrian/Precambrian, 12. Glacier areas. Right: 1. Quaternary, 2. Neogene, 3. Paleogene, 4. Paleogene/Intrusive, 5. Cretaceous, 6. Cretaceous/Jurassic, 7. Jurassic, 8. Triassic/Jurassic, 9. Permian/Igenous, 10. Carboniferous, 11. Carboniferous 12. Carboniferous/Igneous, 13. Devonian, 14.Silurian, 15. Cambrian/Precambrian, 16. Glacier areas*

The precipitation information was not selected as an ideal variable in most of the models because of the of zonation effects in the resulting landslide susceptibility map which introduce incoherent values. However, the precipitation increase and enhance others predictor factors. This situation is evident in the logistic regression when the extraction of the precipitation variable from the models decrease the probability of landslide occurrence due to the distance to a fault.

The selection of the appropriate scale and kernel size for the computation of the thematic variables derived from DEM is another factor to discuss. Some authors implement techniques like random forest (Paudel *et al.*, 2016) to determine the optimal scale of the DEM to compute each of the geomorphological parameters. In this study, the same DEM is used to compute all the variables; however, different kernel sizes are used for the parameters. The definition of the kernel size for parameters like TPI, SR, ERR, and local relief allow the selection of which features are highlighted or identify by the variable. For example, for the TPI, large kernel sizes will reveal major landscape units, while smaller values highlight smaller features such minor valleys and rides.

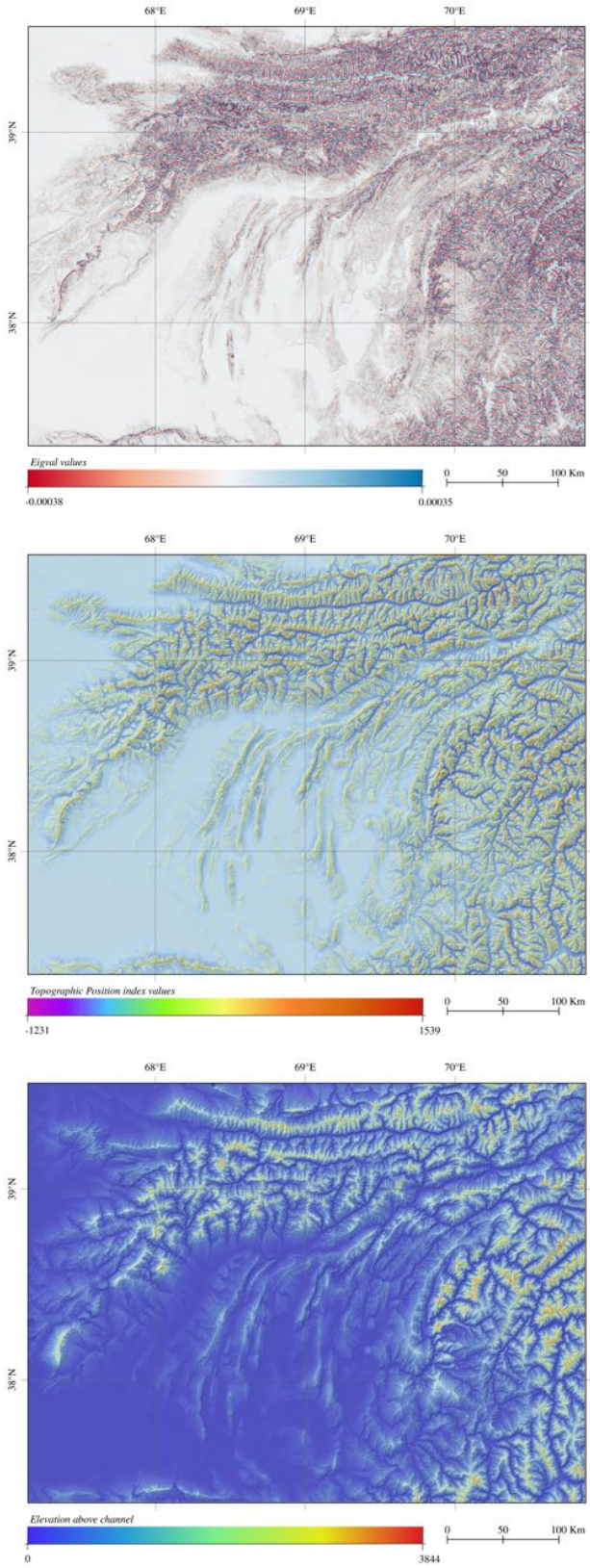


Figure 5.4: *Comparison between the landscape features identified by the EigenValue, TPI and elevation above channel.*

The eigenvalues are used as a significant predictor for more than one model. It can be associated with the ability of this variable to discriminate ridges and valleys, similarly to the TPI that is used for most of models, but represents lower importance or significance. The similarity between the landscape features that are depicted for different variables crates redundant information that is deprecated by the statistical models. Similar observations are made with the elevation above channel, that separates the bottom of the valleys form the areas located higher to them (figure 5.4).

Nevertheless, it is essential to recognize that from each of the different variable information can be extracted. Also, the spatial association of the values and the landslide performed differently. For the EigenValues, the significant landslide density is located in the valley; however, from the TPI is possible to discriminate also the position of the slope where the landslides occur. The last information can also be supported by the areas identified by the elevation above the channel as more related to the landslide occurrence.

Similar behaviour is observed with the slope, considered one of the most important predictor variables, however, for this study, other geomorphological parameters are identified as more significant predictors. Because high slopes dominate the area, geomorphological variables that discriminate between different topographic domains provide more pieces of information. The Surface roughness, for example, highlights those areas where the erosion is modelling the landscape, thus, the unstable areas. Additionally, because the slope is a function of the change of the elevation, other parameters can provide similar information to the slope.

The ERR, SI, and local relief enhance similar areas (figure 5.6) Those parameters seek to the identification of the possible tectonic influence in the landslide evolution and can be easily exchanged between each other, however, concerning statistical behaviour, the local relief is identified as a variable that can be easily predicted from other variables (Multi-collinearity). For the logistic regression approach, the ERR and SI are associated with similar odds ratios, while for the random forest approach, the SI is a relevant parameter. It is because of the nature of the SI that efficiently discriminated poorly incised surface and dissected landscape.

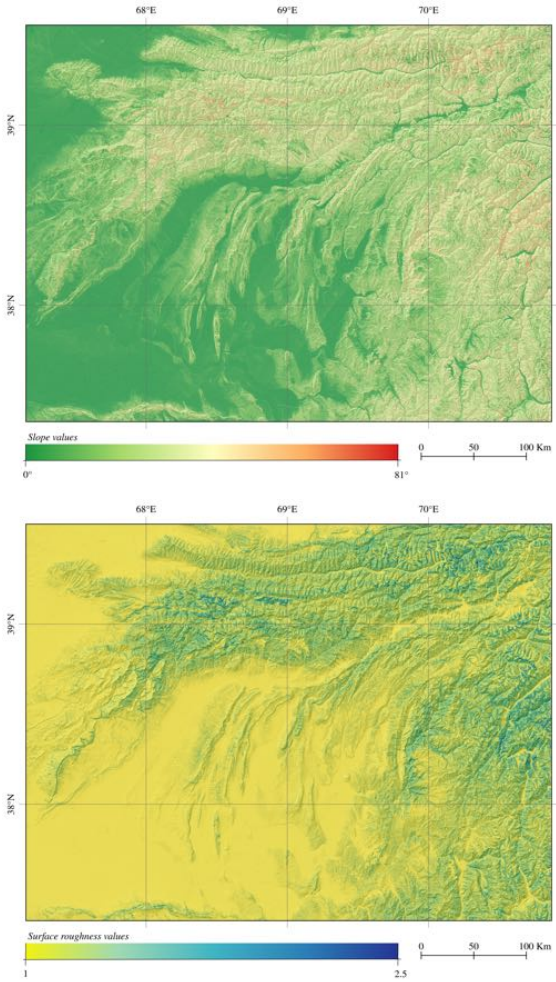


Figure 5.5: *Comparison between the landscape features identified by the slope and the SR.*

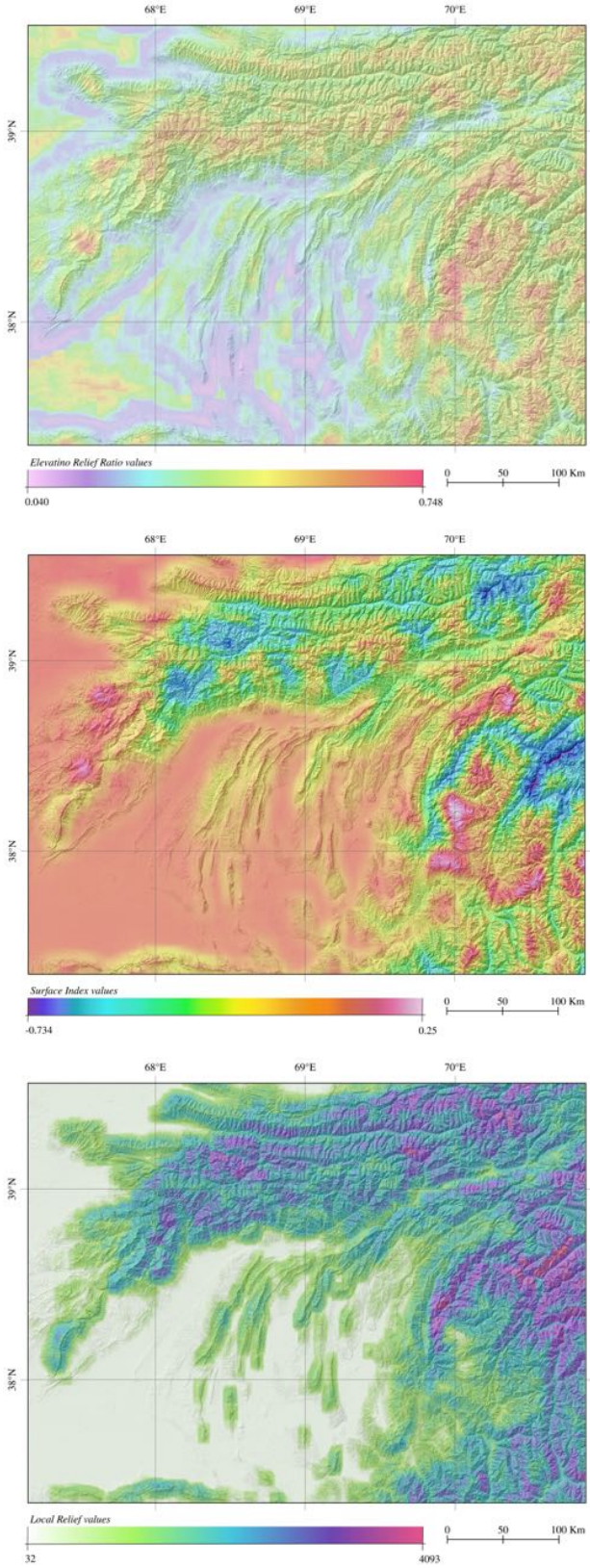


Figure 5.6: *Comparison between the landscape features identified by the EigenValue, TPI and elevation above channel.*

5.4 Landslide susceptibility models

Three different statistically based approaches have been implemented to obtain a landslide susceptibility assessment for the area of Western Pamir and South Tien Shan. The first model, WOE, is characterized by its easy implementation and the calculation of the weights related to each factor is well documented in the literature. Also, the results of the model are easy to understand, and the implementation of the workflow is very intuitive. However, the model can easily overestimate or underestimates susceptibilities based on the number of landslides per classes when they are not evenly distributed, making it less suitable for areas with incomplete landslide catalogue. In order to avoid it different methodologies were tested; however, the implementation of the approach using a majority of continuous variables becomes time-consuming. On the other hand, in order to determine the appropriate combination of variables, a sequential and lengthy process need to be done to obtain the best model. Finally, the resulting LS maps are not comparable in terms of degree of susceptibility, because the weight contras (W_c) is not standardized; adding another methodological step to the process.

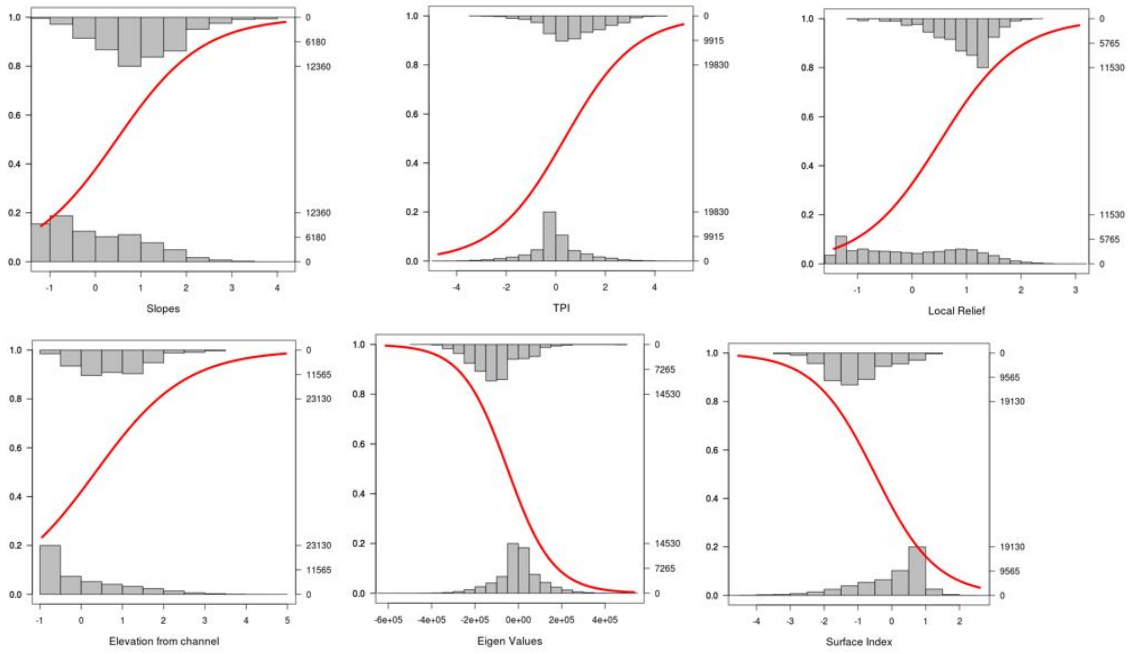


Figure 5.7: *Comparison of the binary variables and its fit within the logistic regression function*

Logistic regression allows faster implementation of the landslide susceptibility through a selection of the relevant variables based on the multi-collinearity results. However, it is possible to recognize that not all the variables fit the sigmoid equation or the logistic regression model, introducing the question of whether this model represents the real behaviour of the variables. The slope, TPI, Local relief, and elevation above the channel present behaviour that fits tightly to the logistic regression (figure 5.7). Similarly, eigenValues and Surface index fits a transposed sigmoid function because more landslides occur for the smaller values of the parameters.

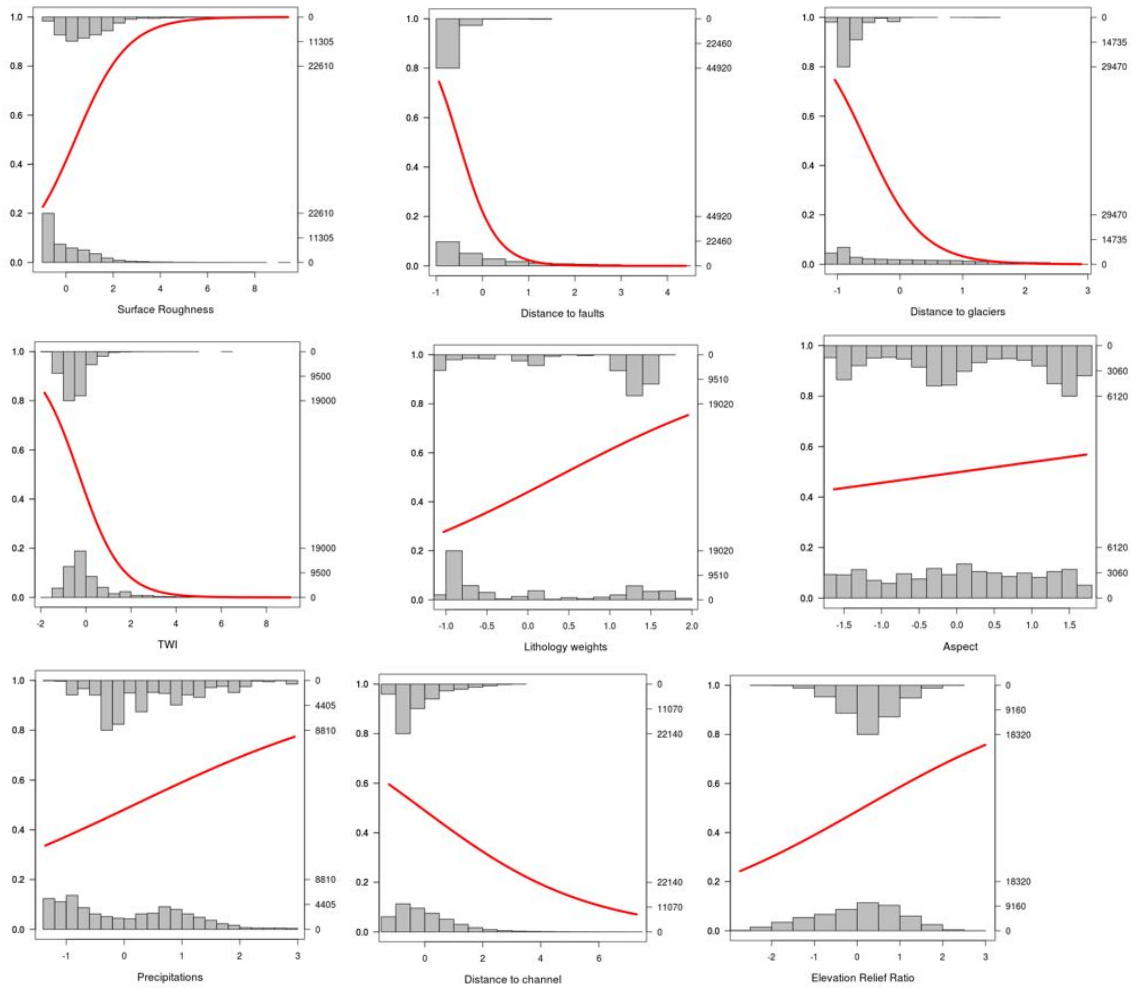


Figure 5.8: *Comparison of the binary variables and its fit with the logistic regression function*

However, the rest of the variables does not fit the sigmoid equation, and they fit better to others functions. Surface roughness, distance to faults, distance to glaciers and TWI fits better an exponential function. This behaviour can explained by the presence of landslides in a more limited range of values as is possible to observe in the upper histogram, while, the weights for the lithological units, aspect, precipitation, distance to channel and elevation relief ratio could be more associated with a linear regression function (figure 5.8).The fact that not all the variables fit in a logistic regression function opens the discussion about whether the input data are well adapted to the model. The exploration of other types of functions should be explored.

Neverthless, concerning the predictive power, the results of LR are higher than the ones obtained by WOE; however, all of the outcomes are characterized by a ROC < 0.9. The resulting landslide susceptibility map depicts the probability of landslide occurrence. Because the values are always in the range of 1 to 100, comparing between models is straightforward. One of the disadvantages of the LR approach is the overestimation of the results when an inadequate landslide catalogue is available as well as when there are more variables than observations

The Random forest approach is designed as a more robust technique that enables the production of landslide susceptibility maps with a high predictive capability following a very easy workflow. RF can handle different type of data types as well as variables with conditional dependence or multicollinearity. Quantification of the importance of the variables is given as a the main output, being a handy tool to improve the model. The resulting LS map represents the landslide density in a range of 1 to 100, being possible the comparison with other models. However, the results are limited by the number of training points, and it is denoted in the LS susceptibility map, where the areas where less or null training points are located, are overestimate.

5.5 Landslide susceptibility results

5.5.1 Selection of the best model

The best model is the combination of predictive variables that lead to the landslide susceptibility map that can identify with the higher accuracy areas where future landslide will occur.

One of the most common evaluations of the predictive capability of a model is using cross-validation methods. The ROC curve is selected as a measure of the accuracy of each model in order to compare between them and select the one with the best performance. However, a measure of the error of the model is taking in account too. The variation of the resulting outputs of a model can be used to analyse how stable and adaptive it is along the possible overfitting. Thus, each model has been implemented in an iterative way, and the main statistics of the results are computed.

The best model is selected based on the results of the iterative process ($i = 50$). In the figure 5.9 is possible to perceive the variation of the ROC with the different implementations. Based on the stability and the higher accuracy of the model 31 -2 it is selected as the best model obtained by the WOE approach.

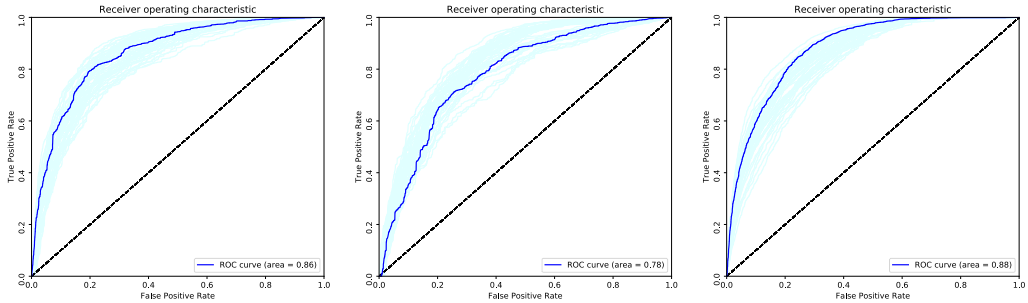


Figure 5.9: *Statistics used to select the best model resulting from the implementation of the WOE approach. Mean prediction accuracy calculated based on the implementation of the model 50 times. Left: Accuracy assessment for the model 12 -3 (11 thematic variables). Center: Accuracy assessment for the model21- 3 (7 variables). Right: Accuracy assessment for the model31 -2 (6 variables).*

A similar methodology is implemented for the five models defined for the logistic regression approach. However, the performance of the models is very similar, so, an additional statistical measure is added. The 2nd standard deviation (model error) is plotted against the mean probability to visualize the dispersion of the model error too. Model number 4 is selected as the best model obtained by the implementation of the Logistic regression.

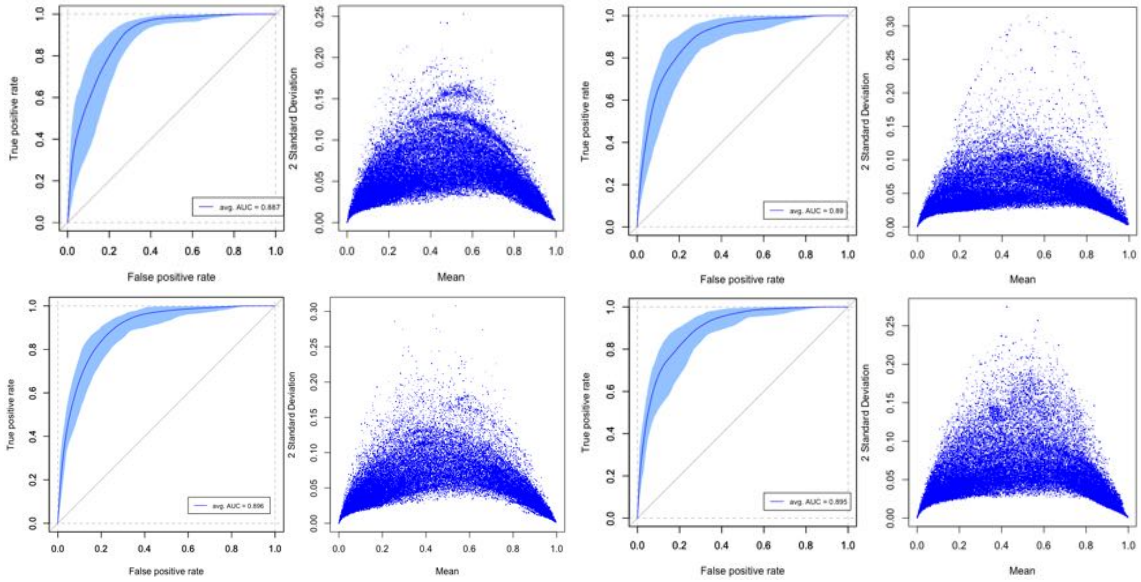


Figure 5.10: *Statistics used to select the best model resulting from the implementation of the LR approach. Prediction accuracy and error when the model is compute 100 times. Top-left: Accuracy assessment and error dispersion for the model 1 (8 variables). Top-Right: Accuracy assessment and error dispersion for the model 2 (7 variables). Bottom-left:Accuracy assessment and error dispersion for the model for the model 3 (9 variables). Bottom-right: Accuracy assessment and error dispersion for the model for the model 4 (7 variables).*

Equivalently, the models created by the random forest approach are tested using the same statistics (figure 5.11). The model 1 is selected as the best mode because of its good predictive accuracy as well as less error dispersion, leading to stability of the model in term of changing the training point. The resulting landslide susceptibility map present consistency between the distribution of the values and just a few areas can be identified as incoherent.

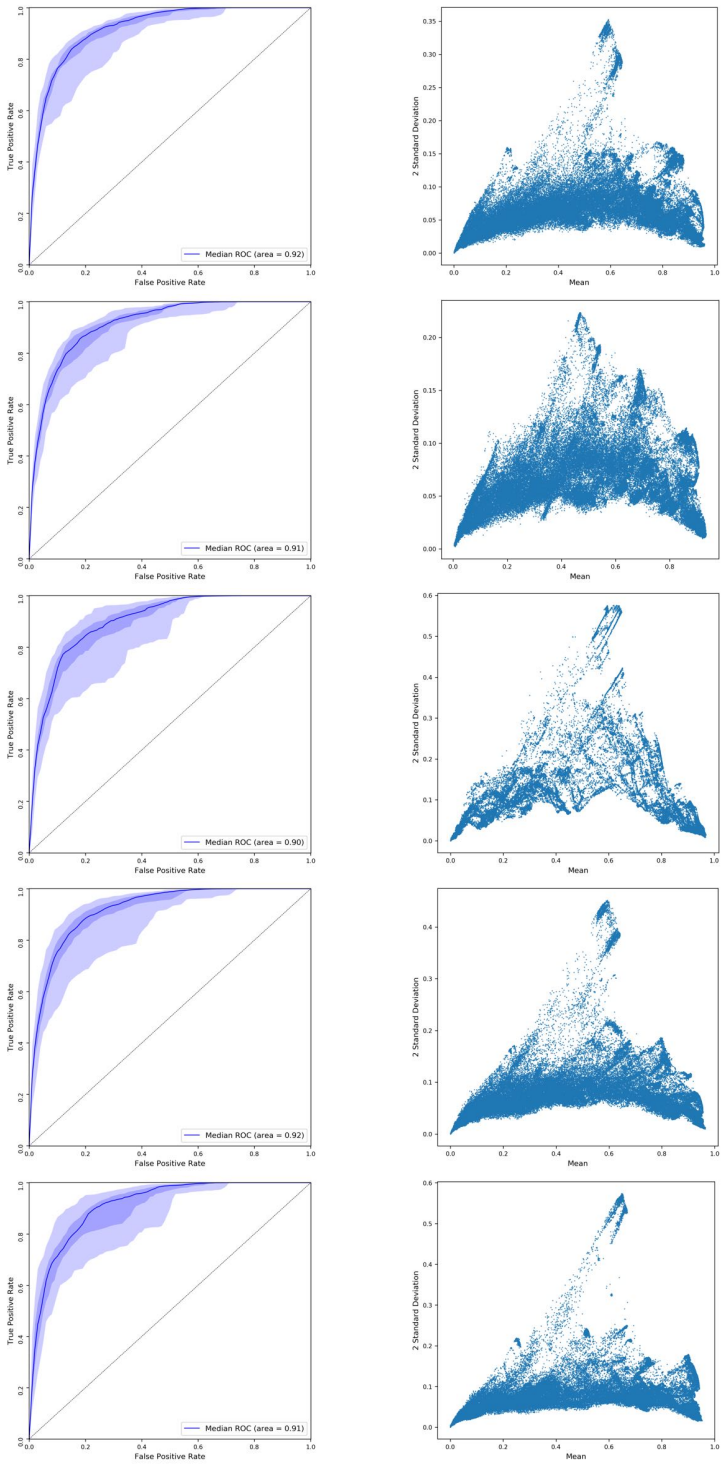


Figure 5.11: *Statistics used to select the best model resulting from the implementation of the RF approach. Prediction accuracy and Model error (OOBE) dispersion represented by the second standard deviation vs the mean when the model is compute 100 times. Top-left: Statistics for the model 1 (8 variables). Top-Right: Statistics for the model 2 (8 variables). Center-left: Statistics for the model 3 (4 variables). Center-right: Statistics for the model 4 (9 variables). Bottom: Statistics for the model 5 (7 variables)*

5.5.2 Best models differences

One of the objectives of applying different methods is to understand if the different approaches identify the same areas in order to add an extra level of confidence to the results. In order to compare the results from the different approaches, all of the outcomes are classified in 5 classes of landslide susceptibility. For the WOE result, natural breaks are used for the classification, while for LR and RF, equal intervals are implemented.

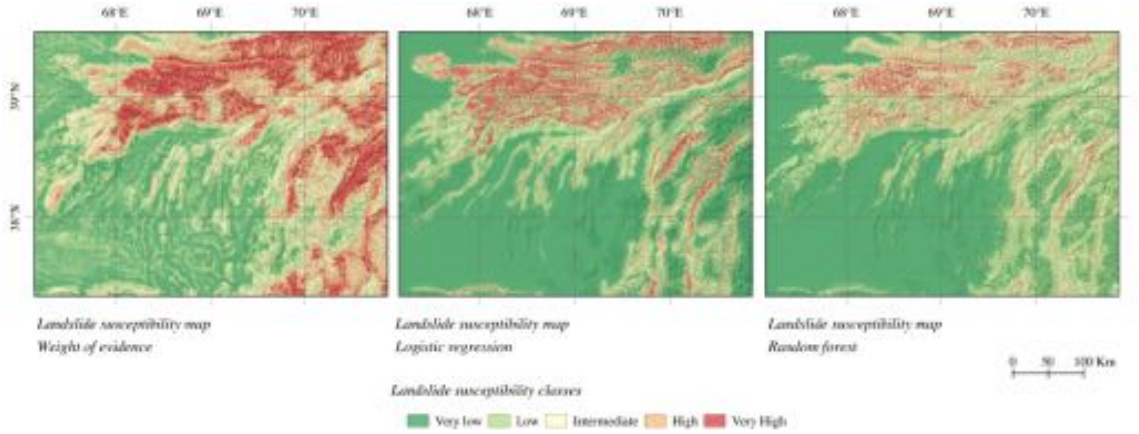


Figure 5.12: *Landslide susceptibility map for the study area computed by three different approaches.*

In general terms, it is possible to conclude that similar areas are identify as high to very high landslide susceptibility for all the approaches. The WOE results distinguish areas with a homogeneous distribution of the values in the Tien Shan and the Pamir. On the other hand, the logistic regression map enhances some areas in the Pamir as high to very high susceptibility; however, the distribution tends to be striped probably marked by the geology. The Random forest results evidence fewer areas as very high susceptibility, while an increment of intermediate to high LS zones are denoted. The results reveal the influence of the landslide catalogue in the classification that results in fewer areas associated with very high LS for the models with higher predictive accuracy.

The WOE model presents more areas classified as high landslide susceptibility, compared to the results of the LR (3 to 4- Purple). The main areas are located in Pamir. The area number 1 in figure 5.13 is probably associated with the lithological information, because a different grouping of the rocks was used for the WOE implementation and LR. The area 2 can be influenced by the characteristics of the SI used in the WOE but not in the LR. This area is characterized by very low SI values indicating the predominance of a dissected landscape. Also, very high values of LR are associated, where a substantial number of landslides are mapped.

Similarly, essential differences in the landslide susceptibility categories are found between the WOE and the RF approach. Even though the total area with robust different classification is less, in the comparison to the WOE against LR, they are consistent concerning spatial distribution. The RF and the WOE were implemented using the same lithological features; thus, a difference in the classification based on the materials is less plausible. The decreasing in the areas with strong differences could be related to the use

of similar geomorphological variables like TPI and SI (figure 5.13).

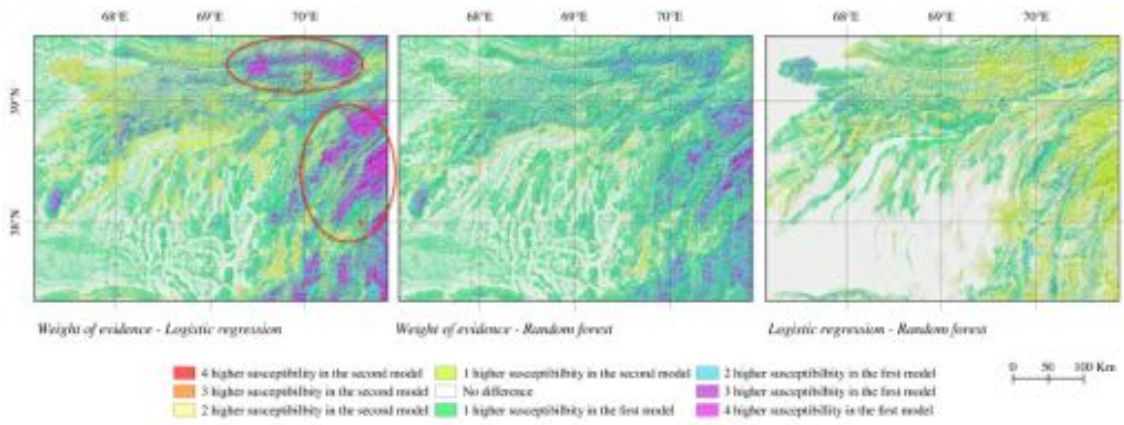


Figure 5.13: *Comparison maps between the resulting landslide susceptibility maps. Red circles are areas to be discuss in detailed.*

Contrary, sharp differences between the RF model and LR are drastically limited. However, the general landslide susceptibility for most of the Pamir area is in disagreement by the difference of 2 levels. It means that for LR the area presents a very low LS, for RF the area could low or intermediate LS. The two models have in common the use of slope and eigenValues as geomorphological predictive variables (figure 5.13).

Those differences are also represented in the abundance of the landslide susceptibility classes in the resulting maps. The figure 5.14 restate the higher number of very high LS areas delimited by the WOE models compares to the other two models. On the other hand, the LR result is characterized by a high number of very low LS. The WEO presents a homogeneous distribution of the classes. It could be related to the method implemented to select the classes. Natural breaks are used to define the landslide susceptibility for the WOE, while for the RF and the LR, five equal intervals are used.

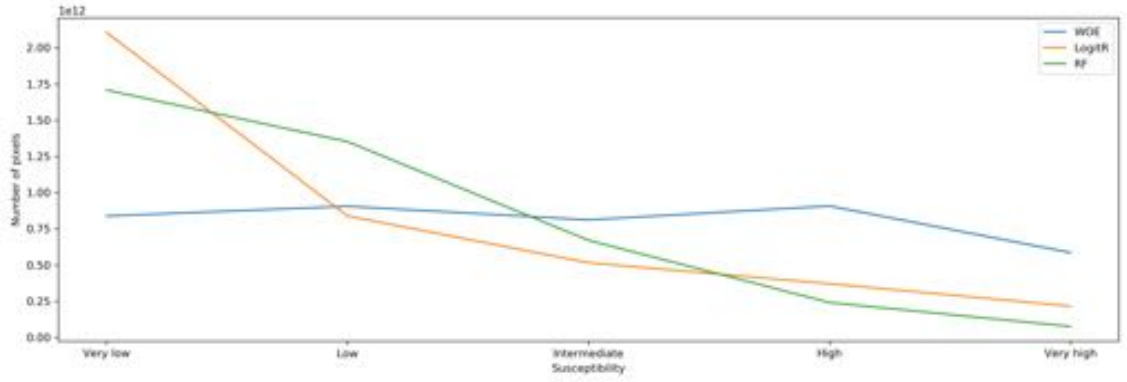
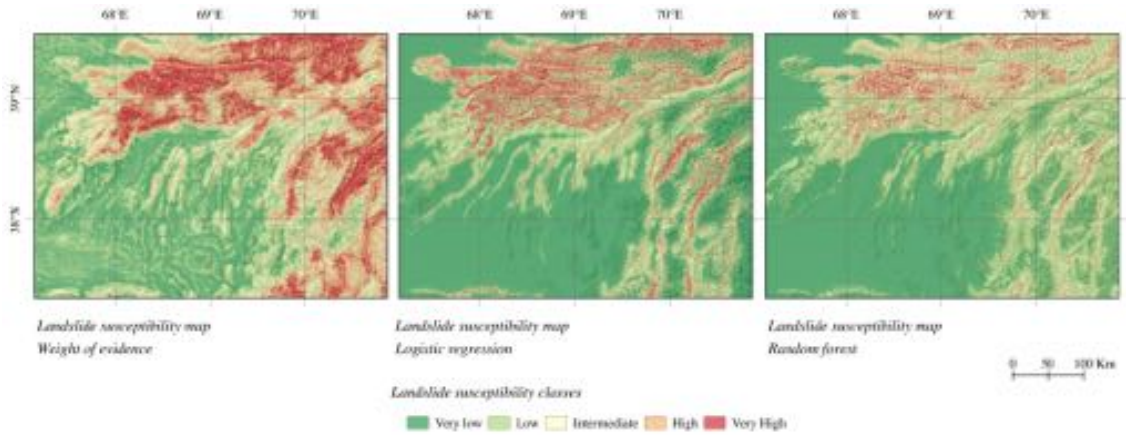


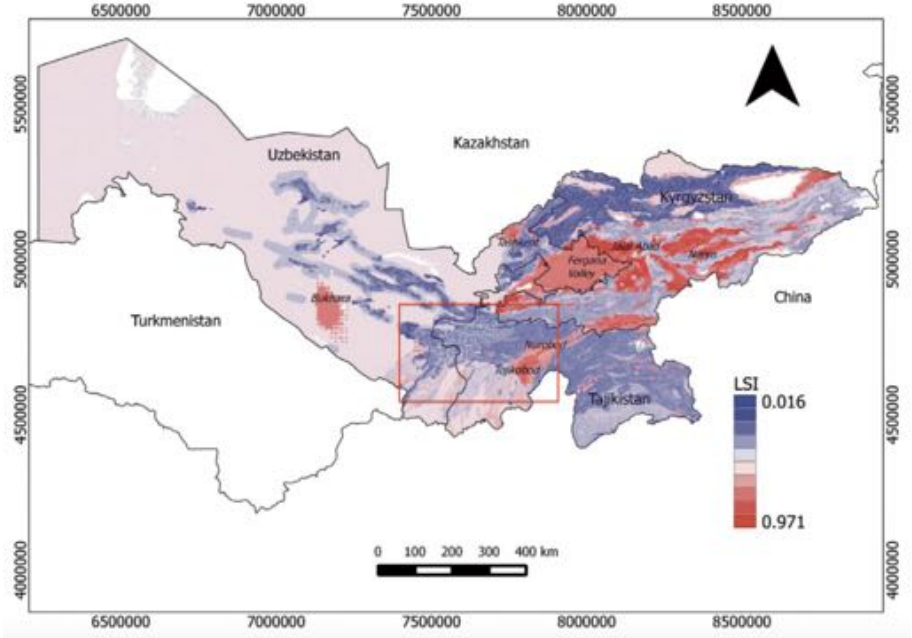
Figure 5.14: *Abundance of the landslide susceptibility class based on the model implemented.*

5.5.3 *Compatibility with previous studies*

Previous landslide susceptibility assessments have been implemented in the area using different techniques. (Saponaro *et al.* , 2015) implemented a weight of evidence model that covers all the area of study; however, the presented results are not standardized making the comparison among models difficult. On the other hand, the scales of implementations vary from this study, changing the level of detail. However, a general pattern of very high LSI values are reported by Saponaro *et al.* (2015) for the Pamir area, while intermediate values are associated with the Tadjik basin, and low LSI values are reported for the Tien Shan. The results of the WOE for this study characterized the Pamir with high LSI values; as well as the Tien Shan. Also, the Tadjik basin is defined by the lowest values of landslide occurrence for all the results obtained during this study, contrary to the information presented by (Saponaro *et al.* , 2015) .

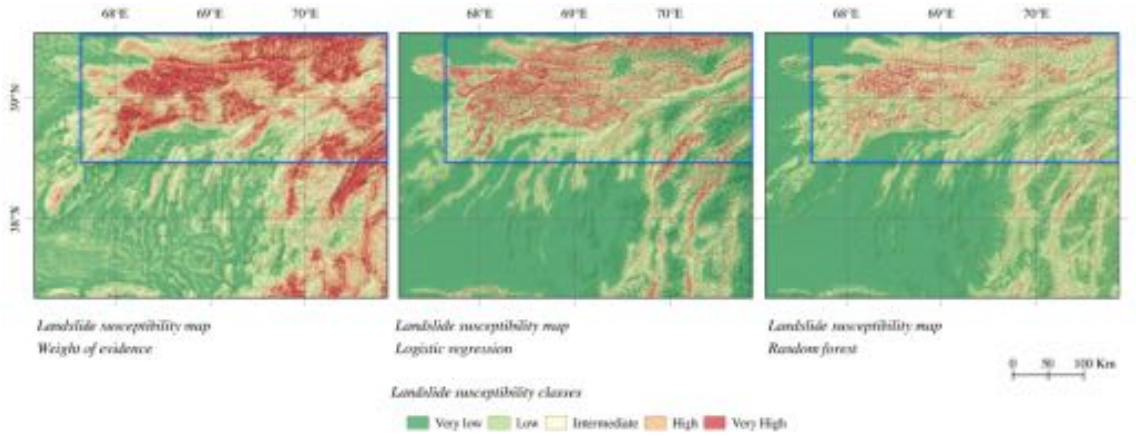


(a) *Landslide susceptibility maps for the area of study.*

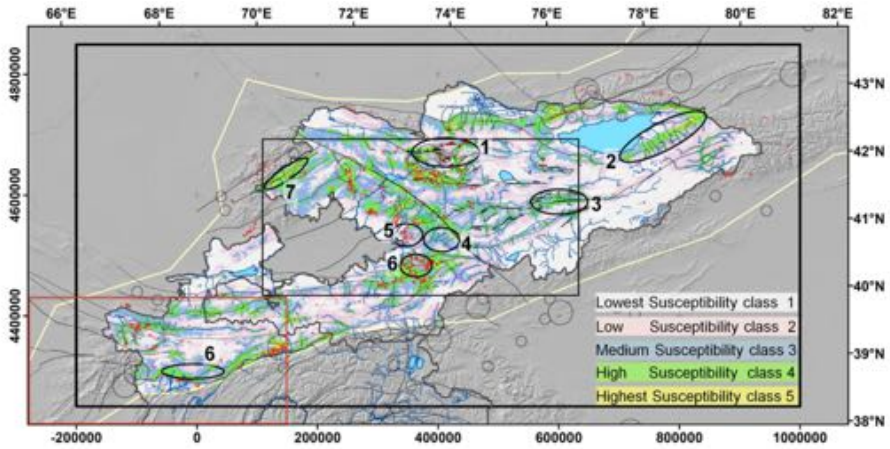


(b) *Distribution of the landslide susceptibility index. The red square indicate the area of study of this work.*

The most recent attempt to assess the landslide susceptibility in the area is the implementation of a landslide factor method for the Tien Shan. This result was obtained using the same base landslide catalogue used for this study. The area is divided into five classes, similar to the approached followed by this work. The results from (Havenith *et al.*, 2015b) are characterized by a predominance of lowest to low susceptibility class contrary to the different results obtained here. A common important landslide susceptibility spot is located in the south margin of the Sarafshon river; however, the explanation can be related more to the bias in the landslide catalogue. The landslide factor method identifies as high susceptibility the areas where a remarkable number of landslides are located. It is restricted to the data collected leading to very limited results; however, the author claims that the method is used as a partial estimation of the landslide susceptibility.



(a) *Landslide susceptibility maps for the area of study.*



(b) *Distribution of the landslide factor. The red square indicate the area of study of this work.*

Chapter 6

CONCLUSIONS

This work presents a landslide susceptibility assessment for the South Tien Shan and Western Pamir in Central Asia. A landslide catalogue was created by the compilation of previous information and the completion by manual delimitation of mass wasting, mainly for the Pamir area. The catalogue has coverage of 0.02% and introduces different biases in the model. First, the area of Tien Shan has better representation than the Pamir area not only regarding the variety of landslides sizes but also in the distribution. A substantial landslide density is identified in the North-West area of Tien Shan, in the valley of the Zarafshon. Also, the areas around Dushanbe, the capital of Tajikistan are characterized by an important landslide density. Intermediate to small size landslides are located along the Vakhsh thrust system (MPT); while small landslides characterized the Darvaz fault system surroundings.

Thematic variables grouped in 5 clusters were created. Geological information is harmonized to obtain a lithological regional map of the area. Precipitation information is re-sampling. Satellite images are used to calculate the NDVI values. Euclidean distances to Glacier and Fault are computed, and DEM derivatives are created. A large number of geomorphological parameters derived from DEM are generated in order to complement the knowledge of the study area. Some of those indices have never been implemented for landslide susceptibility before. The elevation above the channel is an approach to understand the influence of the gradient and the potential energy. The topographic position index separates between ridges and valley bottoms; similarly, the eigenvalues depict the curvature of the valleys and ridges. The surface roughness highlight areas with erosional processes, while the local relief is related to the river incision. The elevation relief ratio reflects the conditions of stability of the landscape while the surface index discriminates between erosional and steady-state landscapes.

Three statically based approaches are implemented to calculate landslide susceptibility. The WOE shows good performance; however, the methodology is time-consuming. LR is presented as a faster option to compute landslide susceptibility; however, a more in-depth discussion about the fitting of the model to the thematic variables is required. The RF approach is a more robust method with which good landslide susceptibility maps are obtained.

The selection of the more relevant variables to obtain the optimal landslide susceptibility assessment are chosen based on a sequential process. The lithological information constrains the areas of very high landslides susceptibility because of a significant number

of landslides associated with the sedimentary marine sequences deposited in the Jurassic and the Devonian mostly. The distance to fault is a factor that is used in all the models because there is a positive spatial association between the closeness to an active fault and the landslide occurrence. Geomorphological parameters like EigenValues, TPI, SI, local relieve, and elevation above the channel proved to be useful proxies to map geomorphological characteristics related to the slope instabilities. However, special care is needed when selecting the variables in order to avoid multicollinearity.

The influence of the glacial processes in the landslide occurrence must be study in detail for the area. During this study a positive spacial association was found; however, the method used to represent the variable creates incoherences in the landslide susceptibility resulting map. Contrary, aspect, NDVI, and TWI are identified as variables with minimal influence in the landslide susceptibility modelling. It is because of the lack of a predominant pattern related to the landslide occurrence. For the aspect, a similar number of landslides are related to the different orientation. On the other hand, grasslands characterized areas affected by mass wasting along areas that not. Similarly, the values of TWI are homogeneously distributed for areas with the presence or absence of landslides.

Bibliography

Ahnert, Frank. 1984. Local relief and the height limits of mountain ranges. *American Journal of Science*, **284**(9), 1035–1055.

Aizen, Elena M, Aizen, Vladimir B, Melack, John M, Nakamura, Tsutomu, & Ohta, Takeshi. 2001. Precipitation and atmospheric circulation patterns at mid-latitudes of Asia. *International Journal of Climatology*, **21**(5), 535–556.

Akgun, Aykut. 2012. A comparison of landslide susceptibility maps produced by logistic regression, multi-criteria decision, and likelihood ratio methods: a case study at İzmir, Turkey. *Landslides*, **9**(1), 93–106.

Akgun, Aykut, Dag, Serhat, & Bulut, Fikri. 2008. Landslide susceptibility mapping for a landslide-prone area (Findikli, NE of Turkey) by likelihood-frequency ratio and weighted linear combination models. *Environmental Geology*, **54**(6), 1127–1143.

Andreani, Louis, Stanek, Klaus P, Gloaguen, Richard, Krentz, Ottomar, & Domínguez-González, Leomaris. 2014. DEM-based analysis of interactions between tectonics and landscapes in the Ore Mountains and Eger Rift (East Germany and NW Czech Republic). *Remote Sensing*, **6**(9), 7971–8001.

Andreani, Louis, Pohl, Erik, Shahzad, F, Koucha, L, & Gloaguen, Richard. 2018. TecGEMS: a python-based toolbox for tectonic geomorphology. *Under development*.

Armaş, Iuliana. 2012. Weights of evidence method for landslide susceptibility mapping. Prahova Subcarpathians, Romania. *Natural Hazards*, **60**(3), 937–950.

Arrowsmith, J R, & Strecker, Manfred R. 1999. Seismotectonic range-front segmentation and mountain-belt growth in the Pamir-Alai region, Kyrgyzstan (India-Eurasia collision zone). *Geological Society of America Bulletin*, **111**(11), 1665–1683.

Baratov, RB. 1966. *Intruzivnyye kompleksy yuzhnogo sklona Gissarskogo khrebeta i svyazannyye s nimi orudeneniya*.

Barredo, JoséI, Benavides, Annetty, Hervás, Javier, & van Westen, Cees J. 2000. Comparing heuristic landslide hazard assessment techniques using GIS in the Tirajana basin, Gran Canaria Island, Spain. *International journal of applied earth observation and geoinformation*, **2**(1), 9–23.

Bindi, D, Abdrakhmatov, K, Parolai, S, Mucciarelli, M, Grünthal, G, Ischuk, A, Mikhailova, N, & Zschau, J. 2012. Seismic hazard assessment in Central Asia: Outcomes from a site approach. *Soil dynamics and earthquake engineering*, **37**, 84–91.

Bonham-Carter, Graeme F. 1994. Geographic information systems for geoscientists-modeling with GIS. *Computer methods in the geoscientists*, **13**, 398.

Boroumandi, Mehdi, Khamehchiyan, Mashalah, & Nikoudel, Mohammad Reza. 2015. Using of analytic hierarchy process for landslide hazard zonation in Zanjan Province, Iran. *Pages 951–955 of: Engineering Geology for Society and Territory-Volume 2*. Springer.

Brabb, Earl E. 1985. Innovative approaches to landslide hazard and risk mapping. *Pages 17–22 of: International Landslide Symposium Proceedings, Toronto, Canada*, vol. 1.

Breiman, Leo. 2001. Random forests. *Machine learning*, **45**(1), 5–32.

Brookfield, ME. 2000. Geological development and Phanerozoic crustal accretion in the western segment of the southern Tien Shan (Kyrgyzstan, Uzbekistan and Tajikistan). *Tectonophysics*, **328**(1-2), 1–14.

Burtman, Valentin Semenovich, & Molnar, Peter Hale. 1993. *Geological and geophysical evidence for deep subduction of continental crust beneath the Pamir*. Vol. 281. Geological Society of America.

Burtman, VS. 2000. Cenozoic crustal shortening between the Pamir and Tien Shan and a reconstruction of the Pamir–Tien Shan transition zone for the Cretaceous and Palaeogene. *Tectonophysics*, **319**(2), 69–92.

Carranza, Emmanuel John M, & Sadeghi, Martiya. 2010. Predictive mapping of prospectivity and quantitative estimation of undiscovered VMS deposits in Skellefte district (Sweden). *Ore Geology Reviews*, **38**(3), 219–241.

Carrara, A, Cardinali, M, Detti, R, Guzzetti, F, Pasqui, V, & Reichenbach, P. 1991. GIS techniques and statistical models in evaluating landslide hazard. *Earth surface processes and landforms*, **16**(5), 427–445.

Chung, Chang-Jo F, Fabbri, Andrea G, *et al.* . 1999. Probabilistic prediction models for landslide hazard mapping. *Photogrammetric engineering and remote sensing*, **65**(12), 1389–1399.

Dai, FC, & Lee, CF. 2001. Frequency–volume relation and prediction of rainfall-induced landslides. *Engineering geology*, **59**(3-4), 253–266.

De Reu, Jeroen, Bourgeois, Jean, Bats, Machteld, Zwertvaegher, Ann, Gelorini, Vanessa, De Smedt, Philippe, Chu, Wei, Antrop, Marc, De Maeyer, Philippe, Finke, Peter, *et al.* . 2013. Application of the topographic position index to heterogeneous landscapes. *Geomorphology*, **186**, 39–49.

Evans, Stephen G, Roberts, Nicholas J, Ischuk, Anatoli, Delaney, Keith B, Morozova, Galina S, & Tutubalina, Olga. 2009. Landslides triggered by the 1949 Khait earthquake, Tajikistan, and associated loss of life. *Engineering Geology*, **109**(3-4), 195–212.

Federal State Budgetary Institution A.P. Karpinsky Russian Geological Research Institute (FGUP VSEGEI). 2018. *Cartographic resources on regional geology*. data retrieved from `http://webmapget.vsegei.ru/index.html`.

Frangi, Alejandro F, Niessen, Wiro J, Vincken, Koen L, & Viergever, Max A. 1998. Multiscale vessel enhancement filtering. *Pages 130–137 of: International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer.

Froude, Melanie J, & Petley, David N. 2018. Global fatal landslide occurrence from 2004 to 2016. *Natural Hazards and Earth System Sciences*, **18**(8), 2161–2181.

Fuchs, MC, Gloaguen, R, Merchel, S, Pohl, E, Sulaymonova, VA, Andermann, C, & Rugel, G. 2015. Millennial erosion rates across the Pamir based on 10 Be concentrations in fluvial sediments: dominance of topographic over climatic factors. *Earth Surf Dyn Discuss*, **3**, 83–128.

German Aerospace Center (DLR), Italian Space Agency (ASI). 2000. *Shuttle Radar Topography Mission*. data retrieved from, https://www.dlr.de/eoc/en/desktopdefault.aspx/tabid-5515/9214_read-17716/.

Grohmann, CH, Smith, MJ, & Riccomini, C. 2009. Surface roughness of topography: a multi-scale analysis of landform elements in Midland Valley, Scotland. *Proceedings of geomorphometry*, **31**, 140–148.

Gruber, FE, & Mergili, M. 2013. Regional-scale analysis of high-mountain multi-hazard and risk indicators in the Pamir (Tajikistan) with GRASS GIS. *Natural Hazards and Earth System Sciences*, **13**(11), 2779–2796.

Guha-Sapir, Debarati, Hoyois, Philippe, Wallemacq, Pasacline, & Below, Regina. 2017. *Annual Disaster Statistical Review 2016*. Tech. rept. Center for research on the epidemiology of disasters.

Guo, Wanqin, Liu, Shiyin, Xu, Junli, Wu, Lizong, Shangguan, Donghui, Yao, Xiaojun, Wei, Junfeng, Bao, Weijia, Yu, Pengchun, Liu, Qiao, *et al.* . 2015. The second Chinese glacier inventory: data, methods and results. *Journal of Glaciology*, **61**(226), 357–372.

Guzzetti, F, Galli, M, Reichenbach, P, Ardizzone, F, & Cardinali, M. 2006. Landslide hazard assessment in the Collazzone area, Umbria, Central Italy. *Natural Hazards and Earth System Science*, **6**(1), 115–131.

Guzzetti, Fausto, Mondini, Alessandro Cesare, Cardinali, Mauro, Fiorucci, Federica, Santangelo, Michele, & Chang, Kang-Tsung. 2012. Landslide inventory maps: New tools for an old problem. *Earth-Science Reviews*, **112**(1-2), 42–66.

Havenith, H-B, Strom, A, Jongmans, D, Abdrakhmatov, A, Delvaux, D, & Tréfois, P. 2003. Seismic triggering of landslides, Part A: Field evidence from the Northern Tien Shan. *Natural Hazards and Earth System Science*, **3**(1/2), 135–149.

Havenith, Hans-Balder, & Bourdeau, Céline. 2010. Earthquake-induced hazards in mountain regions: a review of case histories from Central Asia–an inaugural lecture to the society. *Geologica Belgica*, **13**, 135–150.

Havenith, Hans-Balder, Abdrakhmatov, Kanatbek, Torgoev, Isakbek, Ischuk, Anatoly, Strom, Alexander, Bystrickỳ, Erik, & Cipciar, Andrej. 2013. Earthquakes, landslides, dams and reservoirs in the Tien Shan, central Asia. *Pages 27–31 of: Landslide Science and Practice*. Springer.

Havenith, Hans-Balder, Strom, Alexander, Torgoev, Isakbek, Torgoev, Almazbek, Lamair, Laura, Ischuk, Anatoly, & Abdrakhmatov, Kanatbek. 2015a. Tien Shan geohazards database: Earthquakes and landslides. *Geomorphology*, **249**, 16–31.

Havenith, Hans-Balder, Torgoev, Almazbek, Schlögel, Romy, Braun, Anika, Torgoev, Isakbek, & Ischuk, Anatoly. 2015b. Tien Shan geohazards database: Landslide susceptibility analysis. *Geomorphology*, **249**, 32–43.

Highland, Lynn, Bobrowsky, Peter T, *et al.* . 2008. *The landslide handbook: a guide to understanding landslides*. US Geological Survey Reston.

Huggett, Richard. 2016. *Fundamentals of geomorphology*. Routledge.

Hungr, Oldrich, Leroueil, Serge, & Picarelli, Luciano. 2014. The Varnes classification of landslide types, an update. *Landslides*, **11**(2), 167–194.

Ischuk, Anatoli, Bendick, Rebecca, Rybin, Anatoly, Molnar, Peter, Khan, Shah Faisal, Kuzikov, Sergey, Mohadjer, Solmaz, Saydullaev, Umed, Ilyasova, Zhyra, Schelochkov, Gennady, *et al.* . 2013. Kinematics of the Pamir and Hindu Kush regions from GPS geodesy. *Journal of geophysical research: solid earth*, **118**(5), 2408–2416.

Ishihara, Kenji, Okusa, Shigeyasu, Oyagi, Norio, & Ischuk, Anatoliy. 1990. Liquefaction-induced flow slide in the collapsible loess deposit in Soviet Tajik. *Soils and foundations*, **30**(4), 73–89.

Käßner, Alexandra, Ratschbacher, Lothar, Jonckheere, Raymond, Enkelmann, Eva, Khan, Jahanzeb, Sonntag, Benita-Lisette, Gloaguen, Richard, Gadoev, Mustafo, & Oimahmadov, Ilhomjon. 2016. Cenozoic intracontinental deformation and exhumation at the northwestern tip of the India-Asia collision—southwestern Tian Shan, Tajikistan, and Kyrgyzstan. *Tectonics*, **35**(9), 2171–2194.

Kayastha, Prabin, Dhital, Megh Raj, & De Smedt, Florimond. 2013. Application of the analytical hierarchy process (AHP) for landslide susceptibility mapping: a case study from the Tinau watershed, west Nepal. *Computers & Geosciences*, **52**, 398–408.

Khasanov, A Kh. 1975. Overall sequence and time of formation of igneous rock and ore-metasomatite associations in the Gissar- Alay region, southern Tien Shan. *Pages 91 – 94 of: Dokl. Akad. Nauk SSSR*, vol. 223.

Kirschbaum, Dalia, Stanley, Thomas, & Zhou, Yaping. 2015. Spatial and temporal analysis of a global landslide catalog. *Geomorphology*, **249**, 4–15.

Kirschbaum, Dalia Bach, Adler, Robert, Hong, Yang, Hill, Stephanie, & Lerner-Lam, Arthur. 2010. A global landslide catalog for hazard applications: method, results, and limitations. *Natural Hazards*, **52**(3), 561–575.

Koehrsen, William. 2007. *Random forest simple explanation*.

Korup, Oliver, & Tweed, Fiona. 2007. Ice, moraine, and landslide dams in mountainous terrain. *Quaternary Science Reviews*, **26**(25-28), 3406–3422.

Kriegel, David, Mayer, Christoph, Hagg, Wilfried, Vorogushyn, Sergiy, Duethmann, Doris, Gafurov, Abror, & Farinotti, Daniel. 2013. Changes in glacierisation, climate and runoff in the second half of the 20th century in the Naryn basin, Central Asia. *Global and planetary change*, **110**, 51–61.

Kurenkov, SA, & Aristov, VA. 1995. On the time of formation of the Turkestan paleoocean crust. *Geotectonics*, **29**(6), 469–477.

Kutuzov, Stanislav, & Shahgedanova, Maria. 2009. Glacier retreat and climatic variability in the eastern Terskey–Alatoo, inner Tien Shan between the middle of the 19th century and beginning of the 21st century. *Global and Planetary Change*, **69**(1-2), 59–70.

Lee, Jung-Hyun, Sameen, Maher Ibrahim, Pradhan, Biswajeet, & Park, Hyuck-Jin. 2018. Modeling landslide susceptibility in data-scarce environments using optimized data mining and statistical methods. *Geomorphology*, **303**, 284–298.

Lee, S. 2005. Application of logistic regression model and its validation for landslide susceptibility mapping using GIS and remote sensing data. *International Journal of Remote Sensing*, **26**(7), 1477–1491.

Lee, Saro, Choi, Jaewon, & Min, Kyungduck. 2002. Landslide susceptibility analysis and verification using the Bayesian probability model. *Environmental Geology*, **43**(1-2), 120–131.

Leith, William, & Simpson, David W. 1986. Seismic domains within the Gissar-Kokshal seismic zone, Soviet Central Asia. *Journal of Geophysical Research: Solid Earth*, **91**(B1), 689–699.

Liu, Dongliang, Li, Haibing, Sun, Zhiming, Cao, Yong, Wang, Leizhen, Pan, Jiawei, Han, Liang, & Ye, Xiaozhou. 2017. Cenozoic episodic uplift and kinematic evolution between the Pamir and Southwestern Tien Shan. *Tectonophysics*, **712**, 438–454.

Marchesini, I, Ardizzone, F, Alvioli, M, Rossi, M, & Guzzetti, F. 2014. Non-susceptible landslide areas in Italy and in the Mediterranean region. *Natural Hazards and Earth System Sciences*, **14**(8), 2215–2231.

Maussion, Fabien, Scherer, Dieter, Mölg, Thomas, Collier, Emily, Curio, Julia, & Finkelnburg, Roman. 2014. Precipitation seasonality and variability over the Tibetan Plateau as resolved by the High Asia Reanalysis. *Journal of Climate*, **27**(5), 1910–1927.

Mergili, M, & Schneider, JF. 2011. Regional-scale analysis of lake outburst hazards in the southwestern Pamir, Tajikistan, based on remote sensing and GIS. *Natural Hazards and Earth System Sciences*, **11**(5), 1447–1462.

Mohadjer, S, Bendick, R, Ischuk, A, Kuzikov, S, Kostuk, A, Saydullaev, U, Lodi, S, Kakar, DM, Wasy, A, Khan, MA, *et al.* . 2010. Partitioning of India-Eurasia convergence in the Pamir-Hindu Kush from GPS measurements. *Geophysical Research Letters*, **37**(4).

Nadim, Farrokh, Kjekstad, Oddvar, Peduzzi, Pascal, Herold, Christian, & Jaedicke, Christian. 2006. Global landslide and avalanche hotspots. *Landslides*, **3**(2), 159–173.

Natalya Mikhailova, N.N Poleshko, I.L Aristova A.S Mukambayev G.O Kulikova. 2015. *EMCA Central Asia Earthquake catalogue v1.0*. data retrieved from GFZ Data Services, http://doi.org/10.5880/GFZ.EWS.2015.001.

National Aeronautics and Space Administration (NASA), National Geospatial intelligence agency (NGA). 2000. *Shuttle Radar Topography Mission*. data retrieved from, https://lta.cr.usgs.gov/SRTM1Arc.

Negredo, Ana M, Replumaz, Anne, Villaseñor, Antonio, & Guillot, Stéphane. 2007. Modeling the evolution of continental subduction processes in the Pamir–Hindu Kush region. *Earth and Planetary Science Letters*, **259**(1-2), 212–225.

Nuimura, T, Sakai, A, Taniguchi, K, Nagai, H, Lamsal, D, Tsutaki, S, Kozawa, A, Hoshina, Y, Takenaka, S, Omiya, S, *et al.* . 2015. The gamdam glacier inventory: a quality-controlled inventory of Asian glaciers. *Cryosphere*, **9**(3).

Paudel, Uttam, Oguchi, Takashi, & Hayakawa, Yuichi. 2016. Multi-resolution landslide susceptibility analysis using a DEM and random forest. *International Journal of Geosciences*, **7**(05), 726.

Pike, Richard J, & Wilson, Stephen E. 1971. Elevation-relief ratio, hypsometric integral, and geomorphic area-altitude analysis. *Geological Society of America Bulletin*, **82**(4), 1079–1084.

Pohl, Eric, Gloaguen, Richard, & Seiler, Ralf. 2015. Remote sensing-based assessment of the variability of winter and summer precipitation in the Pamirs and their effects on hydrology and hazards using harmonic time series analysis. *Remote Sensing*, **7**(8), 9727–9752.

Pourghasemi, Hamid Reza, Pradhan, Biswajeet, & Gokceoglu, Candan. 2012. Application of fuzzy logic and analytical hierarchy process (AHP) to landslide susceptibility mapping at Haraz watershed, Iran. *Natural hazards*, **63**(2), 965–996.

Rabus, Bernhard, Eineder, Michael, Roth, Achim, & Bamler, Richard. 2003. The shuttle radar topography mission—a new class of digital elevation models acquired by spaceborne radar. *ISPRS journal of photogrammetry and remote sensing*, **57**(4), 241–262.

Raja, Nussaïbah B, Çiçek, Ihsan, Türkoğlu, Necla, Aydin, Olgu, & Kawasaki, Akiyuki. 2017. Landslide susceptibility mapping of the Sera River Basin using logistic regression model. *Natural Hazards*, **85**(3), 1323–1346.

Raup, Bruce H, Kieffer, Hugh H, Hare, Trent M, & Kargel, Jeffrey S. 2000. Generation of data acquisition requests for the ASTER satellite instrument for monitoring a globally distributed target: Glaciers. *IEEE Transactions on Geoscience and Remote Sensing*, **38**(2), 1105–1112.

Rautian, Tatyana, & Leith, William. 2002. Composite regional catalogs of earthquakes in the former Soviet Union. *US Geological Survey Open File Report*, **2**, 500.

Regmi, Netra R, Giardino, John R, & Vitek, John D. 2010. Modeling susceptibility to landslides using the weight of evidence approach: Western Colorado,USA. *Geomorphology*, **115**(1-2), 172–187.

Reichenbach, Paola, Rossi, Mauro, Malamud, Bruce, Mihir, Monika, & Guzzetti, Fausto. 2018. A review of statistically-based landslide susceptibility models. *Earth-Science Reviews*.

Rogozhin, Ye A. 1993. Southern Tien Shan folding. *Geotectonics*, **27**, 51–61.

Ruleman, CA, Crone, AJ, Machette, MN, Haller, KM, & Rukstales, KS. 2007. *Map and database of probable and possible Quaternary faults in Afghanistan*. Tech. rept. Geological Survey (US).

Saito, Hitoshi, Nakayama, Daichi, & Matsuyama, Hiroshi. 2010. Relationship between the initiation of a shallow landslide and rainfall intensity—duration thresholds in Japan. *Geomorphology*, **118**(1-2), 167–175.

Saponaro, Annamaria, Pilz, Marco, Bindi, Dino, & Parolai, Stefano. 2015. The contribution of EMCA to landslide susceptibility mapping in Central Asia. *Annals of Geophysics*, **58**(1).

Schenider, JF, Gruber, FE, & Mergili, M. 2013. Impact of large landslides, mitigation measures. *Ital J Eng Geol Environ*, **6**, 73–84.

Schumm, Stanley A. 1956. Evolution of drainage systems and slopes in badlands at Perth Amboy, New Jersey. *Geological society of America bulletin*, **67**(5), 597–646.

Schurr, Bernd, Ratschbacher, Lothar, Sippl, Christian, Gloaguen, Richard, Yuan, Xiaohui, & Mechie, James. 2014. Seismotectonics of the Pamir. *Tectonics*, **33**(8), 1501–1518.

Shahid Ullah, Kanat Abdrakhmatov, Alla Sadykova Roman Ibragimov Anatoly Ishuk Danciu Laurentiu Stefano Parolai Dino Bindi Marc Wieland Massimiliano Pittore. 2015. *EMCA Central Asia seismic source model v1.0*. data retrieved from GFZ Data Services, http://doi.org/10.5880/GFZ.EWS.2015.002.

Shahzad, Faisal, & Gloaguen, Richard. 2011. TecDEM: A MATLAB based toolbox for tectonic geomorphology, Part 2: Surface dynamics and basin analysis. *Computers & geosciences*, **37**(2), 261–271.

Shi, Yafeng, Liu, Chaohai, & Kang, Ersi. 2009. The glacier inventory of China. *Annals of Glaciology*, **50**(53), 1–4.

Smith, Mark W. 2014. Roughness in the earth sciences. *Earth-Science Reviews*, **136**, 202–225.

Strahler, Arthur N. 1952. Hypsometric (area-altitude) analysis of erosional topography. *Geological Society of America Bulletin*, **63**(11), 1117–1142.

Strom, Alexander. 2010. Landslide dams in Central Asia region. *Journal of the Japan Landslide Society*, **47**(6), 309–324.

T Bolch, N Mölg, H Frey F Paul. *DynRG-Tip, Aksu-Tarim-RS and Glaciers_cci* . Projects executed by Technical University of Dresden, Germay and University of Zürich, Switzerland.

Tarbuck, Edward J., Lutgens, Frederick K, *et al.* . 2014. *Earth: An Introduction to Physical Geology*. Pearson.

Thurman, Michael. 2011a. *Natural Disaster Risks in Central Asia: A Synthesis*. Tech. rept. UNDP/BCPR, Regional Disaster Risk Reduction Advisor, Europe and CIS.

Thurman, Michael. 2011b. *Natural Disaster Risks in Central Asia: A Synthesis*.

Trentin, Romario, & de Souza Robaina, Luís Eduardo. 2018. STUDY OF THE LANDFORMS OF THE IBICUÍ RIVER BASIN WITH USE OF TOPOGRAPHIC POSITION INDEX. *Revista Brasileira de Geomorfologia*, **19**(2).

Trigila, Alessandro, Iadanza, Carla, Esposito, Carlo, & Scarascia-Mugnozza, Gabriele. 2015. Comparison of logistic regression and random forests techniques for shallow landslide susceptibility assessment in Giampilieri (NE Sicily, Italy). *Geomorphology*, **249**, 119–136.

Tseng, CM, Lin, CW, & Hsieh, WD. 2015. Landslide susceptibility analysis by means of event-based multi-temporal landslide inventories. *Natural Hazards & Earth System Sciences Discussions*, **3**(2).

Ulomov, Valentin I, Group, GSHAP Region 7 Working, *et al.* . 1999. Seismic hazard of northern Eurasia. *Annals of Geophysics*, **42**(6).

Ulomov, V.I. 1999. *Global Seismic Hazard Assessment Program*. Tech. rept. International Lithosphere program/ International Council of Scientific Unions (ICSU).

U.S. Geological Survey. 2015. *Landsat 8 OLI/TIRS Level-2 Data Products - Surface Reflectance*. https://lta.cr.usgs.gov/L8Level2SR.

Varnes, David J. 1958. Landslide types and processes. *Landslides and engineering practice*, **29**(3), 20–45.

Wieczorek, GERALD F, & Guzzetti, F. 1999. A review of rainfall thresholds for triggering landslides. *Pages 407–414 of: Proc. of the EGS Plinius Conference, Maratea, Italy*.

Yalcin, A, Reis, S, Aydinoglu, AC, & Yomralioglu, T. 2011. A GIS-based comparative study of frequency ratio, analytical hierarchy process, bivariate statistics and logistics regression methods for landslide susceptibility mapping in Trabzon, NE Turkey. *Catena*, **85**(3), 274–287.