FACULTY OF SCIENCES
DEPARTMENT OF GEOGRAPHY

LANDSLIDE SUSCEPTIBILITY PREDICTION IN A MOUNTAINOUS CATCHMENT: THE NARANJO BASIN, WESTERN GUATEMALA

Nick Kenner ESTRADA OROZCO

DISSERTATION PRESENTED
IN ACCORDANCE WITH THE REQUIREMENTS
FOR THE DEGREE OF DOCTOR OF SCIENCES

JULY 2017

FACULTY OF SCIENCES
DEPARTMENT OF GEOGRAPHY

LANDSLIDE SUSCEPTIBILITY PREDICTION IN A MOUNTAINOUS CATCHMENT: THE NARANJO BASIN, WESTERN GUATEMALA

Nick Kenner ESTRADA OROZCO

**DISSERTATION PRESENTED
IN ACCORDANCE WITH THE REQUIREMENTS
FOR THE DEGREE OF DOCTOR OF SCIENCES**

JULY 2017

Photos on the cover © Estrada 2017 – Overview of Guatemalan landslide issues

Members of the Jury

Prof. Dr. Alain DEMOULIN – University of Liège – Co-Advisor, Secretary

Dr. Olivier DEWITTE – Royal Museum for Central Africa

Prof. Dr.Matthieu KERVYN DE MEERENDRE – Vrije Universiteit Brussel

Prof. Dr. Raúl Francisco PINEDA LÓPEZ – Universidad Autónoma de Querétaro

Prof. Dr. Serge SCHMITZ – University of Liège – Advisor

Prof. Dr. Matthias VANMAERCKE – University of Liège – Chairman

Acknowledgements

First and foremost, I want to thank my adviser, Prof. Dr. Serge Schmitz, for his support, patience, encouragement, and kindness especially during the tough times of my Ph.D. study. I thank him for believing in me. Above all, I am deeply grateful to him for offering me the opportunity to pursue my doctoral studies in Liège.

I would like to express my sincere gratitude to my co-advisor, Prof. Dr. Alain Demoulin, for his constant and productive guidance throughout this research by commenting on my writings and offering his suggestions.

I would also like to acknowledge Dr. Olivier Dewitte for sharing his time and impressive expertise. His papers were of great assistance.

My special thanks also go to Prof. Dr. Yves Cornet for the critical comments that helped me construct my database.

I gratefully acknowledge the funding source, the CUD-PIC Project and the Laboratory for the analysis of places, landscapes and European countryside (LAPLEC) for allowing me to develop my research and make my Ph.D. work possible.

I am especially grateful to the Dubois Family for their kindness during my stay in Liege, particularly Charline Dubois, who always encouraged and gave me extraordinary help to finish my Ph.D. She was always there for me when I needed help and moral support.

My special gratitude is also extended to Vincent Vanderheyden for providing valuable comments to my research; Ruihao Qiu for helping me develop the script used for data processing. I am also fortunate to have a great group of friends at the Department of Geography - my climatologist friends Sébastien Doutreloup and Alexandre Belleflamme, my LAPLEC colleagues Dieudonné Lekane, Adel Boussaid, Lou Ann Ocampo and Viviane Mallo. I thank them all for providing me with an enjoyable and motivating environment for my research, as well as warm friendship. I am grateful for the time spent with roommates and friends, especially Tiziana Svaldo Lanero and Andy Taitakwong. My gratitude also goes to the many friends whom I cannot name here individually. Thank you for all the kindness you showed me during my doctoral studies.

Finally I would like to dedicate this work to my mother: Irma Y. Orozco J. for her unconditional and endless love, support and sacrifices throughout all my life, especially during the challenging period. She was my source of motivation and determination. She always had faith in me and encouraged me to give the best of myself since I was a child.

I would like to warmly thank my wife Karina and my children Luna and Tristan, they knew more than anyone what this research meant to me. It was indeed valuable to experience the persistent inspiration and support of such a marvelous family. I thank my nephew Daniel who have been a part of my life since was born, my sister Nicté and her husband Estuardo, for the support and care they gave my family in my absence and my brother Hosy and his family. My sincerest gratitude for the continuing support to all of them. I have finally achieved my dream in Belgium.

Summary

LANDSLIDE SUSCEPTIBILITY PREDICTION IN A MOUNTAINOUS CATCHMENT: THE NARANJO BASIN, WESTERN GUATEMALA

In tropical mountain areas, landslides triggered by heavy rainfall represent a natural hazard, especially where meteorological events such as tropical cyclones are recurrent. Rainfall-triggered landslides may also increase in highly seismic active areas. This is the case of Guatemala, located in the Circum-Pacific Belt and between Pacific and Atlantic Oceans, sources of cyclonic events.

Although Guatemala is a region where landslides are a widespread phenomenon, landslide inventories, maps and quantitative susceptibility assessments that are useful for land use planners and decision makers are hitherto non-existent.

The aim of this research is to produce quantitative landslide susceptibility assessments, using logistic regression multivariate statistical method. The study area is a watershed located in the department of San Marcos (western Guatemala), impacted by a tropical storm event in October 2005.

A total of 766 landslides were identified and mapped using orthoimages from 2005. Then, 99 landslides were mapped in 2011 based on field data. The main landslide type is shallow landslide (61 % in 2005), while 39 % of those landslides from 2005 evolved into debris flows.

In total, susceptibility models using multivariate probabilistic approach were developed for shallow landslides, evaluating two different strategies for the sample size of non-landslides events and three different numbers of input variable in the models. Susceptibility models were developed for debris flows dataset and the union of both dataset (shallow and debris flows). The comparison of the models and the associated susceptibility maps highlighted 6 significant input variables that are associated with landslide occurences - elevation, slope, aspect, profile curvature, planform curvature and distance to roads. Performance comparisons of models were also carried out. To validate the performance of the model results, the ROC curve was used, as well as the four-fold and confusion matrix plots. A susceptibility map was generated to display the results of the models in terms of probability values.

A proposal and discussion on the operational use of susceptibility maps where cutoff values can be chosen to define the lowest and highest landslide susceptibility were also made. These will help land use planners in decision-making and in implementing protective measures.

Résumé

PRÉDICTION DE LA SUSCEPTIBILITÉ DES GLISSEMENTS DE TERRAIN DANS UN BASSIN VERSANT MONTAGNEUX: LE BASSIN HYDROGRAPHIQUE NARANJO, GUATEMALA OCCIDENTAL

Dans les régions montagneuses de climat tropical, les glissements de terrain provoqués par de fortes précipitations représentent un danger naturel, en particulier lorsque des événements météorologiques comme les tempêtes interviennent de manière récurrente. Les glissements de terrain déclenchés par les précipitations augmentent d'autant plus que la région est soumise à une forte activité sismique. C'est le cas du Guatemala, pays situé dans la ceinture circum-pacifique et entre les océans Pacifique et Atlantique, vecteurs de cyclones tropicaux.

Bien que le Guatemala soit une région où les glissements de terrain sont des phénomènes répandus, leurs inventaires, leurs cartographies et leurs études statistiques, utiles pour les planificateurs et les décideurs en aménagement du territoire, sont restés jusqu'à présent inexistantes. L'objectif de cette recherche est de produire des évaluations quantitatives de la susceptibilité au glissement de terrain, en utilisant la régression logistique, méthode statistique multivariée. La zone d'étude est un bassin hydrographique situé à San Marcos, dans l'ouest du Guatemala, région impactée par une tempête tropicale en octobre 2005.

Au total, 766 glissements de terrain ont été identifiés et cartographiés à l'aide d'orthoimages de 2005 et 99 glissements de terrain de 2011 ont été identifiés grâce à la reconnaissance sur le terrain. Les principaux types de glissements de terrain sont des glissements de type peu profond (61 % en 2005, nommés *shallow landslides*) dont certains évoluent en des coulées (39 % en 2005, nommés *debris flows*).

Au total, des modèles de susceptibilité utilisant une approche probabiliste multivariée ont été développés pour les glissements de terrain peu profonds, modèles basés sur deux stratégies différentes liées à la taille de l'échantillon référençant des glissements de terrains (*event*) et des événements sans glissements de terrain (*non-event*), et sur la combinaison de trois nombres différents de variables. Des modèles de susceptibilité ont été développés pour la base de données des coulées et pour les deux ensembles de données (glissements peu profonds et coulées) réunis. La comparaison des modèles et des cartes de susceptibilité révèle l'importance de six variables explicatives significativement associées à l'occurrence du glissement de terrain (élévation, pente, orientation de la pente, profil, courbure planiforme et distance à la route). Des comparaisons de performances ont été réalisées et, pour valider la performance des modèles, nous utilisons une *ROC curve*, *four-fold plot*, et une matrice de confusion. Une carte de susceptibilité montre les résultats des modèles en affichant les probabilités d'occurrence.

Finalement, nous proposons et discutons une utilisation opérationnelle des cartes de susceptibilité où nous choisissons les valeurs limites (*Cutoff values*) pour définir les susceptibilités les plus faibles et les plus élevées. L'exploitation de ces cartes sera utile aux planificateurs afin de les aider dans leurs prises de décisions et dans la mise en place de mesures de protection.

Resumen

PREDICCIÓN DE LA SUSCEPTIBILIDAD A DESLIZAMIENTOS EN UNA MONTAÑOSA CUENCA HIDROGRÁFICA: LA CUENCA DEL RIO NARANJO, OCCIDENTE DE GUATEMALA

En áreas montañosas tropicales, los deslizamientos de tierra provocados por fuertes lluvias representan una amenaza natural, especialmente en donde eventos meteorológicos como los huracanes son recurrentes. Deslizamientos provocados por lluvias pueden incrementarse en áreas con alta actividad sísmica. Éste es el caso de Guatemala, localizada en el Cinturón de fuego del pacífico y entre los océanos pacífico y atlántico, origen de huracanes.

A pesar que Guatemala es una región en donde los deslizamientos son un fenómeno generalizado, inventarios de deslizamientos, mapas y evaluaciones cuantitativas de susceptibilidad, útiles para los planificadores de uso de la tierra y responsables de la toma de decisiones, son hasta ahora casi inexistentes.

El objetivo de esta investigación es producir una evaluación cuantitativa de deslizamientos de tierra, utilizando un método estadístico multivariado como la regresión logística. El área de estudio es una microcuenca localizada en el departamento de San Marcos (en el occidente de Guatemala), impactada por una tormenta tropical en Octubre de 2005.

Un total de 766 deslizamientos fueron identificados y mapeados utilizando orthoimagenes del 2005. Además, 99 deslizamientos fueron mapeados en el 2011 sobre el terreno. Los principales tipos de deslizamientos son los deslizamientos superficiales (61 % en 2005, nombrados *shallow landslides*) y otros que evolucionaron a flujos de detritos (39 % en 2005, nombrados *debris flows*).

Modelos de susceptibilidad, utilizando una enfoque probabilístico multivariado, fueron desarrollados para los deslizamientos superficiales, evaluando dos diferentes estrategias de tamaño de muestra referente a la ocurrencia de deslizamientos y a la no ocurrencia de deslizamientos, y sobre la combinación de tres números diferentes de variables. Modelos también fueron desarrollados para los flujos de detritos y la unión de ambos tipos (deslizamientos superficiales y flujos de detritos). La comparación de los modelos y los mapas de susceptibilidad, revelan la importancia de 6 variables significativas asociadas a la ocurrencia de los deslizamientos de tierra (elevación, pendiente, orientación de la pendiente, perfil de la curvatura, plano de la curvatura, y la distancia a los caminos). Comparaciones del desempeño de los modelos han sido realizadas y para validar su desempeño se utilizaron la curva ROC, *four-fold* plot, y la matriz de confusión. Mapas de susceptibilidad muestran los resultados de los modelos visualizando las probabilidades de ocurrencia.

Finalmente, proponemos y discutimos una utilización operacional de los mapas de susceptibilidad en donde elegimos los valores de corte (*Cutoff values*) para definir las áreas con las susceptibilidades más bajas y más altas. Esto será de utilidad para los planificadores del uso de la tierra, ayudándoles a la toma decisiones y a la implementación de medidas de mitigación.

Table of contents

1 INTRODUCTION AND OBJECTIVES

Mountain areas, especially in the tropical zone, are particularly subject to landslides which are often associated with a trigger. Earthquakes and rainfall are the most common triggering mechanisms (Harp *et al*., 1978; Keefer, 1984; Keefer, 2002; Bommer & Rodriguez, 2002; Galli *et al*., 2008; Korup *et al*., 2010; Lei, 2012; Keefer, 2013; Parker *et al*., 2016). Slope failures generally occur within minutes after an earthquake trigger, and days to weeks after an intense rainfall trigger (Malamud *et al*., 2004). Human activities can also be the cause such as the construction of roadways and building excavations, mining, and vegetation removal (Jaboyedoff *et al*., 2016; Gill & Malamud, 2017).

In some cases, it can be expected that earthquakes do not trigger landslides directly but act as a preparatory factor, with landslides being more easily triggered by rainfall when it occurs after an earthquake (Chang *et al*., 2007; Zhang *et al*., 2014). This is because seismic shaking induces dynamic stress and thereby reduces shear strength and elevates pore pressure (Chung *et al*., 2014). The role of seismicity as preparatory factor has a great importance for the economy and society (Vanmaercke *et al*., 2017; Efremidis *et al*., 2017).

Parameters like rainfall duration, rainfall intensity, antecedent rainfall and cumulative rainfall, have been employed to define rainfall-thresholds and establish the relationship between rainfall and landslides in various parts of the world (Sidle *et al*., 2006; Dahal & Hasegawa, 2008; Guzzetti *et al*., 2008a; Sidle & Bogaard, 2016). Different approaches suggest that specific rainfall intensity and duration circumstances trigger shallow landslides and debris flows. In mountain regions where meteorological events such as tropical cyclones are recurrent, the consequences can be devastating (Dahal & Hasegawa, 2008; Cascini *et al*., 2011; Alcantara-Ayala, 2004).

Different catalogs with information about historical occurrence of landslides at global scale have been compiled by various authors. Geographically speaking, all continents and countries are not equal. Cavallo and Noy (2010) recall that “the overwhelming majority of people affected and killed by natural disasters reside in developing countries”. According to Alcantara-Ayala and Oliver-Smith (2014), Asia has been the continent the most affected by landslides, followed by America, Europe, and Africa. Between the years 2000 and 2011, around 4,041,000 people were affected by landslides worldwide. The global landslide catalog compiled by Kirschbaum *et al*. (2015), with 5741 points of references, evaluated the spatial and temporal trends in landslide activity from 2007 to 2013. Globally, landslides were reported most frequently from July to September. Most events occurred in North America and Southeast Asia. The Central American and Caribbean region have a high number of landslides reports and fatalities during the months of September and October, when tropical cyclones are more frequent (Kirschbaum *et al*., 2010). According to Sepúlveda and Petley (2015), the Enhanced Durham Fatal Landslide Database – EDFLD – recorded a total of 611 fatal landslides causing 11,631 deaths in Latin America and the Caribbean in the 10-year period between 2004 and 2013. Fatal landslides occurred in 25 countries of which seven in Central America, nine in South America and nine in the Caribbean. Almost 90 % of the recorded cases in the EDFLD were triggered by heavy rainfall, of which 15 % were clearly identified as related to tropical cyclones (hurricanes). The monthly distribution of landslides and the average monthly precipitation are strongly correlated.

Central America is situated in the Circum-Pacific Belt, also known as the Pacific Ring of Fire. Earthquakes and volcanic activity are particularly frequent along the Cocos-North American and Cocos-Caribbean subduction zones. Ten major volcanoes and up to 300 smaller emission centers are concentrated in the Quaternary volcanic arc of Guatemala. The Central American tectonic setting induces very high seismicity (Alcantara-Ayala, 2009).

Guatemala, in Central America, has been severely affected by hazards including earthquakes, volcanic activity, floods, tropical cyclones, storms, and landslides during the past and present century (Alcantara-Ayala, 2009). Indeed, the global assessment reports on disasters and risk reduction as elaborated by the United Nations International Strategy for Disaster Reduction Secretariat (UNISDR, 2009, 2011, 2013, 2015) ranked Guatemala among the top ten countries on the Mortality Risk Index for landslides triggered by high precipitation, in the top ten countries on the Mortality Risk Index for earthquakes, and among the top 15 countries for multi-hazard average annual social costs.

Devastating earthquakes have impacted Guatemala; the most destructive was the 7.5 magnitude event with the epicenter located on the Motagua fault zone which occurred on 4, February 1976 (Porfido *et al*., 2015; Harp *et al*., 1978; Husid, 1978). This earthquake triggered landslides over an area of approximately 16,000 km^2 and created the greatest human and economic losses to date. More recently, a 7.4 magnitude earthquake struck off the Pacific coast of Guatemala on 7 November 2012, killing 44 people. Guatemalan authorities revealed that 127 municipalities in 11 departments were affected (OCHA, 2012).

Rainfall induced landslides represent the major hazard in Guatemala. The rainfall intensity that accompanied hurricane Mitch in October and November of 1998 affected Guatemala and triggered thousands of landslides. In the same way that hurricane Mitch hit the country, tropical storm Stan occurred in Central America and Mexico and reached the Highlands of Guatemala. From October 1st to 9th the continuous heavy rainfall resulted in destructive flooding and landslides (Bucknam *et al*., 2001; Coe *et al*., 2004; Silverman, 2011; CEPAL, 2005). Figure 1 shows the temporal distribution of the earthquakes and tropical cyclones (hurricanes) that occurred from June 1986 to December 20016.

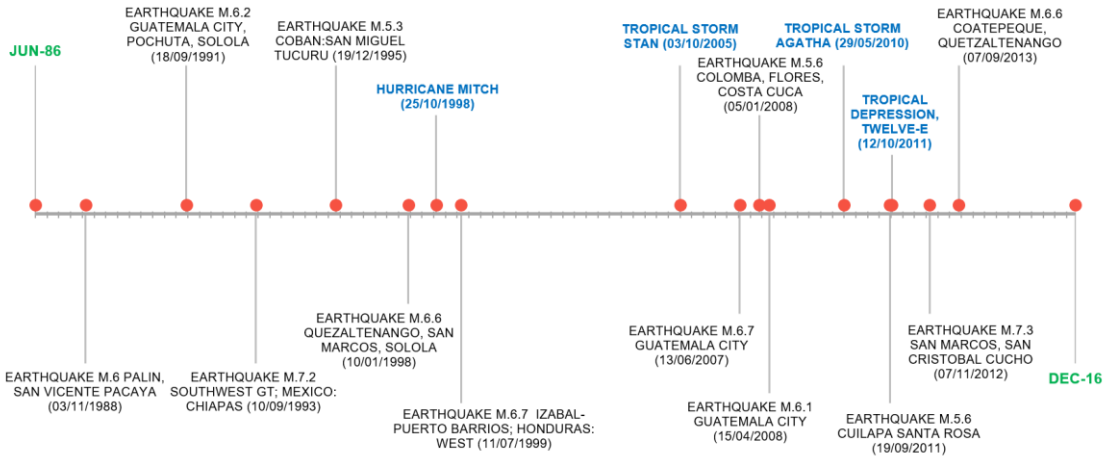


Figure 1 — Timeline of earthquakes and rainfall events, based on the Registration Spatial Distribution of Seismic Events in Guatemala. From the National Institute of Seismology, Volcanology, Meteorology and Hydrology (INSIVUMEH, 2006) & the Geophysical Data Center of the National Oceanic and Atmospheric Administration (NOAA, 2016).

In Guatemala, landslides susceptibility is a relatively underdeveloped area of research. This constitutes a real challenge and contributes to the originality of this research. The data-poor context is explained by some reasons. First, the information about landslide types is very difficult to collect; there are no reliable quantitative data about landslides because erosion processes, plowing, and the rapidly growing vegetation make traces of landslides less visible in the landscape. Second, the data quality of predisposing factors is poor or scarce. Third, the disturbances caused by human activities due to population expansion in the mountainous regions such as the construction of roads that destabilize slopes, breaking natural habitats , and the transformation of landscapes are not computed.

One of the main problems in Guatemala is that landslides occur in densely populated areas where mitigation strategies are yet to be set up. The first need is therefore to understand the parameters that explain the spatial occurrence of the landslide processes triggered by tropical cyclones in those populated areas. For that purpose, landslides inventories and prediction maps are required.

In Guatemala, more precisely in the department of San Marcos and in the Naranjo watershed which are the study areas of this research, mapping and statistical studies are almost inexistent. Yet, creating landslide susceptibility maps is important to land use planners and decision makers to reduce disaster risk. This study will scrutinize hillslope processes and produce susceptibility maps based on the modeling of landslide events occurred in the past.

Quantitative data about landslides in the studied watershed are not available, as well as the clear insight on the local causes of the landslides related to the 2005 tropical storm Stan. However, despite the poor data conditions, it is important to define a consistent probabilistic susceptibility model to understand how the variables are involved in the process.

As reviewed by Corominas *et al*. (2014), methods for modeling landslide susceptibility are numerous. The methods can be qualitative (inventory-based and knowledge-driven methods) and quantitative (data-driven statistical methods and physically based models). Logistic regression is one of the most frequently chosen data-driven methods to assess landslide susceptibility (Korup & Stolle, 2014; Corominas *et al*., 2014). This regression, multivariate statistical method, will be applied here to assess landslide susceptibility in order to understand the relationships between predisposing factors and the occurrence of different landslides types in this mountain watershed. The impact of predictor variable selection, sample size and sample stratification strategy on model performance will be assessed in order to improve the resulting susceptibility models.

The main aim of this research is to produce quantitative landslide susceptibility assessments in the high part of the Naranjo Watershed, San Marcos. This area is representative of the mountain environment in terms of natural and anthropological conditions found in the western part of Guatemala. The selected area is known for having been struck by numerous landslides that were triggered by rainfall.

The specific objectives of this research are:

- To create inventories of landslides in the specified area of interest;
- To analyze the spatial distribution of landslides and their predisposing factors;
- To understand the factors explaining the spatial occurrence of landslides triggered by tropical cyclones in the populated mountain areas in Guatemala.

Landslide susceptibility maps that can be used even by non-experts will be provided to aid in sustainable land management.

2 LANDSLIDES: STATE OF THE ART

2.1 Landslides definition

Landslides are defined as “the gravitational movement of a mass of rock, debris or earth down a slope” (Cruden & Varnes, 1996; Hungr *et al*., 2014). Landslides are a complex process that play a major role in the evolution of landforms (Malamud *et al*., 2004). Landslides may be caused by various factors ranging from natural to anthropogenic origins (see section 2.2).

In international literature, there are many approaches for the classification of landslides (Varnes, 1978; Hutchinson, 1968, 1988; Cruden & Varnes, 1996; Crozier & Glade, 2005; Turner & Schuster, 1996; Hungr, 2005; Hungr *et al*., 2014). The most widely accepted classification of landslides is that of Varnes (1978), with modifications proposed by Cruden and Varnes (1996) and recently updated by Hungr *et al*. (2014), as shown in the Table 1.

Table 1 — Classification of landslides. Updated by Hungr et al. (2014) and based on Cruden and Varnes (1996).

Type of movement	Rock	Soil
Fall	1. Rock/ice fall	2. Boulder/debris/silt fall
Topple	3. Rock block topple 4. Rock flexural topple	5. Gravel/sand/silt topple
Slide	6. Rock rotational slide 7. Rock planar slide 8. Rock wedge slide 9. Rock compound slide 10. Rock irregular slide	11. Clay/silt rotational slide 12. Clay/silt planar slide 13. Gravel/sand/debris slide 14. Clay/silt compound slide
Spread	15. Rock slope spread	16. Sand/silt liquefaction spread 17. Sensitive clay spread
Flow	18. Rock/ice avalanche	19. Sand/silt/debris dry flow 20. Sand/silt/debris flowslide 21. Sensitive clay flowslide 22. Debris flow 23. Mud flow 24. Debris flood 25. Debris avalanche 26. Earthflow 27. Peat flow
Slope deformation	28. Mountain slope deformation 29. Rock slope deformation	30. Soil slope deformation 31. Soil creep 32. Solifluction

Beyond the material and landslides process, a third parameter, the position of the basal shear surface, allows to distinguish shallow and deep-seated landslides. In mountain regions, shallow, debris flows and deep-seated types are a common phenomenon (Glade, 2003; Sidle & Bogaard, 2016).

2.2 Causes of landslides

Landslides occur naturally, but they are also related to anthropogenic influences on the environment, changing the extent to which it is susceptible to slope failure and thus increase or decrease landslide occurrence or its size (Jaboyedoff *et al*., 2016; Gill & Malamud, 2017). Sometimes, landslides are caused, or aggravated, by a combination of the two factors (Van Asch *et al*., 1999; Holm *et al*., 2003; Alcantara-Ayala, 2004; Alcantara-Ayala *et al*., 2006; Mikoš *et al*., 2004; Paolini *et al*., 2005; Ardizzone *et al*., 2007; Dahal & Hasegawa, 2008; Highland & Bobrowsky, 2008; Korup *et al*., 2010; Turner *et al*., 2010; Weng *et al*., 2011; Lin *et al*., 2011; Mondini *et al*., 2011; Cascini *et al*., 2011; Broothaerts *et al*., 2012).

The occurrences are linked to three mechanisms that can occur either alone or in combination: (1) rainfall, (2) seismic activity, and (3) volcanic activity. The effects of these causes may vary widely and depend on factors such as the degree of the slope, morphology, soil type, underlying geology (Highland & Bobrowsky, 2008).

In tropical mountain areas, rainfall events can induce hillslope instability that may be devastating (Alcantara-Ayala, 2004). The rainfall intensity, duration, accumulative and earlier rainfall are the parameters commonly investigated and related to landslide initiation (Dahal & Hasegawa, 2008; Salciarini *et al*., 2006; Guzzetti *et al*., 2008a; Cascini *et al*., 2011; Sidle & Bogaard, 2016).

The largest earthquakes can trigger tens of thousands of landslides throughout regions (Keefer, 2013). Landslides caused by earthquakes may be triggered in dry material, particularly in high seismic risk areas. However, seismic disturbance usually combines with pore-water pressures to generate the destabilization of slopes (Keefer, 2002). For example, Lin *et al*. (2006) investigated the influence of the Chi-Chi earthquake on subsequent rainfall-induced landslides by comparing the occurrence of landslides in the Choushui River watershed in Taiwan. The study revealed that the density of rain-induced landslides increased substantially after the earthquake and that the landslide locations changed as well. In 2008, the Wenchuan earthquake in China triggered more than 15,000 landslides, much of the landslides were reactivated and the earthquake has a significant impact on the subsequent rain-induced landslides. During the rainy seasons after the earthquake, the density of landslides increased while the area of landslides decreased significantly and the type of landslides transformed from debris slides to debris flows (Zhang *et al*., 2014).

The precipitation coming from storms during short periods of very intense rainfall trigger shallow landslides and debris flow. These events are an important issue in urban planning: the more mountains are urbanized and inhabited, the higher the risk for people living in those areas (Vega & Hidalgo, 2016). Population growth and urban sprawl on new land also create new neighborhoods, towns, and cities that are the primary vector by which humans and their activities contribute to the occurrence of landslides.

In fact, disturbing or changing drainage patterns, destabilizing slopes, and removal of vegetation are common human-induced factors that may initiate landslides. Other examples include oversteepening of slopes by undercutting the bottom, and loading the top of a slope to exceed the bearing strength of the soil. However, landslides may also occur in once-stable areas due to

other human activities such as irrigation, lawn watering, reservoir draining or creation, leaking pipes, and improper excavating or grading on slopes (Highland & Bobrowsky, 2008).

2.3 Landslides inventories

Mapping landslides and producing landslide inventory maps have received special attention from a wide range of specialists. These maps provide effective and easily understandable products for experts such as geomorphologists and for non-experts, including decision makers, planners, and civil defense managers (Guzzetti *et al*., 2003; Galli *et al*., 2008; Moosavi *et al*., 2014).

The basis of all regional landslides studies, as an elementary step, are the landslide inventory maps, with information about landslide locations and their specific characteristics for the assessment of landslide susceptibility or landslide risk (Galli *et al*., 2008; Guzzetti *et al*., 2012). This could also be used to predict future patterns of instability from the past and present distribution of landslides (Guzzetti *et al*., 1999).

Landslide inventory may be compiled at different scales, from the local to the national, using a variety of techniques, including the analysis of stereoscopic aerial photographs, geomorphological field mapping, engineering geological slope investigations, and the examination of historical archives. Selection of a specific technique depends on the purpose of the inventory, the extent of the study area, and the scale of the base maps. Of course, a combination of these techniques is often used. (Guzzetti *et al*., 2000, 2012; Malamud *et al*., 2004; Ardizzone *et al*., 2007).

The limitations of landslide inventories from aerial photographs or derived products such as orthoimages result to subjectivity and the difficulty of measuring reliability. The reliability of geomorphological inventories depends on the type and scale of the aerial photographs used; the experience and skill of the interpreter; the scale of the final map; the complexity of the geological, morphological, and land-use setting; and the persistence of land slide morphology within the landscape (Guzzetti *et al*., 2000).

In the last decade, researchers generally mapped landslides using vector-based representations of the landslide data, which are represented by points, lines or polygons and interpretation of high resolution post-event color orthoimages (Brenning, 2005; Galli *et al*., 2008; Van Westen *et al*., 2008; Chung & Fabbri, 2003; Trigila *et al*., 2015; Hussin *et al*., 2016).

Shallow landslides and debris flow represent a destructive natural hazard and a disturbance of the ecosystem in steep mountainous areas. Research papers concerning areas affected by shallow landslides have stated that many of them evolve into debris flows. The potential debris flow hazard as a consequence of shallow landslides has thus motivated the systematic mapping of those shallow landslides and their susceptibility assessment (Dietrich *et al*., 1995; Dahal & Hasegawa, 2008; Salciarini *et al*., 2006; Guzzetti *et al*., 2008b; Mondini *et al*., 2011; Dislich & Huth, 2012; Parker *et al*., 2016; Fan *et al*., 2017).

2.4 Predisposing factors

Landslides susceptibility assessment involves identifying the causative factors that contribute to the instability of the slope. In landslides prediction, it is necessary to assume that landslide occurrence is determined by landslide predisposing factors, and that future landslides will occur under the same conditions as past landslides (Lee & Talib, 2005). It is not possible to have a prescribed uniform list of predisposing factors since it depends on the scale of analysis, the characteristics of the study area, the landslide type and the failure mechanisms (Crozier & Glade, 2005; Van Westen *et al*., 2008; Corominas *et al*., 2014). Van Westen *et al*. (2008) gave a schematic overview of the main data layers and their relevance in landslide susceptibility. This list was complemented and updated by Corominas *et al*. (2014). Both indicated that the list is not exhaustive, and that it is important to select the specific factors that are related to the landslide types and failure mechanisms in each particular environment.

The predisposing factors were grouped according to topography, geology, soil, hydrology, geomorphology, land use, earthquakes, volcanoes, weather and climatic conditions details, and anthropogenic factors. Details of each group are specified in Corominas *et al.* (2014).

2.5 Landslides susceptibility assessment

According to Lee and Talib (2005), the relationship between areas where a landslide has occurred and landslide predisposing factors can be distinguished from the relationship between areas without past landslides and landslide predisposing factors. The frequency ratio method can provide a simple geospatial assessment tool to calculate the probabilistic relationship between dependent (landslide) and independent variables (predisposing factor) (e.g. Lee & Talib, 2005; Ozdemir, 2011; Shabanzadeh *et al*., 2011; Yalcin *et al*., 2011; Kannan *et al*., 2013; Ozdemir & Altural, 2013; Solaimani *et al*., 2013; Wu *et al*., 2016; Wang *et al*., 2015).

Landslide susceptibility, proposed by Brabb (1984) quoted by Rossi *et al*. (2010a), is the spatially relative likelihood of a landslide occurring in an area on the basis of local terrain conditions but it does not consider the temporal probability of failure and the magnitude of the expected landslide. So, assessment of landslide susceptibility involves determining “where” landslides are expected and for this reason, landslide susceptibility is different from landslide hazard (Guzzetti *et al*., 2000, 2005, 2009; Rossi *et al*., 2010a; Corominas *et al*., 2014).

Several methods and techniques (qualitative or quantitative and direct or indirect) have been proposed and tested in order to evaluate landslide susceptibility, but no general agreement has been reached yet about the most adapted methods to assess the phenomena. A literature review proposed by Guzzetti *et al.* (1999) showed a predilection for indirect and quantitative approaches based on statistical (i.e. probabilistic) or process-driven (i.e. geotechnical) modeling. But susceptibility assessment methods for areas where landslides have previously occurred are different from methods for areas where landslides might occur but where no existing landslide is known. With more complex methods being applied at larger scales, amount of data required greatly increases (Corominas *et al*., 2014).

The methods used for landslide susceptibility analysis are usually based on widely accepted basic assumptions such as homogeneity of the geological conditions, similar probability of

slope failure, or similar sizes of landslides, (Corominas *et al*., 2014). According to Guzzetti et al., (1999) the main principles of landslide susceptibility analysis are:

- Most of the slope failures produce discernible morphological features that can be recognized, classified and surveyed both in the field or through remote sensing, chiefly aerial photographs.
- Conditions that caused landslides instability directly or indirectly linked to slope failure, can be collected and used to set up predictive models of landslide occurrence (Dietrich *et al*., 1995).
- Next slope failures would occur under conditions which led to past and present instability (Varnes, 1978; Carrara *et al*., 1991; Hutchinson, 1995; Hutchinson & Chandler, 1991).
- Landslide occurrence can be inferred by different approaches (heuristic investigations, inferred from physical models, probabilistic models, among others) (Guzzetti *et al*., 1999).

In knowledge-driven or heuristic methods, the landslide susceptibility map can be prepared directly in the field by expert geomorphologists, or created in the office as a derivative map of a geomorphological map and depend on how much the investigator understands the geomorphological processes at work (Guzzetti *et al*., 1999; Van Den Eeckhaut, 2006; Corominas *et al*., 2014).

Physically-based landslide susceptibility assessment methods are based on the modelling of slope failure processes. The methods are only applicable over small areas when the geological and geomorphological conditions are fairly homogeneous and the landslide types are simple (Rossi *et al*., 2010b; Guzzetti *et al*., 1999; Corominas *et al*., 2014).

In data-driven landslide susceptibility assessment methods, the combinations of factors that have triggered but also landslides triggered in the past are evaluated statistically. The output may be expressed in terms of probability. These methods are termed “data-driven”, as data from past landslide occurrences are used to obtain information on the relative importance of the factor maps and classes (Corominas *et al*., 2014). Forms of statistical methods are used to estimate the relative contributions of the factors responsible for slope instability and to create some predictions based on these factors. These methods can be categorized into two subgroups, bivariate and multivariate methods (Soeters & van Westen, 1996). Bivariate methods assume that the factors are not correlated with each other whereas, in multivariate methods, all factors are treated together and their interactions are handled differently (Süzen & Doyuran, 2004). Bivariate statistical methods are exploratory preliminary analyses to the use of multivariate statistical methods (Corominas *et al*., 2014). Multivariate statistical models evaluate the combined relationship of a dependent variable (landslide occurrence) to a series of independent variables (controlling factors). In this type of analysis, all relevant factors are measured and sampled either on a grid basis or in slope morphometric units (see below). For each of the sampling units, the presence or absence of landslides is determined. The resulting matrix is then analyzed using multiple regression, multiple logistic regression, discriminant analysis, random forest or active learning, etc. In Table 2 some data-driven approaches are mentioned (Corominas *et al*., 2014).

Table 2 — Methods for data-driven landslide susceptibility assessment. According to Corominas et al. (2014).

Data driven approach	Method
Bivariate statistical methods	Likelihood ratio model (LRM) Information value method Weights of evidence modelling Favorability functions
Multivariate statistical method	Discriminant analysis Logistic regression Artificial neural networks

Other statistical methods were included in Costanzo *et al*. (2013) and in Corominas *et al*. (2014), such as conditional analysis, classification and regression trees, including also artificial neural networks. The authors made a summary of the most used statistical techniques and showed that the logistic regression is one of the most suitable and performing methods for the assessment of landslide susceptibility (Rossi *et al*., 2010a; Costanzo *et al*., 2013; Van Den Eeckhaut, 2006; Van Den Eeckhaut *et al*., 2010; Wang *et al*., 2013; Dai & Lee 2002; Guzzetti *et al*., 1999; Melchiorre *et al*., 2008).

Logistic regression is well suited to analyze a binary presence–absence dependent variable and has been used to predict slope instability. This multivariate technique for landslide susceptibility modeling works on any type of independent variable (continuous, discrete, ordinal or nominal scale), regardless of the deviance of the considered predictors and residuals from a normal distribution (e.g. Carrara *et al*., 1991; Guzzetti *et al*., 1999; Hosmer & Lemeshow, 2000; Dai & Lee, 2002; Peng & So, 2002; Ohlmacher & Davis, 2003; Lee *et al*., 2004; Ayalew & Yamagishi, 2005; Guzzetti *et al*., 2005; Nefeslioglu *et al*., 2008; Rossi *et al*., 2010a; Van Den Eeckhaut *et al*., 2010; Ozdemir, 2011; Yalcin *et al*., 2011; Bui *et al*., 2011; King & Zeng, 2001; Costanzo *et al*., 2013; Wang *et al*., 2013; Ozdemir & Altural, 2013; Solaimani *et al*., 2013; Heckmann *et al*., 2014; Trigila *et al*., 2015; Wang *et al*., 2015; Dewitte *et al*., 2015).

The choice of the mapping unit is crucial because it also determines how landslides will be sampled to prepare the training and prediction (validation) subsets for the susceptibility modeling (Van Den Eeckhaut, 2006; Hussin *et al*., 2016). Differences between these units were cited by Guzzetti *et al*. (1999) as follows, among which grid-cells and slope-units were considered very important. In this study, landslides in a deep valley were modeled using DEM.

Grid-cells: they divide the territory into regular squares of pre-defined size and are the reference spatial mapping unit, preferred by raster-based GIS users.

1. Terrain units: they are based on the observation that in natural environments the interrelations between materials, forms and processes result in boundaries which are frequently reflected in geomorphological differences, favored by geomorphologists.
2. Unique-condition units: they imply the classification of each slope-instability factor into a few significant classes which are stored into a single map, or layer that combines the information on all relevant factors.
3. Slope-units: they are automatically derived from high-quality DTMs and partition the territory into hydrological regions between drainage and divide lines.

4. Topographic units: they are defined by the intersections of contours and flow tube boundaries orthogonal to contours.

Multiple factors should be involved in the selection of the appropriate mapping unit: landslide type, investigation scale, resolution, quality and type of the thematic information, analysis tools, each one with their own limitations and advantages in relation with the chosen method for susceptibility evaluation (Guzzetti *et al*., 1999).

3 STUDY AREA

The study area is located in the department of San Marcos in the western part of Guatemala. It is located in the Naranjo River catchment which drains to the Pacific Ocean. It is one of the 38 watersheds in Guatemala. The upper part of this river catchment is the most prone to landslides that are triggered by rainfall. The study focused on this upper part which is representative of other mountain regions in terms of natural and anthropological conditions affected by tropical cyclones.

The study area has steep mountain slopes. The high pressures of human activities on these slopes result in high vulnerability to landslides, a common situation in the region. Remnants of past landslides can be observed in the landscape. Based on personal experience as a native of the region, landslides indeed occur in the area every rainy season. However, the landslides associated with exceptional rainfall event, particularly during the Tropical storm Stan in 2005, were unprecedented.

The study area covers around 16 km². It is located 250 kilometers to the west of Guatemala City, within highlands and mountain chains (Figure 2). In total, 9 villages from two municipalities of the department of San Marcos are included in this area, Chamac, Ixhual, Hierba Buena (Municipality of San Pedro Sacatepéquez), San Ramon, Siete Tambores, Santa Rita, San Antonio, Candelaria Siquival and Barrancas (Municipality of San Antonio Sacatepéquez). The total population projected in 2016 was 9,073 people, corresponding to a 2.5 % growth per year according to the National Institute of statistic, based on the last National Census (2002).

3.1 General features

The study area is located within the Central America Volcanic Arc (CAVA). According to the Guatemalan 1:250,000 geological map, the predominant lithology consists of Tertiary volcanic rocks (Tv) that include andesitic and basaltic tuffs and lahars conglomerates, and Quaternary pyroclastic deposits (Qp) including various pumice deposits. Volcanic rocks produce both soils with a high clay content and sandy soils, respectively locally known as “tierra de barro” and “tierra arenosa”.

Situated within tropical latitudes, the watershed does not experience marked seasonal changes. The seasons are divided into dry, from November to April, and rainy seasons, from May to October. Peel *et al*. (2007) classify the Guatemala climate type as Tropical rainforest (Af) in their updated world map of the Köppen-Geiger climate classification.

In the study area, the mean temperature is 14 °C at 2358 m a.s.l. (Orozco, 2007). Moreover, the data recorded in 15 weather stations located in the south-western Guatemala show that the moist-adiabatic lapse rate is approximately of 5.6 °C / km (Orozco, 2007).

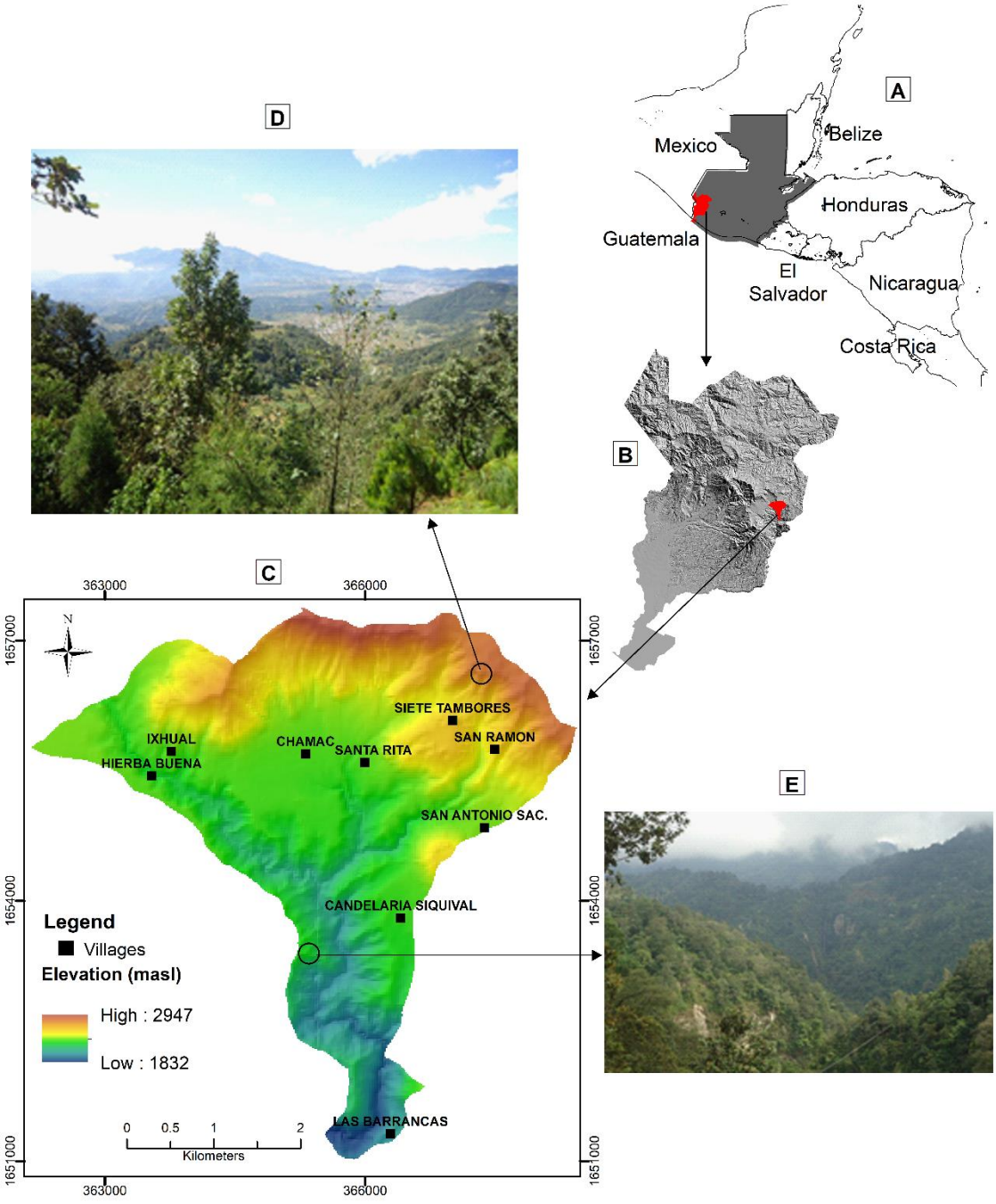


Figure 2 — Study area. A) Guatemala in Central America. B) Department of San Marcos. C) Study area and upper part of the Naranjo River watershed. D) Photo from North-West to South-West (© Estrada, December 2011). E) Photo of the low part from South-East to North-East (© Estrada, December 2011).

On the climograph of San Marcos (Figure 3), rainy season peaks in June and as well as in September.

The meteorological data was obtained from the National Institute of Seismology, Volcanology, Meteorology and Hydrology (INSIVUMEH) for the last 27 years (1990 – 2016), based on the records of their weather station located in the upper part of the San Marcos municipality at 2358 meters above mean-sea level, and with a distance of 15 km from the study area.

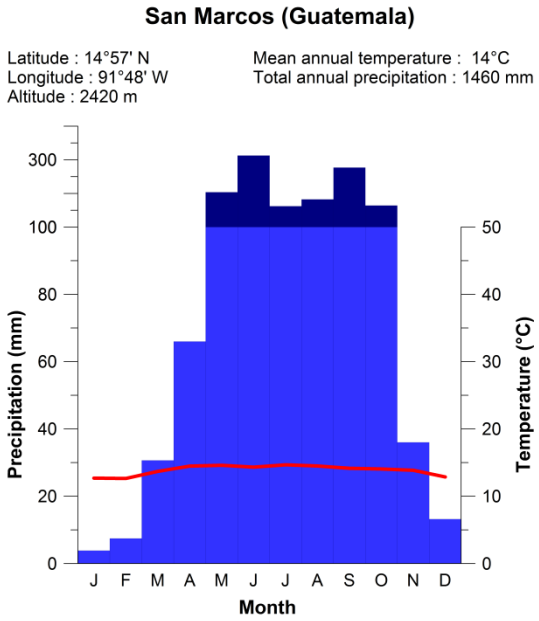


Figure 3 — Precipitation and temperature diagram from the weather station of the city of San Marcos (INSIVUMEH, 2016).

The altitudinal range is approximately 1000 meters. The highest elevation is 2,947 m a.s.l. and the low part is near 1,832 m a.s.l. Dominant drainages run roughly from North to South through steep terrain. The slopes range between 0 and 284 %. However, 80 % of all slopes are of less than 75 %.

The main land cover in the study area is forest with 52 % of the whole surface composed mostly of conifers, mainly pine (*Pinus ayacahuite* and *P. hartwegii*), and subordinate broadleaf trees, mainly oak (*Quercus sp.*) and alder (*Alnus jorullensis H.B.K.*). Agriculture is practiced on 37 % of the total area with maize and beans as main crops and some vegetables like potatoes and carrots as well. Built areas represent 5 % of the study area, pastures 3 % and roads 2 % (Figure 4). Low slope zones are generally built up areas, but houses are also found in high slope zones.

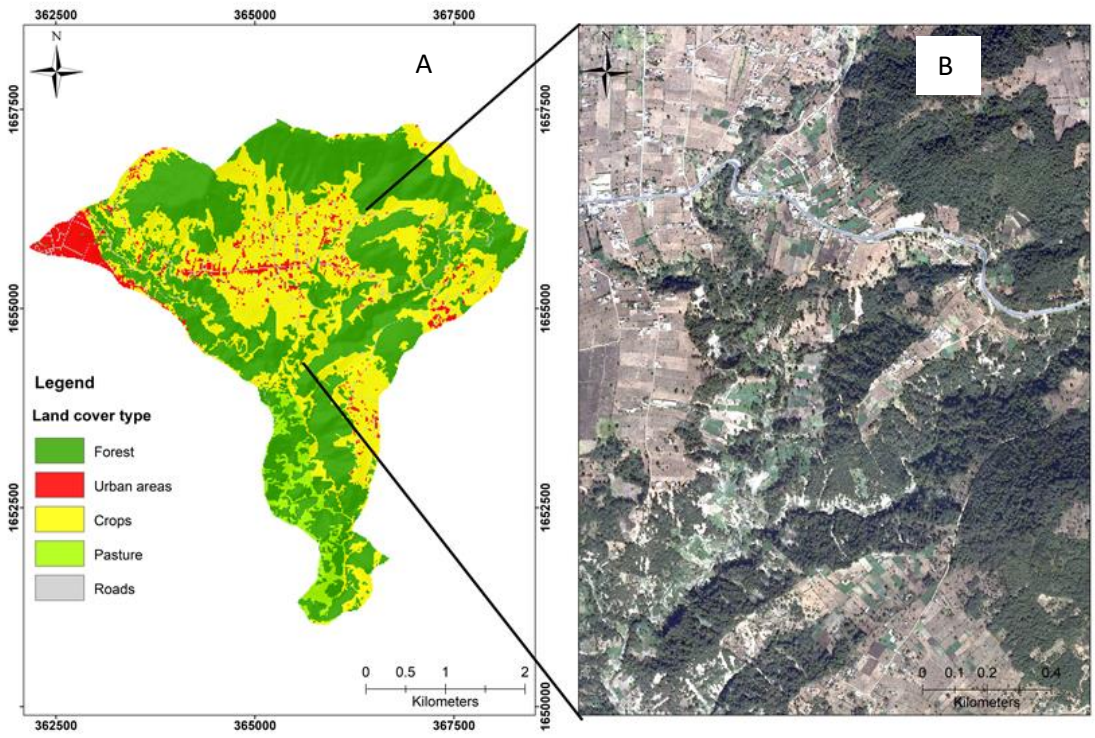


Figure 4 — A) Land cover map. B) Orthoimage showing land cover in a sub-area.

A ~63-km road network is present in this watershed. 85 % are dirt roads passable by four-wheel-drive vehicles and 9.5 kilometers are paved roads including the Inter-American highway which is the main road linking San Marcos to Guatemala City. Roads also create drainage problems, initiating gullies that go through the settlements. The roads are also associated with landslide occurrences in the rainy season.

As reported in the municipal development plan of San Pedro Sacatepéquez and San Antonio Sacatepéquez, approximately 66 % of the people in the municipality of San Antonio Sacatepéquez live in conditions of poverty and 15 % in extreme poverty (according to the international standard of extreme poverty which is set to the possession of less than 1$ a day) (Human Development Report, 1997 in SEGEPLAN, 2010a&b). In San Pedro Sacatepéquez the poverty rate is of 53 % and extreme poverty rate is of 11 % (SEGEPLAN, 2010a-b).

The local communities can be qualified as rural, as 65 % of the whole population is engaged in some form of agriculture. Other economic activities reported include textile industry, local retail trade, construction and public administration (SEGEPLAN, 2010a-b).

3.2 Tropical cyclone events

According to the National Oceanic and Atmospheric Administration (NOAA), the Atlantic tropical storm season runs from June to November, and the Eastern Pacific tropical cyclone season runs from May to November. The Atlantic basin includes the Atlantic Ocean, the Caribbean Sea and the Gulf of Mexico. The tropical cyclone history is shown on Figure 5.

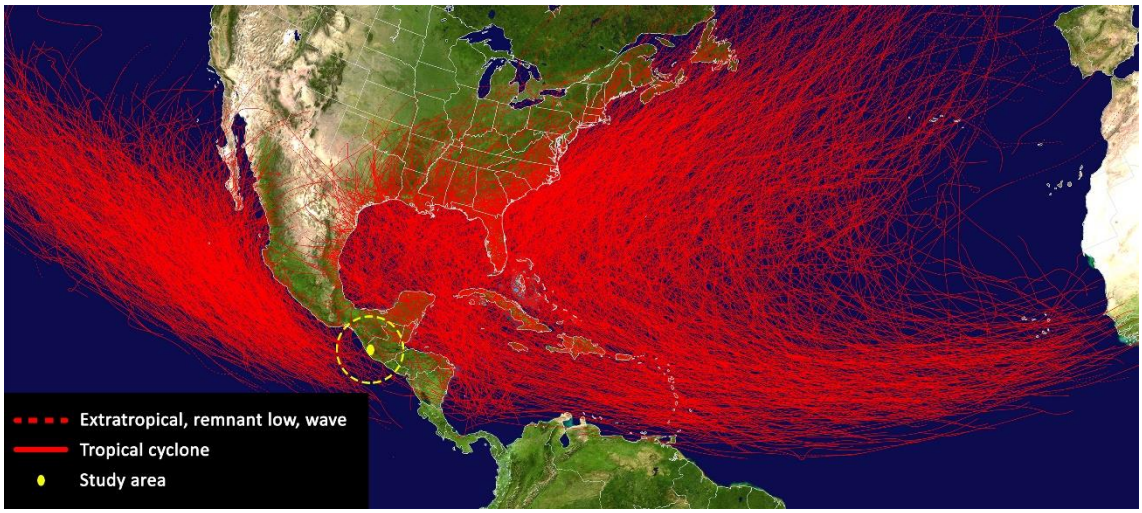


Figure 5 — Tropical cyclone history: all North Atlantic and Eastern North Pacific tropical cyclones. Data from 1949 in the Pacific, from 1851 in the Atlantic to 2014 (Courtesy of NOAA/NWS, 2016). The yellow point locates the study area.

Even if Pacific tropical cyclones generally head away toward the North-West, they may originate near Guatemala. While the volcanic front in this country acts as an orographic barrier to the monsoonal rainy season precipitation, tropical cyclone events coming from the Atlantic, cause heavy rain onshore, beyond the volcanoes with several impacts across the mountainous terrain in Guatemala (Silverman, 2011).

In 1998, hurricane “Mitch” impacted Central America, especially Guatemala. Exceptional rainfall lasted about one week at the end of the rainy season. Rainfall brought by “Mitch” triggered thousands of landslides in eastern Guatemala, which received between 200 mm and 600 mm of rain over a 13-day period between October 25 and November 06. The highest rainfall reached around 400 mm to 600 mm during the whole event; whereas the rainiest month in Eastern Guatemala usually cumulate around 420 mm (Bucknam *et al*., 2001; Silverman, 2011). San Marcos was also impacted by the event but no literature was found concerning this area.

In 2005, another torrential rainfall event affected Guatemala across nearly 70 % of the national territory causing hundreds of deaths, thousands of persons displaced and infrastructure collapses (Picard, 2007). This meteorological event was a tropical storm named “Stan”. On October 2, 2005, a low pressure system developed in the Caribbean Sea and moved towards the southern Gulf of Mexico with a status of hurricane and became the almost stationary tropical storm “Stan” creating a massive low-pressure system over Central America and providing humidity from the Pacific Ocean that resulted in heavy rain across the region. Guatemala was the most severely affected country between October 2 and October 10 (Picard, 2007; CEPAL, 2005).

The torrential rainfalls that accompanied “Stan” triggered thousands of landslides in the occidental part of Guatemala. Tropical storm Stan occurred during the 2005 Atlantic tropical cyclone season, the most active season to date. Initially weak, it eventually reached Category 1 on the Saffir-Simpson scale (Figure 6). The official number of victims and damages established by the National Coordinator for Disaster Reduction (CONRED for the Spanish acronym) (Bulletin

No. 88 dated 22/10/2005) was of 669 deaths, 844 disappearances, 386 injuries, 493,965 people who suffered direct damage, 14,743 provided with emergency shelter, 174 community emergency shelters established and 1,222 communities affected (Picard, 2007; CEPAL, 2005).

Figure 6 — Map showing the trajectory of the tropical storm Stan and how it changes from a tropical depression to a hurricane (INSIVUMEH, 2005). In red, location of the study area.

Rainfall from this storm was exceptional because it was geographically widespread and continued over a period of about 10 days. The cumulated rainfall in the region registered by the National Meteorological Service (INSIVUMEH) at the weather station located in San Marcos was 362.5 mm (378 mm for all October 2005). This prolonged precipitation event occurred at the end of the rainy season when the ground had already a high moisture content. This type of rainfall, on saturated or nearly saturated ground, has the capability to trigger landslides over a large area (INSIVUMEH, 2005)

During the 2010 rainy season in Guatemala, torrential rainfall from the tropical storm “Agatha”, which is considered as one of the worst after “Stan” by governmental authorities, caused severe impact in thousands of families across the country. Tropical Storm “Agatha”, the first Pacific storm of the 2010 season struck Guatemala from 25 to 30 May and the cumulated rainfall registered by the INSIVUMEH weather Station in San Marcos was 290.8 mm. Official figures report 174 deaths, 13 missing people and 154 injured people. The Executive Secretariat of the CONRED registered 397,808 affected people and an additional 133,102 people at risk. Agatha’s heavy rains caused widespread flooding and mudslides in Guatemala (IFRC, 2012; INSIVUMEH, 2010).

In 2011 at the end of the rainy season, between October 10 and October 20, a tropical depression named “Twelve-E” dumped rainfall that surpassed the average for that period. The cumulated rainfall registered by the INSIVUMEH at weather Station in San Marcos was 317 mm during the event (413 mm for all October 2011). Heavy rainfall caused flooding and landslides in Guatemala as the soil was already saturated and river levels were high as a result of previous rainfall (Figure 7). Extensive flooding and multiple landslides affected more than 500,000 persons, of which 82,913 were severely affected and 39 people died. In the housing sector, authorities reported 16,609 homes with minor damage, 5,571 with moderate damage, 753 with severe damage and 5,291 homes still at risk were reported. In addition, the national road network was severely damaged, with estimates indicating 167 sections of roads affected and 9 destroyed, while 55 bridges were damaged and 30 destroyed. Given the situation, the Guatemalan government decreed a State of Public Calamity on October 16 (IFRC, 2013; INSIVUMEH, 2011).

Figure 7 — Pictures of landslides in the study area A) Crops affected by landslides B) Landslide in forest one month after “Tropical Depression 12-E” (© Estrada, November 2011).

3.3 Seismic events

Guatemala is located in a complex tectonic setting composed by three tectonic plates: the Caribbean plate (including southern Guatemala), the Cocos plate to the South-West, and the North American plate to the North (including northern Guatemala) (Figure 8).

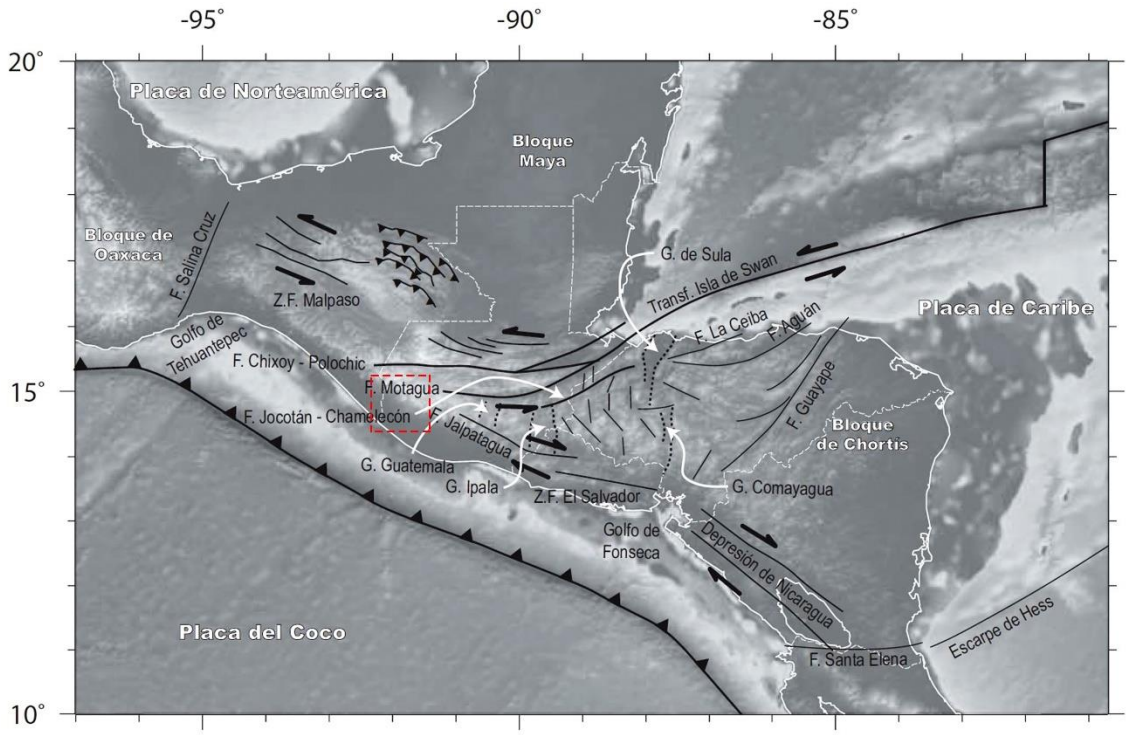


Figure 8 — Tectonic Northern frame (Alvarez-Gomez, 2009). In red, location of the study area.

The subduction is a well-defined zone, between the Cocos, Caribbean and North America plates, convergence rate being estimated at ~7 cm/yr. (Ligorría *et al*., 1995). This zone belongs to the Middle America Subduction Zone, where a triple junction is supposed to be located. The Middle America Subduction Zone represents the source of most of the largest historical earthquakes in Central America (Ligorría *et al*., 1995; Villagran *et al*., 1996).

The study area is situated in the western region of the subduction zone in Guatemala and the state of Chiapas Mexico having a long history of seismic activity (Ligorría et al., 1995; Villagran et al., 1996; Monterroso, 2003). In Guatemala, 70 % of the seismic events occur in relation with the subduction zone, 20 % along the Altiplano faults and 10 % are related to the transform Chixoy-Polochic-Motagua faults (INSIVUMEH, 2006) (Figure 9).

A destructive earthquake occurred in Guatemala on February 4, 1976, with surface wave magnitude M_S 7.5) centered close to Honduras border, some 300 km to the East of my study area. The rupture zone followed the Motagua fault through eastern and central Guatemala and then branched sharply southward about 35 km west of Guatemala City (Matumoto &

Latham, 1976; Husid, 1978). San Marcos Department was affected along with Chimaltenango, Chiquimula, El Progreso, Guatemala, Huehuetenango, Izabal, Sacatepéquez, and Sololá.

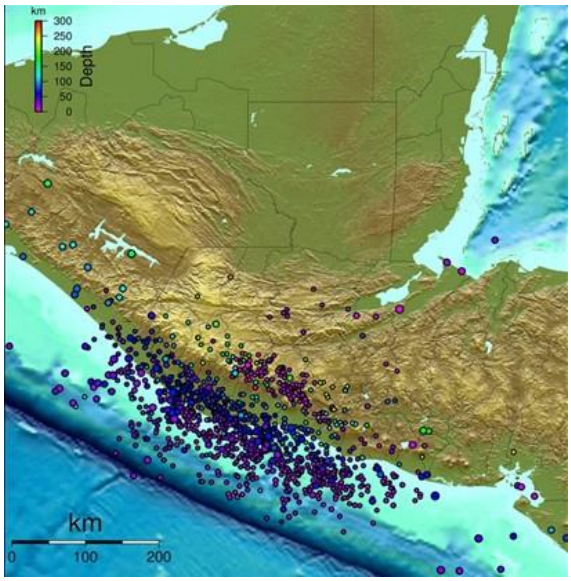


Figure 9 — Seismicity (M>4) in Guatemala among 1984 to 2014 (Yani-Quiyuch, 2015).

On November 7, 2012 at about 10:35 am (local time), a magnitude M_w 7.4 earthquake occurred south of Champerico, SE Guatemala. The location of this event was estimated from the records of more than 700 stations of the Global Seismograph Network. The depth of the hypocenter was estimated at 24.1 km, and the rupture lasted about 32 seconds. The extensive damages in San Marcos, which is ~130 km away from the epicenter, were likely due to the propagation of waves through a nearby volcano, which amplified the waves, or to site effects. This zone is characterized by the presence of large volcanic structures, such as calderas and stratified volcanoes (EERI, 2013). In relation to this event, landslides in the mountains and on steep slopes were officially reported by CONRED along the Pan American Highway CA-1 crossing the study area. However, we noticed two aggravating circumstances. First, the poor design of the slope cuts made along the Pan American Highway and other roads triggered many of the slides. Second, slides along the roads were also caused by the failure of hillsides previously weakened by Tropical Storm “Agatha” in 2010. The municipality of San Cristobal Cucho, in San Marcos Department, reported seven casualties in a family that was buried by a landslide. The total number of landslides linked to this event was not as big as during the 1976 earthquake, but most landslides were aggravated by human-related activities (EERI, 2013).

4 MATERIALS AND METHODS

Thematic maps, DEM and orthoimages were provided by the National Geographic Institute and the Direction of geographic information, strategic and risk management of the Ministry of Agriculture, Livestock and food in Guatemala. The preparation procedure for each data layer is summarized below, based on literature review.

4.1 Landslide inventories: data source and data types:

"For this research, orthoimages with a scale of 1:5,000 were used to perform the restitution of landslides for 2005. Field survey was also conducted to obtain data on 2011 rainfall events. Landslide age is defined as the year of landslide occurrence.

The orthoimages used for the landslide inventory is the product of the project "Obtaining digital images at detail scale of the Republic of Guatemala" developed by the Ministry of Agriculture, livestock and Food in Guatemala. The orthoimages were taken from 01 November 2005 to 28 February 2006. Tropical storm "Stan" rainfall was prolonged over a period of about 10 days in October 2005 triggering landslides in the whole study area. For this reason, it is very likely that the orthoimages contained the landslides triggered by storm "Stan". Moreover, farmers in the area were asked to confirm the dates of the landslide events during the field survey. The characteristics of orthoimages are reported in Table 3.

Table 3 — Characteristics orthoimages used in landslide inventory.

Item	Description
Geodesic coordinate systems	WGS84.
Altimetry origin	Puerto de San Jose 1949 / 50
Projected coordinate system	Guatemala Transverse Mercator –GTM-, this is an adaptation of the UTM projection to avoid working on 2 different zones.
Guatemala Transverse Mercator parameters	-Central Meridian: -90° 30' -K = 0.9998 (scale Factor in Central Meridian) -False east: 500,000 m -False North: 0 m
Photo scale	1:5,000
Pixel size	0.5 meters.
Color depth	-Photo color: 24 bits Equivalent to 8 bits for each channel RGB. -Photo false color: 24 bits equivalent to 8 for each channel IRG. Spectral bands Blue, Green, Red Near Infra-Red
Tiles used in this study	18601_01_ORT_RGB, 18601_06_ORT_RGB, 18601_11_ORT_RGB, 18604_05_ORT_RGB, 18604_10_ORT_RGB, 18604_15_ORT_RGB.

For the purpose of this research, the landslides were initially digitized as polygon features. Then these landslide polygon features were later transformed and sampled as a single pixel (Van Den

Eeckhaut, 2006; Hussin *et al*., 2016; Simon *et al.,* 2017). One single pixel was selected to represent the top-point of one landslide on the initiation area.

The second source of information used to produce the landslide inventory was obtained by undertaking fieldwork after the landslide events that were triggered by the intense rainfall in Guatemala linked with the tropical depression Twelve-E in October 2011. The coordinates at the head of each slide were taken using Garmin Oregon 550.

Focus was given on recent landsides. It was very difficult to collect reliable quantitative data about these landslides because erosion processes and plowing made them less visible? Moreover, the local farmers could not give much information about these landslides during the interviews. Due to limited information about landslides features collected during the field survey, all were classified as shallow landslides. It was not possible to establish the presence of debris flow owing to the quickly growing vegetation.

Landslide classification for this inventory followed the definition updated by Hungr *et al*., (2014) based on Cruden and Varnes (1996), Turner *&* Schuster (1996) and Varnes (1978) for debris flows and shallow landslides, see state of art for more details.

4.2 Predictor variables

Twelve potential predictor variables used as independent variables in the modelling were included: Factors derived from DEM such as elevation, slope angle, slope aspect, plan curvature, profile curvature and contributing drainage area (CDA). Thematic attributes such as lithology, land cover, and estimated variables like Euclidean distance to drainage, Euclidean distance to road, landscape fragmentation and geomorphic zones (region) were derived from specific thematic maps. The characteristics of each data layer is presented in Table 4.

Table 4 — Characteristic of data layers used for the modelling.

Type	Scale	Date	Source
Orthoimages	1:5000	2005	Ministry of Agriculture, livestock and Food in Guatemala –MAGA-.
DEM (elevation, slope angle, slope aspect, planform, profile and (CDA).	1:10000	2005	National Geographic Institute -IGN-.
Lithology	1:250000	2002	Geological map, National Geographic Institute -IGN-.
Land cover	1:5000	2005	Orthoimages –MAGA-.
Euclidean distance to drainage	1:10000	2005	DEM, National Geographic Institute -IGN-.
Euclidean distance to road	1:250000	2002	National Geographic Institute -IGN-. And orthoimages –MAGA-
Landscape fragmentation	1:5000	2005	Land cover, Orthoimages –MAGA-.
Geomorphic zones	1:10000	2013	Field visit, digitized on DEM –IGN-.

Digital Elevation model (DEM)

According to National Geographic Institute -IGN- acronym in Spanish, DEM given for this research was derived from aerial images taken in 2005, using surface analysis tool Match-T software. The photogrammetric flying integrated GPS/INS system with mean square error for planimetric <0.20 m., <0.25 m. for altimetry, and < 0.004° for angles direction, the grid resolution is 15 X 15 meters, scale 1:10:000. The resolution and the intrinsic errors related to DEM have an impact on the accuracy and reliability of the predictor variables (Dewitte *et al.,* 2015), the errors of DEM are inherent to the method used by the National Geographic Institute.

Most DEMs contain numerous topographic depressions which are defined as areas without an outlet and often referred to sinks or pits, which represent topographic errors. Most depression filling algorithms are based on the 1-D single flow direction. It is used in this research previous to derived topographic factors (Zhu *et al*., 2013). This is one of the widely used algorithms that has been implemented in many GIS and hydrological softwares. DEM was used to derive elevation, slope, aspect, profile curvature, plan curvature and contributing drainage area.

Elevation

The elevation or altitude is considered as a factor influencing the climate conditions (spatial variation of precipitation, temperature solar irradiation and others). Elevation can also be linked to lithological properties and therefore can be used as a proxy. The surface of the study area is represented as a DEM (raster file) that contains the elevation value field (Figure 10). A defined interval of 100 meters is used to determinate the threshold values between elevation classes (Dai & Lee, 2002; Bui *et al*., 2011; Wu *et al*., 2016).

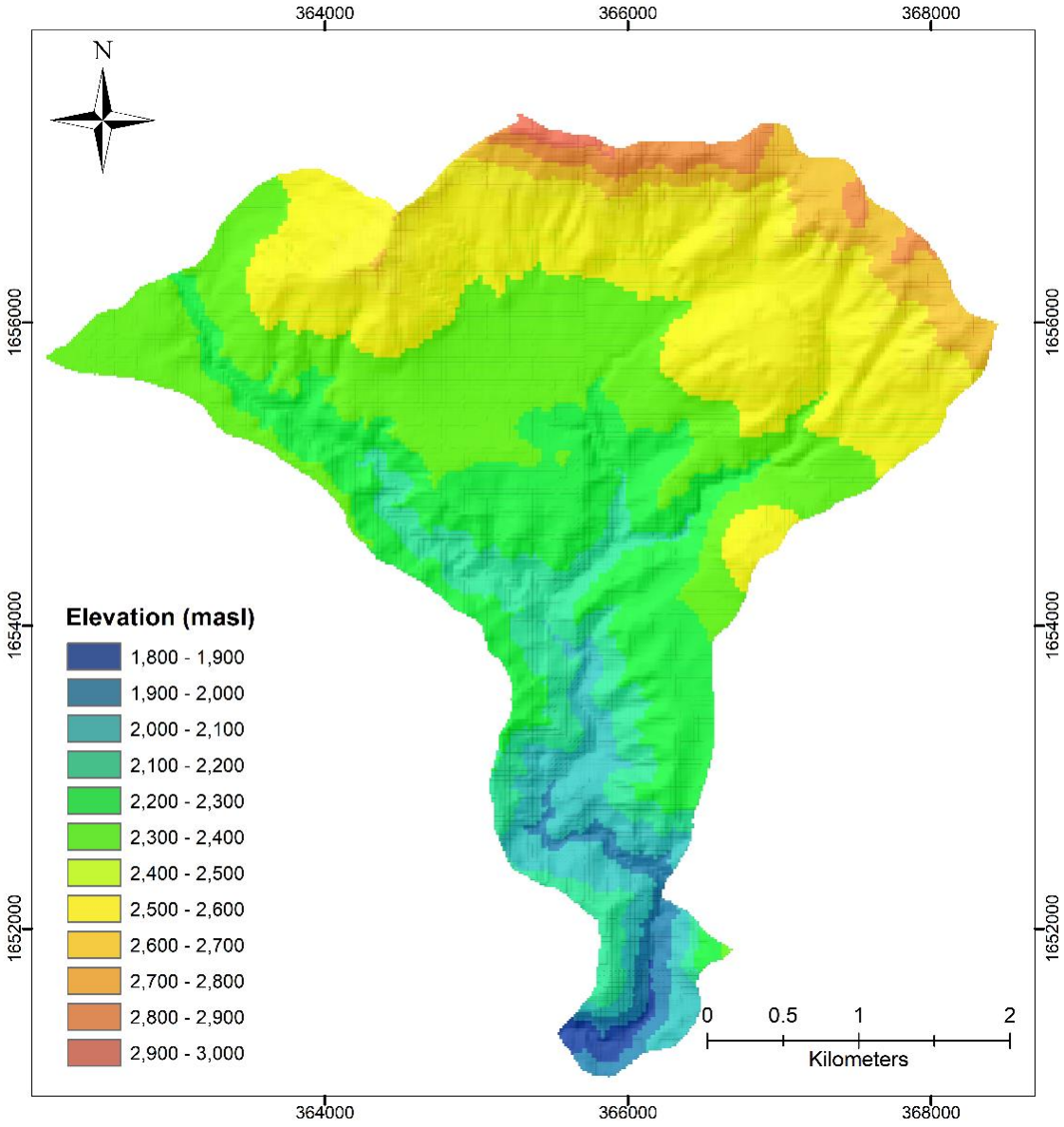


Figure 10 — Elevation map (meters above sea level).

Slope

Slope gradient is defined as the maximum rate of change in altitude and is a key factor in the stability control. There are many ways by which land may fail, depending on the slope angle, the water content, the type of material involved and local environmental factors (Bui *et al*., 2011; Ercanoglu *et al*., 2004; Dai & Lee, 2002). The slope gradient for this study is expressed in percent. For threshold in the map and frequency ratio analysis was reclassified in five classes, 0-25 %, 25-48 %, 48-72 %, 72-102 % and 102-285 % (Figure 11). To determinate the threshold values between classes, natural breaks (Jenks) algorithm was used, this is form of variance-minimization classification, brakes are typically uneven, and are selected to separate values where large changes in value occur (de Smith *et al*., 2007).

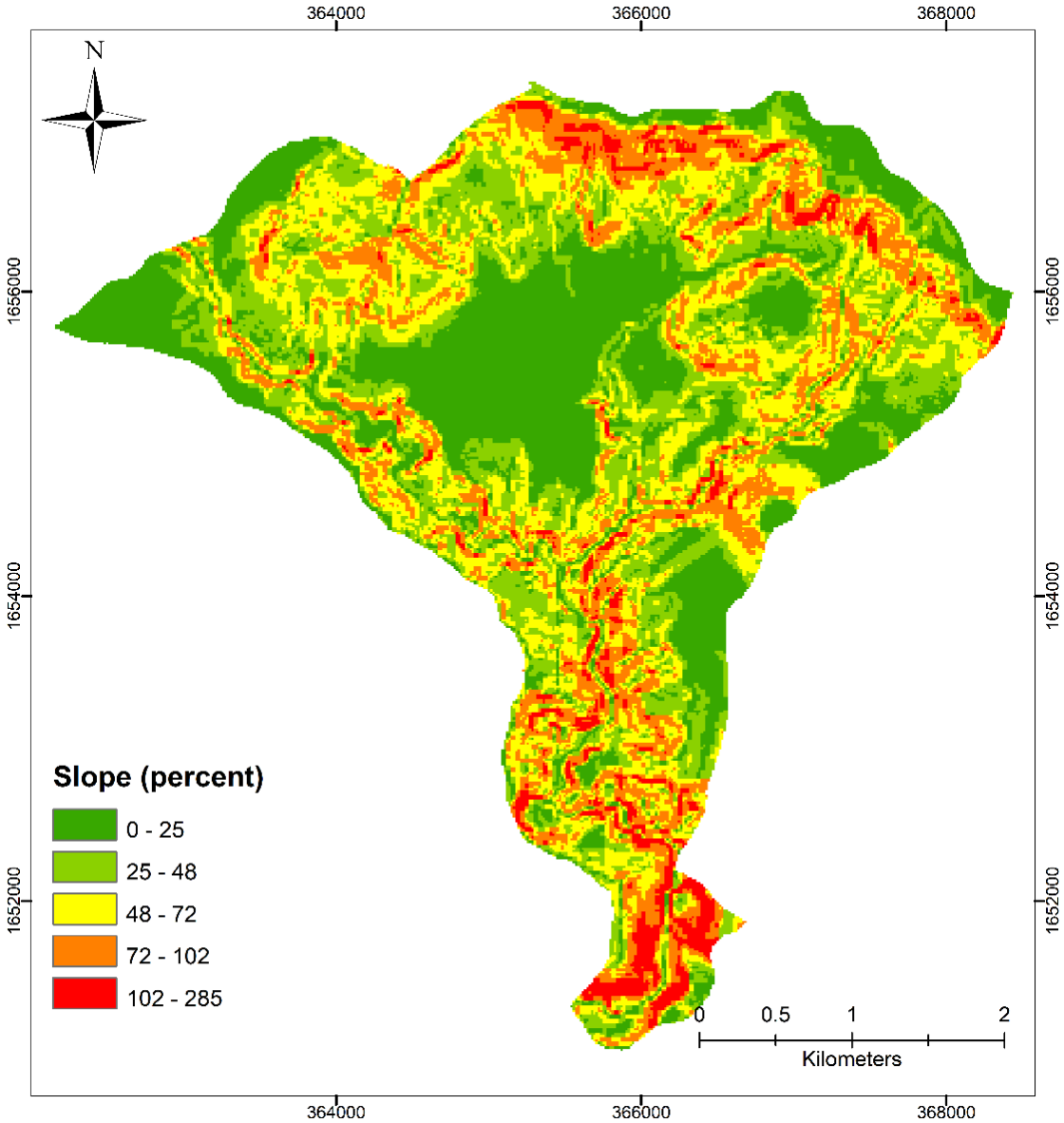


Figure 11 — Slope map.

Slope aspect

The slope aspect is the bearing or azimuth of a maximum slope gradient. In some circumstances, the direction that a slope faces may have an influence on the stability of the slope. Moisture retention and vegetation is reflected by slope aspect, which in turn may affect soil strength and susceptibility to landslides. If rainfall has a pronounced directional component influenced by prevailing winds, the amount of rainfall falling on a slope may vary depending on its aspect (Ohlmacher & Davis, 2003; Dai & Lee, 2002). The slope aspect was computed and expressed in degree from cartographic north in a clockwise direction, ranking from 0° to 360°. The aspect computed, fits a plane to the z-values of a 3 x 3 DEM cell neighborhood around the processing or center cell, using 8-point formulas of Horn (de Smith *et al*., 2007). The aspect factor for this study was divided into eight aspect categories on the basis of the facing direction with respect to the eight main cardinal directions, shown in Figure 12.

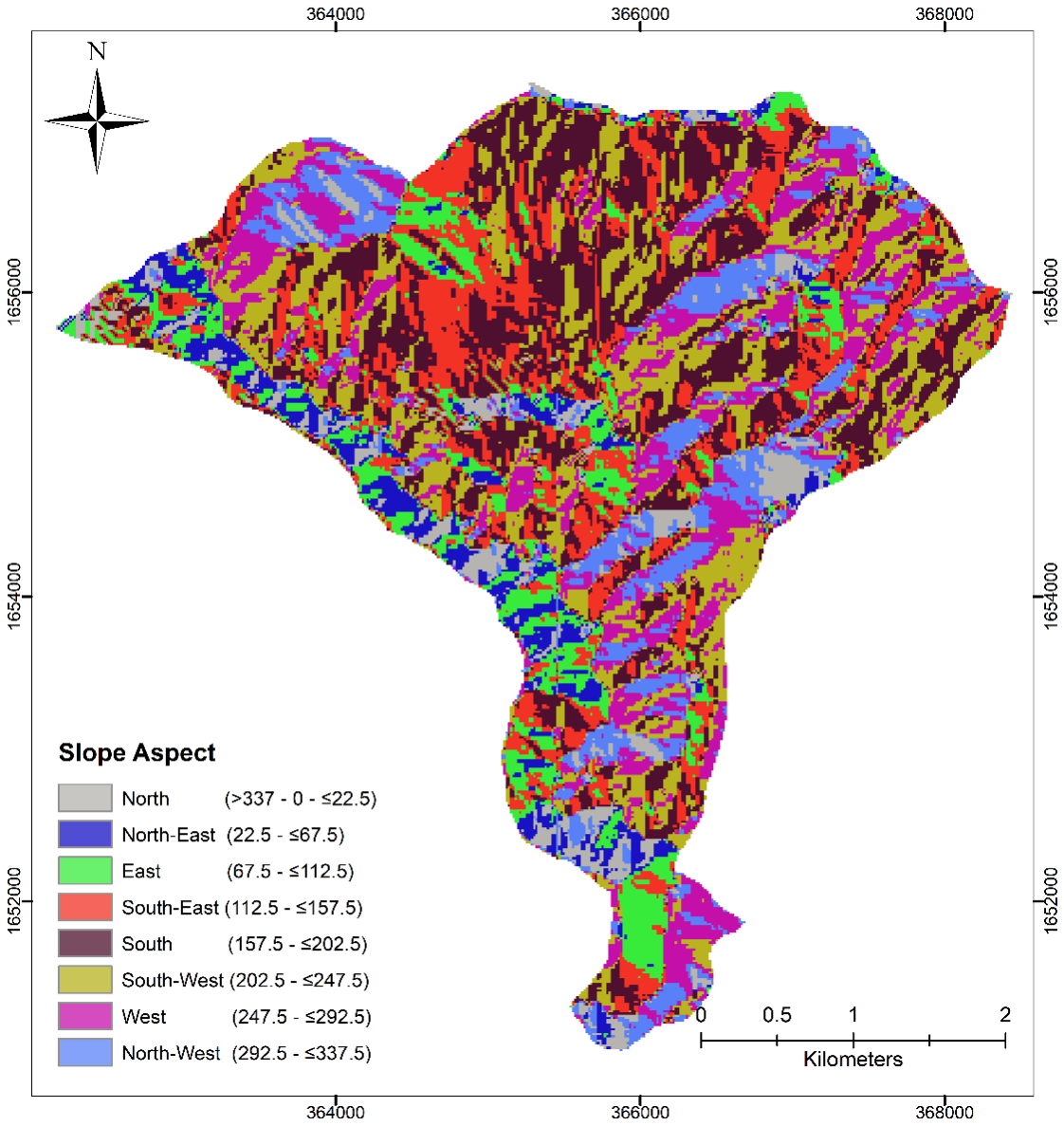


Figure 12 — Slope aspect map.

Curvature

Curvature is the second derivative of a surface, or the measure of the spatial variation of the slope. Curvature is defined as the change in slope angle along a very small arc of the curve. A plane has a null curvature and as the bending becomes more acute the curvature value increases. In a hillslope an infinite number of curvature values can exist, although this characteristic is often used to recognize the gullies. In this study, two curvature components were used: profile curvature and planform curvature (Bui *et al*., 2011; Ohlmacher, 2007; Horton *et al*., 2008; Catani *et al*., 2013; Iovine *et al*., 2013).

Profile curvature

Profile curvature is parallel to the direction of the maximum slope. Negative value means that the surface is upwardly convex at that cell. Positive profile means that the surface is upwardly concave. Zero value indicates that the surface is linear. Profile curvature affects the acceleration or deceleration of flow along the maximum slope direction on the surface. The Profile curvature for frequency ratio analysis was reclassified in three classes: upward (convex), straight (r) and upward (concave), depending to their values profile curvature <-0.02, -0.02 ≤ profile curvature ≥ 0.02; 0.02 ≥ profile curvature, respectively (Figure 13). (Ohlmacher, 2007; Catani *et al*., 2013; Iovine *et al*., 2013; Dewitte *et al*., 2015). In the case of profile curvature, the inherent errors of DEM impact significantly the accuracy of this variable showing some patterns that should be considered in the evaluation criteria of the model.

Planform curvature

Commonly named as plan curvature, planform curvature is the 2D curvature of the intersection of the surface with a horizontal plane. Positive value means the surface is sidewardly convex at that cell. Negative plan means the surface is sidewardly concave at that cell. Zero value means the surface is linear. The influence of plan curvature is the convergence or divergence of water during downhill flow. The plan curvature for frequency ratio analysis was reclassified in three classes: divergent (convex), parallel (r) and convergent (concave), depending to their values (planform curvature < -0.02, -0.02 ≤ planform curvature ≥ 0.02; 0.02 ≥ planform curvature), respectively (Figure 14) (Ohlmacher, 2007; Nefeslioglu *et al*., 2008; Catani *et al*., 2013; Iovine *et al*., 2013).

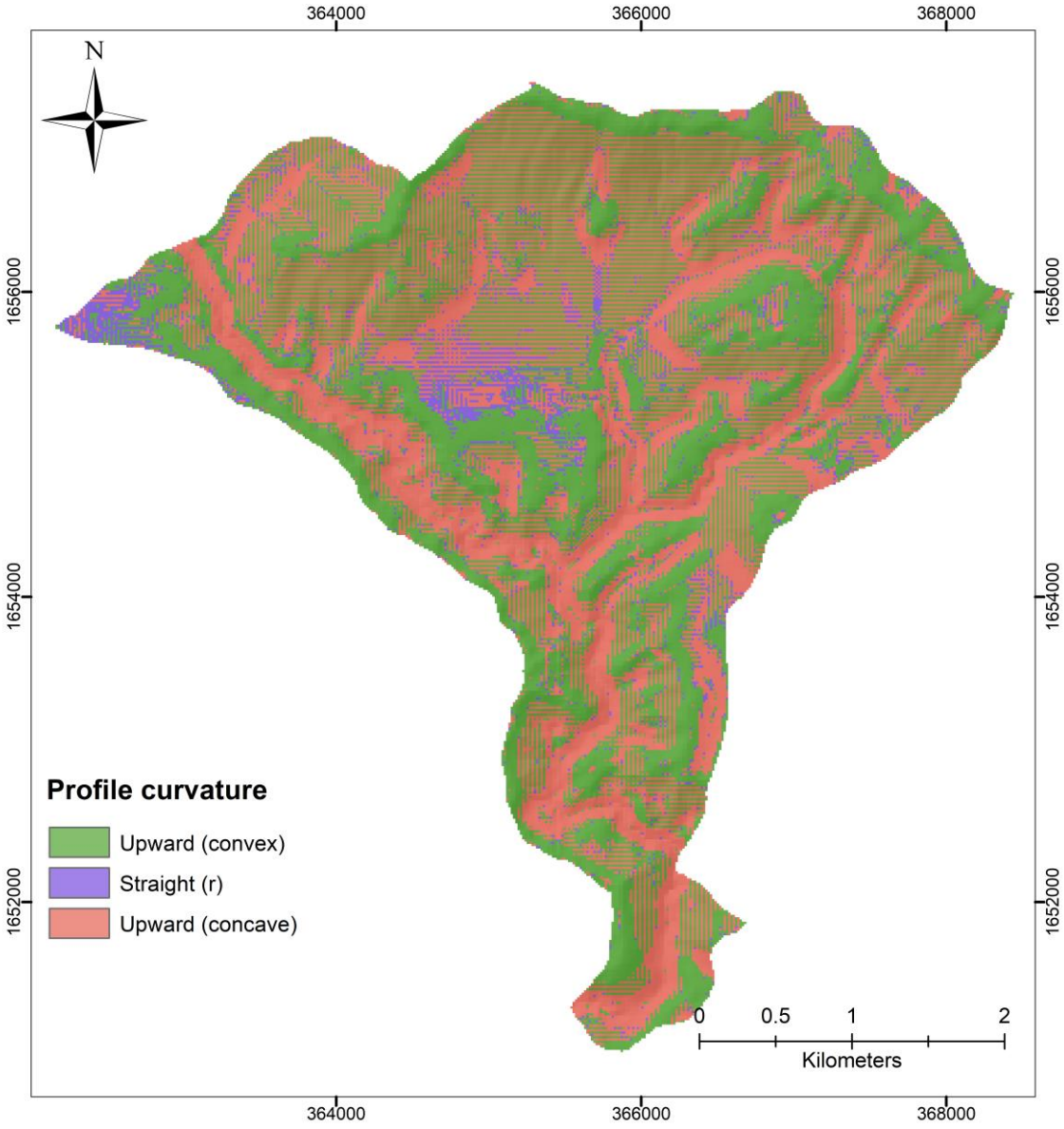


Figure 13 — Profile curvature map.

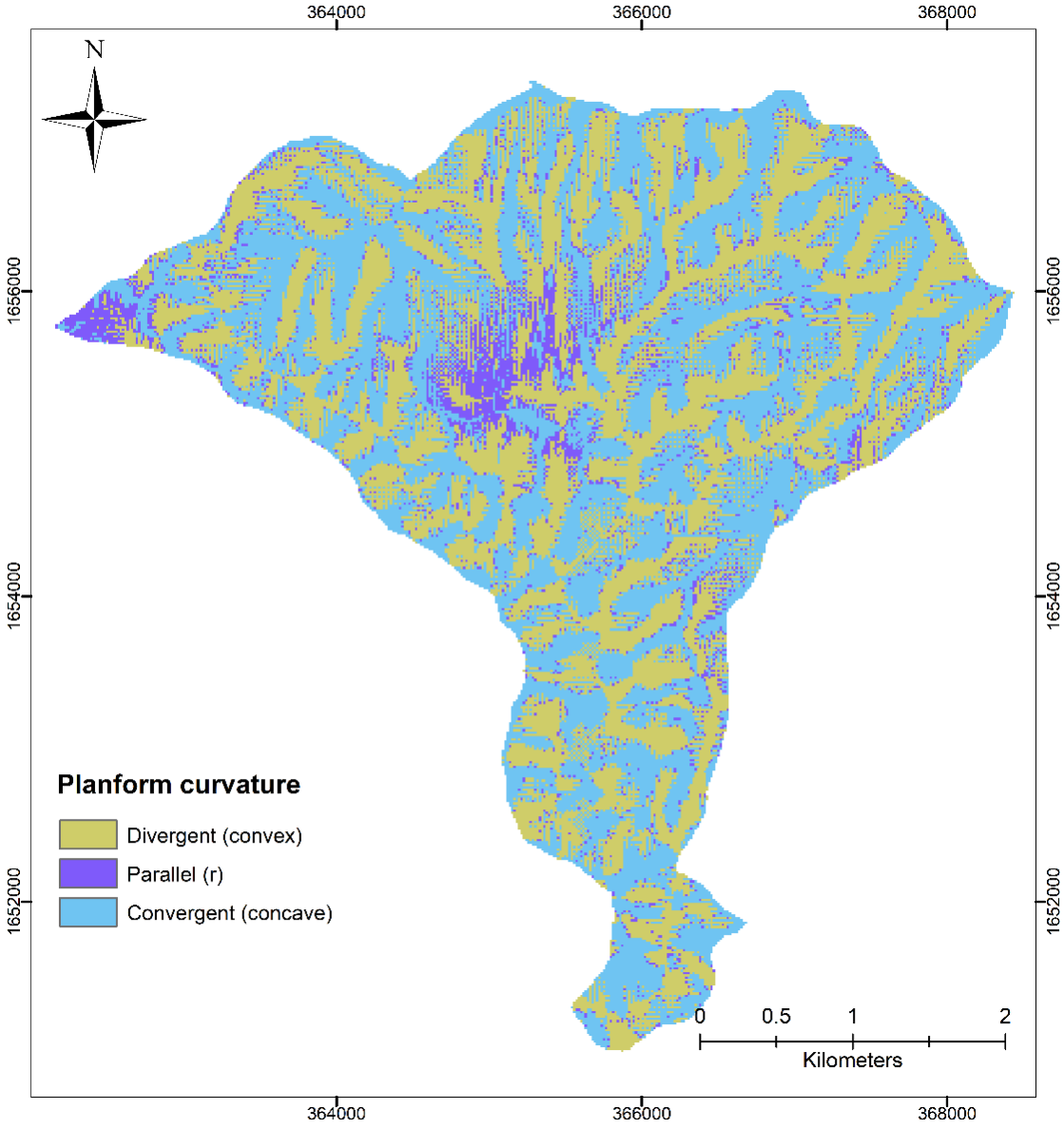


Figure 14 — Planform curvature map.

Contributing drainage area

Once the pits or depressions in DEM had been removed, to ensure that flows be continuous across the surface, it was implemented with the D8-algorithm which involves a 3x3 windows of cells for the locally steepest direction and then coding each cell to indicate which one direction (of the 8 available) the flow is to follow (de Smith *et al.*, 2007). A weight of 1 is applied to each cell and the contributing drainage area is calculated as the cumulated weight of all cells flowing into each downslope cell in the output raster. To reclassify this variable into classes for frequency ratio analysis, eight classes were constructed, 0-2; 2-4; 4-8; 8-16; 16-32; 32-64; 64-128 and larger to 128 cells contributing (Figure 15).

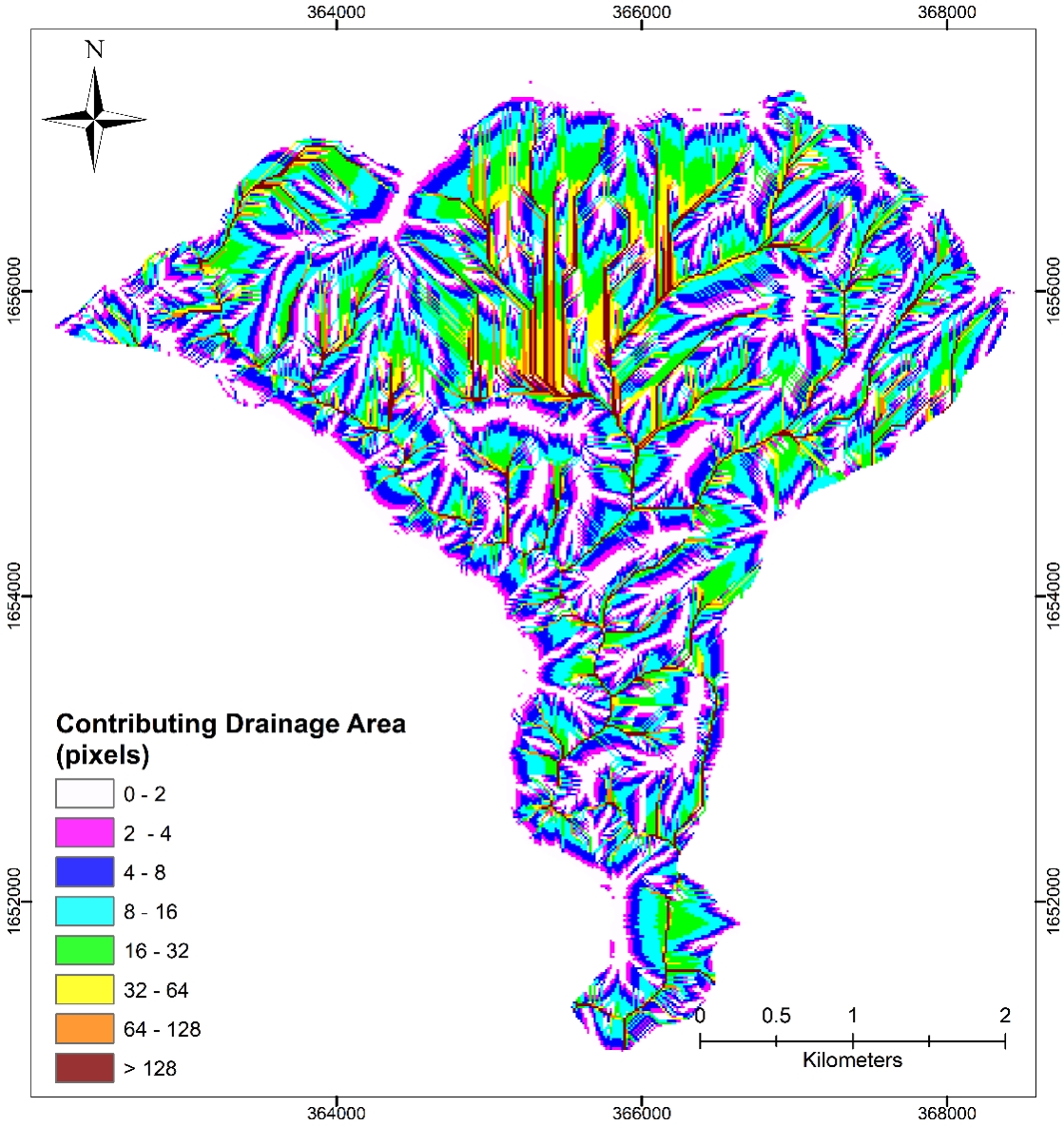


Figure 15 — Contributing drainage area map (cells).

Lithology

It is expected that the properties of the slope-forming materials, such as strength and permeability can be involved in the failure. It directly reflects the geomechanical and hydraulic properties of the bedrock and also influences the characteristics of the soil coverage (Costanzo *et al*., 2013; Catani *et al*., 2013). According to the geological map, the types are volcanic rocks and pumice ash landfills (Figure 16).

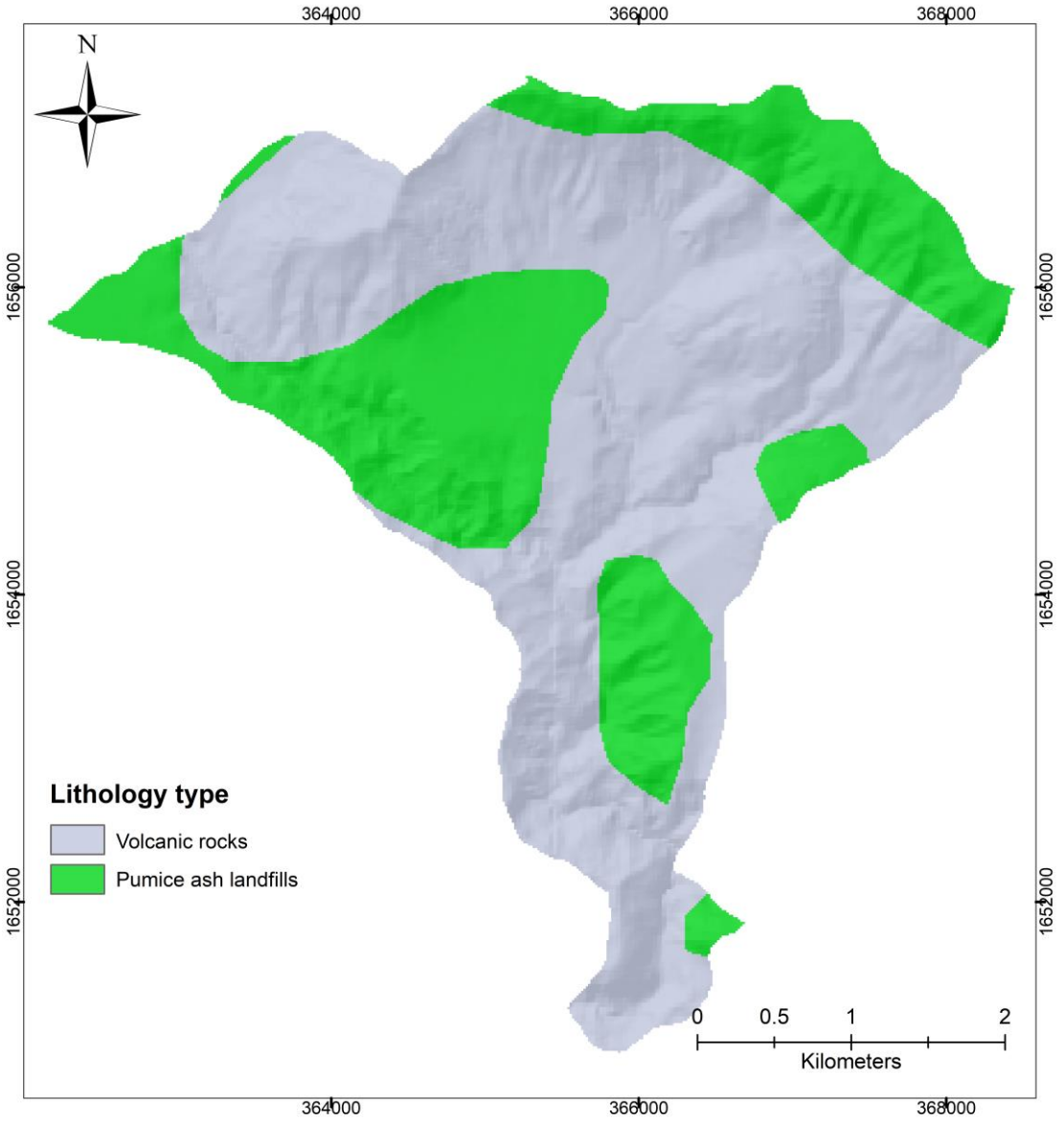


Figure 16 — Lithology map.

Land cover type

Vegetation is an important factor influencing the occurrence and movement of rainfall-triggered landslides and particularly important for shallow landslide susceptibility. Land cover changes have been recognized throughout the world as one of the most important factors influencing the occurrence of landslides, various researchers emphasized the importance of vegetation or land use characteristics on the stability of slopes (Glade, 2003; Horton *et al*., 2008; Pourghasemi *et al*., 2012; Trigila *et al.,* 2015). The forest cover has a positive effect on the slope stability. The most important effect is the extra cohesion of the soil given by the roots increasing soil shear strength (Broothaerts *et al.,* 2012). Thus, we can expect that vegetation protects slopes against soil erosion, surface runoff and shallow landslides (Trigila *et al.,* 2015; Persichillo *et al.,* 2017).

Land cover information of study area was digitalized from 2006 orthoimages and was classified as follows: forest, urban areas, cropland, pasture and roads (Figure 17).

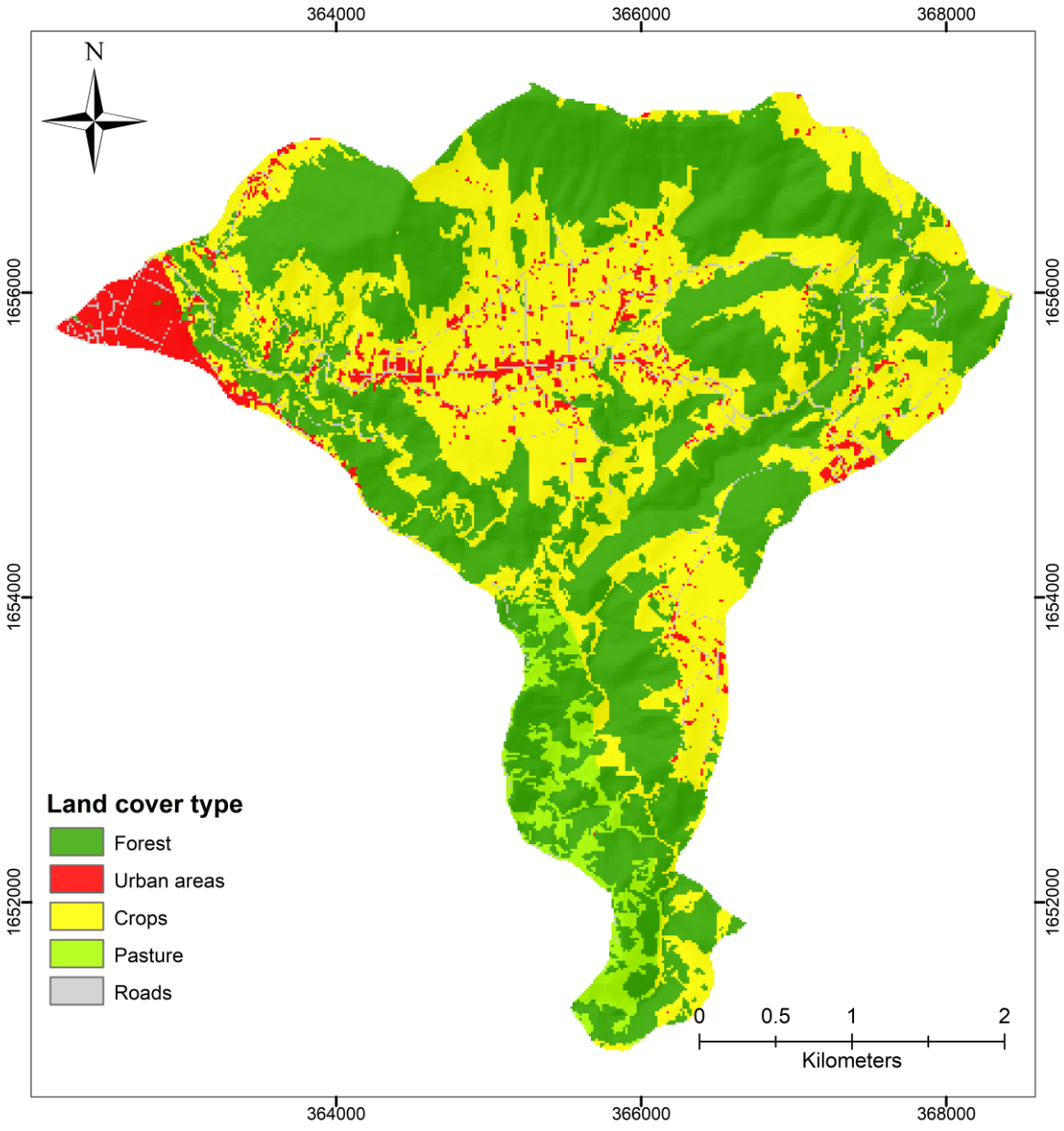


Figure 17 — Land cover map.

Landscape fragmentation

Landscape metrics have been used to assess land cover change. Landscape metrics characterize landscape patterns, such as the number, area, composition, configuration, and connectivity of various patch types (Lin *et al.,* 2010). Landscape fragmentation is used to define the process whereby a large, continuous area one land cover is both reduced in area and divided into two or more fragments (Alcantara-Ayala *et a*l., 2006). In this research it refers to the characteristics associated with the abundance of patch types (number of patches) within a given landscape as a predisposing factor to landslides occurrences (Lin *et al.,* 2010).

The analysis considered a quadratic sliding window on the study area, proposed by Alcantara-Ayala *et al*. (2006). For this research the quadratic windows is 30 x 30 meters overlapping the layer of land cover, thereby the number of land cover patch inside the windows are summed and indicate the degree of fragmentation. Lower value of number patches in the quadratic window indicates that landscape is not fragmented. It is expected that the disturbance produced by various land cover patch types contribute to the landslides occurrences (Figure 18).

The disturbance due to the rapid land cover change is unprecedented in human history especially in developing tropical and subtropical countries. Guatemalan landscapes are continuously changing, threatening global sustainability and livelihood systems. Research concerning landscape fragmentation has included land cover classifications as well as comparisons with vegetation indices. (Jaeger, 2000; Gu *et al*., 2002; Serrano *et al*., 2002; Nagendra *et al*., 2004; Southworth *et al*., 2004; Alcantara-Ayala *et al*., 2006; Baldi *et al*., 2006; Kamusoko & Aniya, 2007; Ming *et al*., 2008; Girvetz *et al*., 2008; Hobbs *et al.*, 2008; Fu *et al*., 2010; Hernandez, 2012; Tomaselli *et al*., 2012).

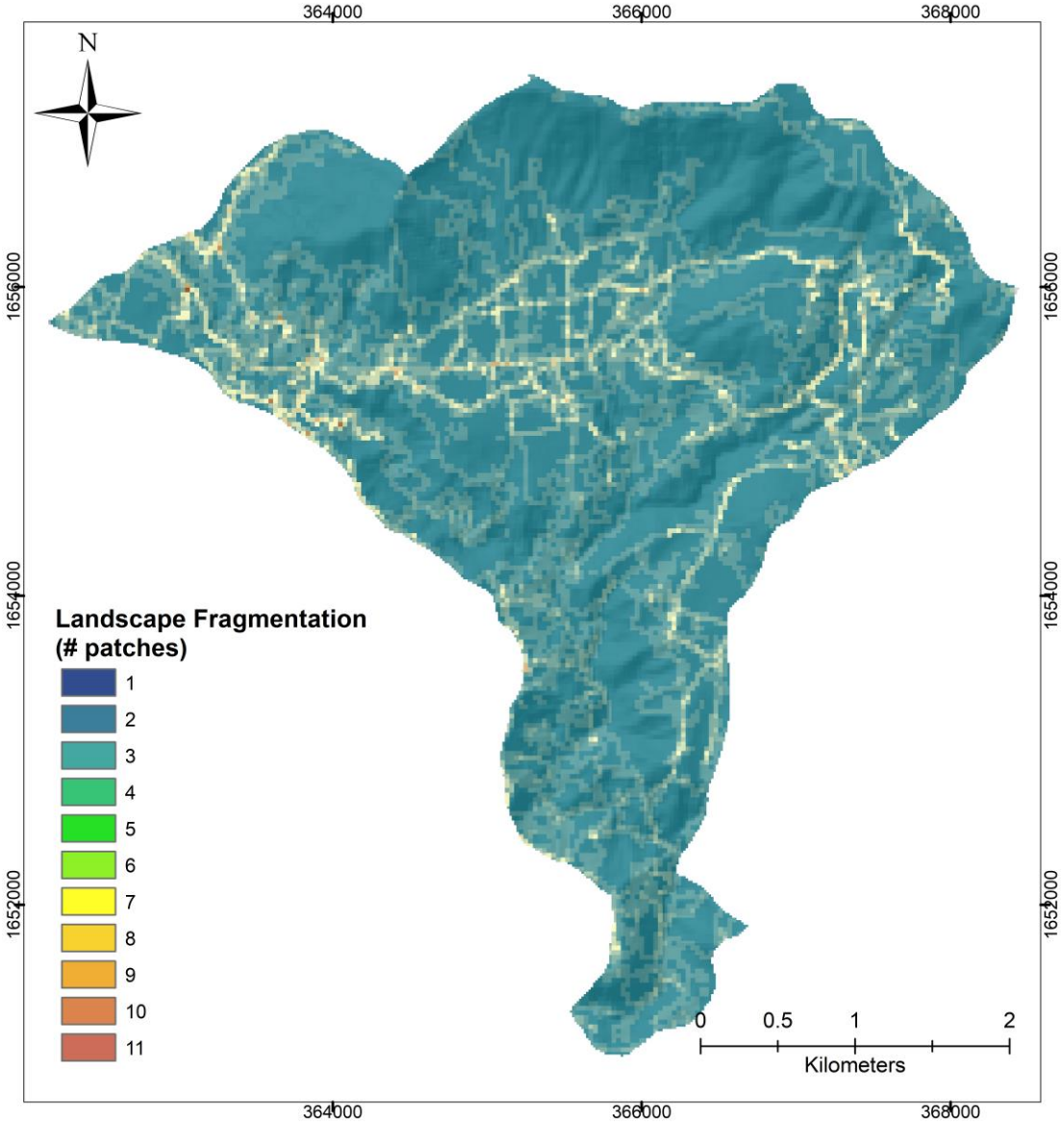


Figure 18 — Landscape fragmentation map.

Euclidean distance to drainage

The proximity to drainage due to vertical incision of valleys, groundwater table, river bank erosion, and slope erosion is an important factor for the occurrence of landslide. Often landslides in hills occur due to the erosional activity associated with drainage (Bui *et al*., 2011; Park *et al*.,2013; Wu *et al*.,2016; Dai & Lee,2002; Kannan *et al*.,2013). Using the stream network constructed on the basis of flow accumulation grid, the distance to drainage was quantified by the Euclidean distance between every pixel and the stream channel network. To reclassify into classes this variable for frequency ratio analysis, geometrical interval was used by minimizing the sum of squares of the number of elements in each class, this method was designed to work on data that are heavily skewed by a preponderance of duplicate values (de Smith *et al*., 2007). The five classes are as follow: 0-58; 58-86; 86-144; 144-269; 269-541 meters. (Figure 19).

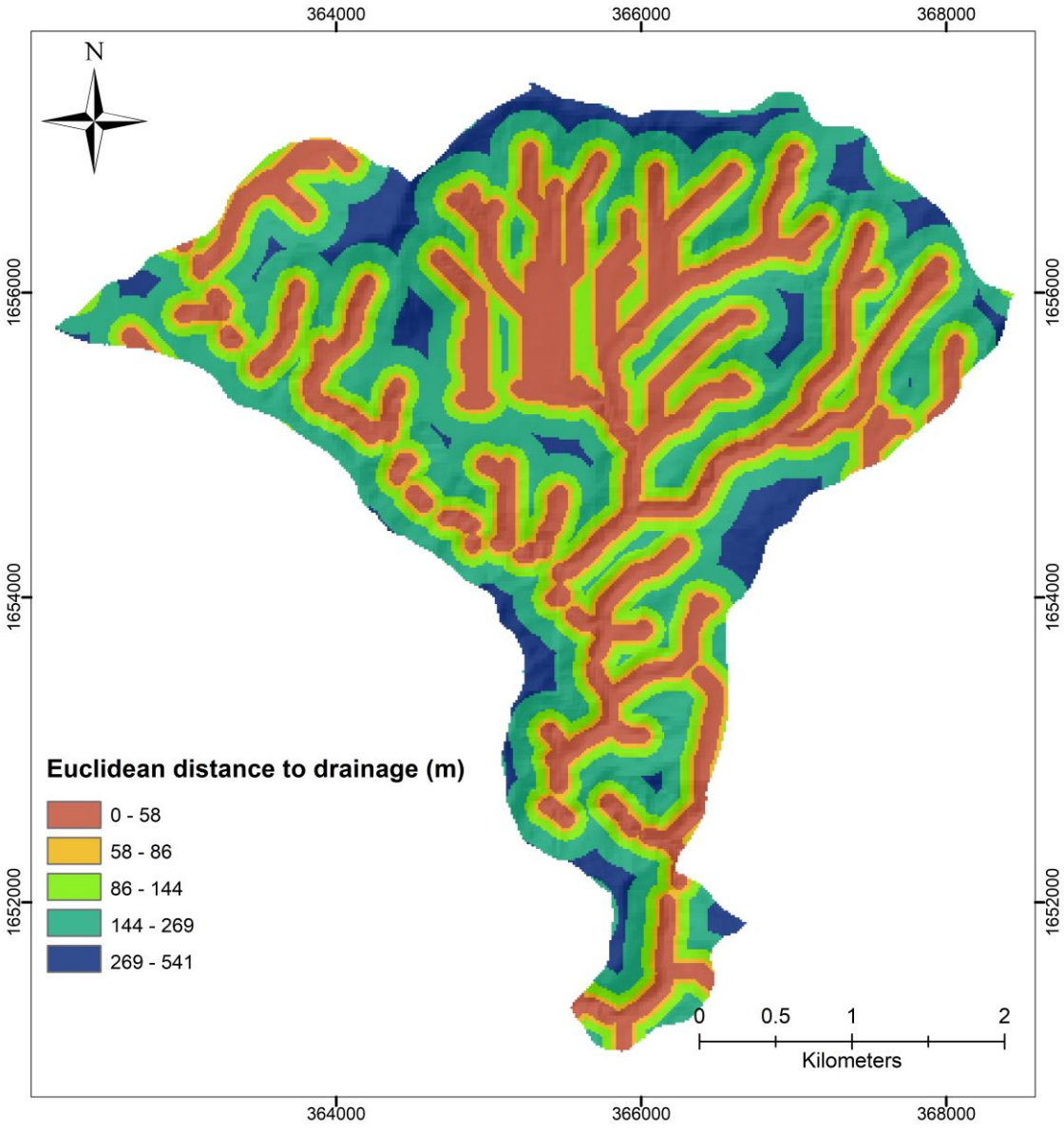


Figure 19 — Euclidean distance to drainage network map (meters).

Euclidean distance to road

The 2002 road network map of Guatemala a scale of 1:250000 was updated with the data obtained from the 2006 orthoimages. The material of the roads (asphalt and dirt roads) were taken into account by tracing a continuous line. The value of this factor was quantified by computing the Euclidean distance from the road network to every pixel. In the same way as distance to drainage, geometrical interval was used to reclassify into classes (de Smith *et al*., 2007). The five classes are: 0-19; 19-58; 58-139; 139-304; 304-645 meters (Figure 20).

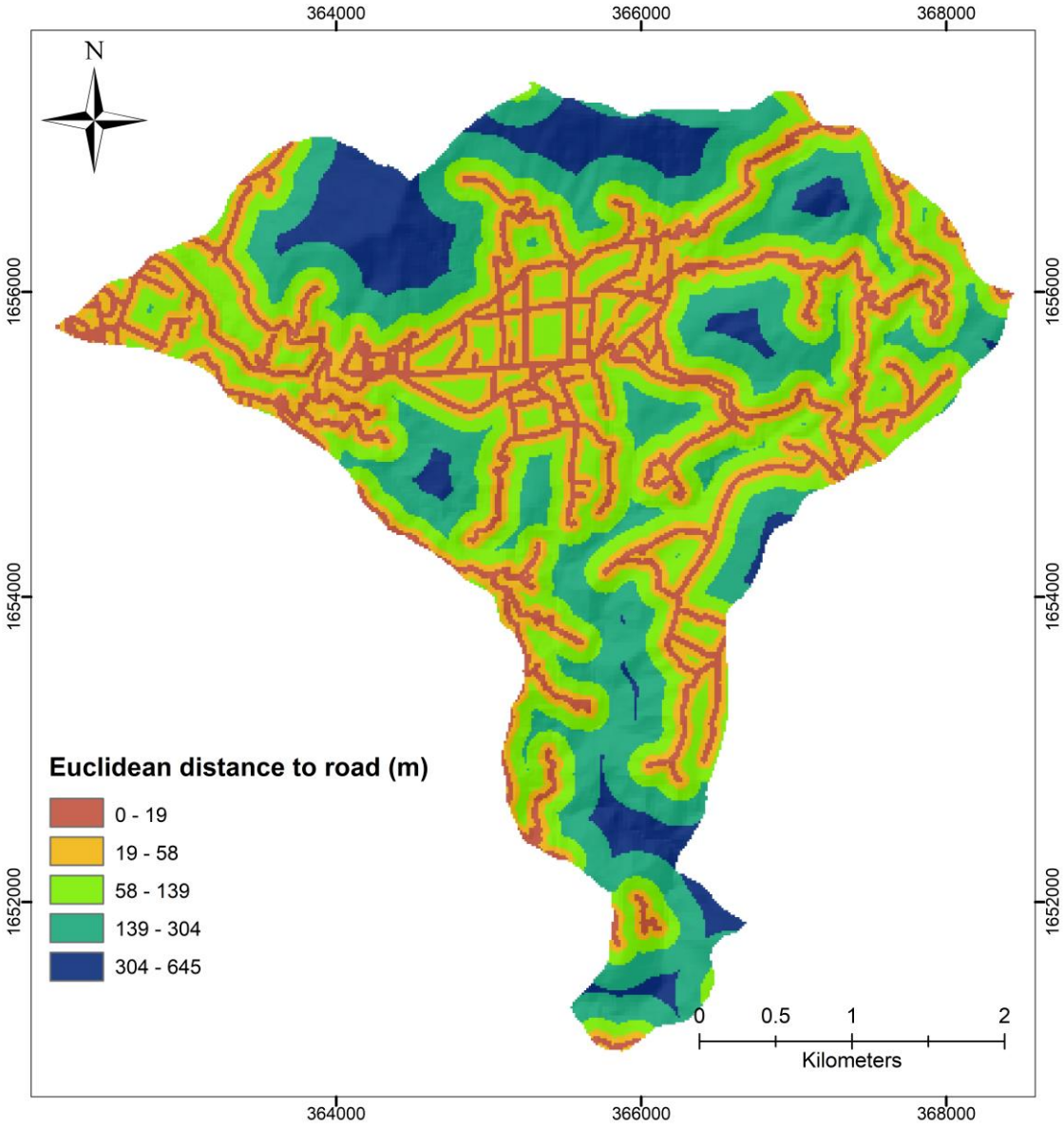


Figure 20 — Euclidean distance to roads map (meters).

Morphometric regions

The type and weighting of each predisposing factor to trigger landslides depends on the environmental setting (e.g. geomorphological evolution and processes) (Comominas *et al.,* 2014). A characteristic of the geomorphology of the micro-watershed is its dramatic steep and broken terrain. On the terrain, three different morphometric region types can be identified: the flat altiplano region, upslope can be observed an "old" topography, and downslope as more recent topography due to river incision into the plateau. So far, the knickpoints do not seem to have reached the old topography. Therefore, for equal slopes, the reply to landsliding may be different in these two regions. We could expect more landslides on the recent slopes (bedrock landslides). On the other end, older slope may be prone to stronger weathering. This factor takes a qualitative value based on different zones delimited on the terrain, High part (Upslope), Flat part, Low part (Downslope) (Figure 21).

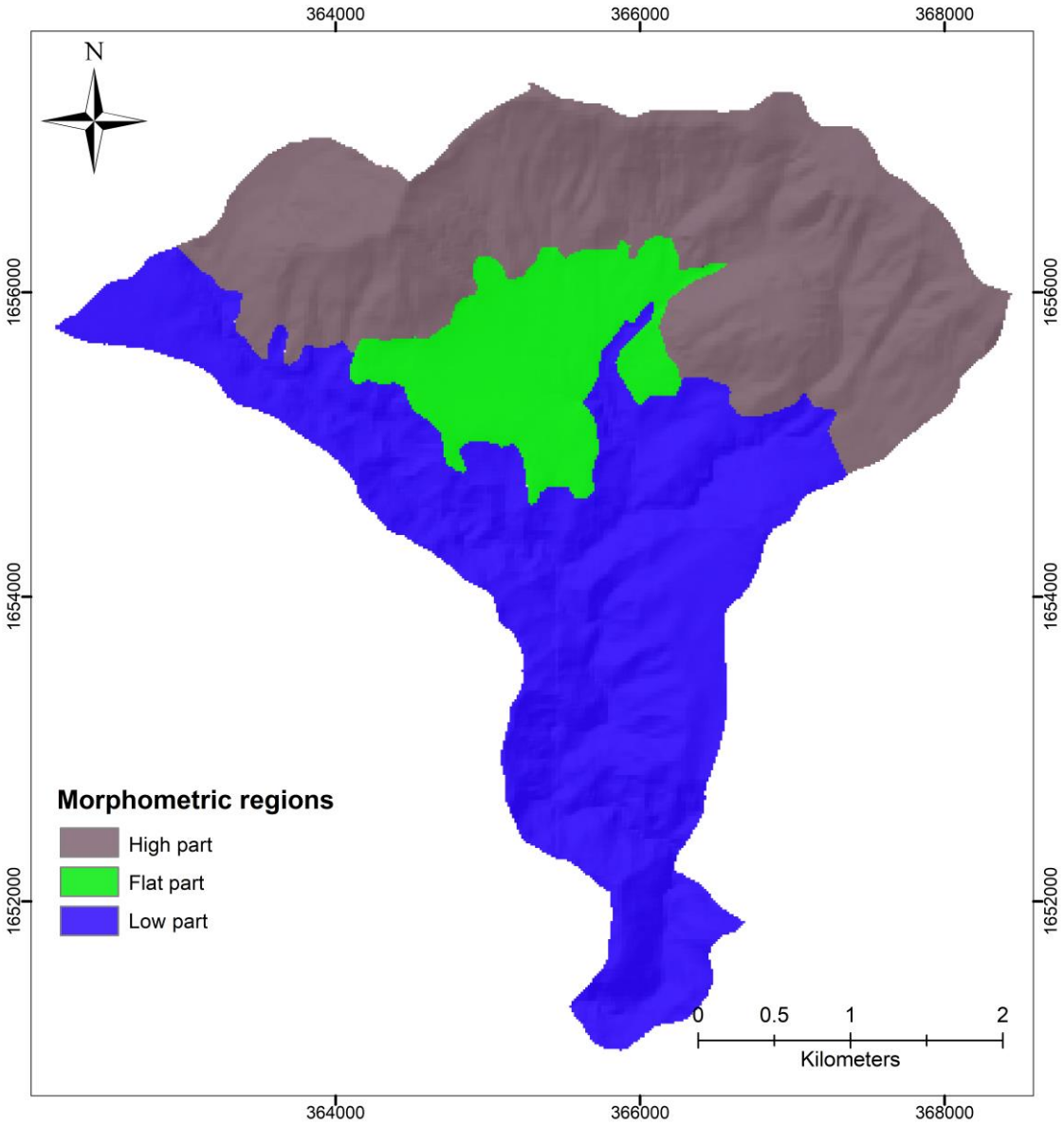


Figure 21 — Morphometric regions map.

4.3 Bivariate analysis

In order to assess the importance of the predisposing factors as precondition for landslides in the study area, the values of these factors affected by landslides were compared to the values for the whole study area. This is an important exploration step of the dataset before adjusting a multivariate regression model to spatially predict landslide occurrences (Lee & Talib, 2005; Bui *et al*., 2011; Yalcin *et al*., 2011; Shabanzadeh *et al*., 2011; Ozdemir & Altural, 2013; Solaimani *et al*., 2013; Pardeshi *et al*., 2013; Wu *et al*., 2016).

The Frequency ratio method is based on relationship between distribution of landslides and causative factors. The frequency ratio is the ratio of the area where landslides occurred in the total study area and is also the ratio of the probabilities of a landslide occurrence to a non-occurrence for a given attribute (Lee & Talib, 2005; Wu *et al*., 2016). The method requires the reclassification into classes of continuous variables. The predisposing factors were classified in different classes. The Jenks natural breaks algorithm was used for slope, defined interval of 100 meters for altitude, multiples of two for contributing drainage area, geometrical interval for distance to roads and drains to determinate the threshold values between classes of continuous variables. Aspect, profile curvature, plan curvature, lithology, land cover and region were classified following the criteria of source data. Details for each variable are explained above. The spatial relationship between landslide locations and each predisposing factor was extracted. The plateau (flat zone) of the study area is removed from the analysis.

The numbers of landslide occurrence in each class is evaluated, and then the frequency ratio value for each factor is calculated as follows (Lee & Talib, 2005; Shabanzadeh *et al*., 2011; Wu *et al*., 2016).

$$Fr_i = \frac{N_{pix(l_i)}/N_{pix(N_i)}}{\sum N_{pix(l_i)}/\sum N_{pix(N_i)}}$$

$N_{pix(l_i)}$: The number of pixels containing landslides in class (i).

$N_{pix(N_i)}$: Total number pixels having class (i) in the whole area of watershed.

$\sum N_{pix(l_i)}$: Total number of pixels containing landslides.

$\sum N_{pix(N_i)}$: Total number of pixels in the whole area of watershed.

Frequency Ratio (Fri) assesses the relative importance of each class with respect to landslides, when the quantity of Fri is greater than one, relationship is positive between landslide occurrence and the class of a data layer (direct correlation) and if this ratio be smaller than one, this relationship is negative (absence of correlation) (Lee & Talib, 2005).

4.4 Susceptibility analysis

4.4.1 Logistic regression

The logistic regression allows to estimate the probability of a binary response on the basis of independent variables (Van Asch *et al*., 1999; Dai & Lee, 2002; Vanacker *et al*., 2003; Ohlmacher & Davis, 2003; Ohlmacher, 2007; Van Den Eeckhaut, 2006; Van Den Eeckhaut *et al*., 2010; Nefeslioglu *et al*., 2008; Guzzetti *et al*., 1999; Rossi *et al*., 2010a; Bui *et al*., 2011; Wang *et al*., 2013; Pardeshi *et al*., 2013; Trigila *et al*., 2015).

For this research, stepwise logistic regression (LR) is used to describe the relation between a dichotomous categorical variable named the dependent variable (Y, presence or absence of a landslide) and a set of explanatory variables named the independent predictor variables (X1, X2,…..Xn). These variables can be continuous or discrete, and do not need to be normally distributed. Categorical variables need to be represented by numerical codes named dummy variable, which is an artificial variable created to represent an attribute with two or more distinct categories/levels. The dependent variable is binary coded as "0" for this case absence of landslide and "1" representing the presence of landslide. The logistic response function can be written as (Hosmer & Lemeshow, 2000):

$$P(Y = 1) = \pi(x) = \frac{1}{[1 + exp^{-(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n)}]}$$

where $\pi(x)$ is the probability of event occurrence in function of the *x* vector of values corresponding to explanatory variables. For this study, the event is the occurrence of a landslide, β_0 is the intercept ($\gamma\ intercept$), and $\beta_i\ with\ i \neq 0$ are the coefficients (first derivative parameters) for the independent variables x_i (predictors). To fit the unknown parameters of the logistic regression model, the values β_i, the maximum likelihood method is used. The significance of coefficients β_i is tested with the Wald test, which is obtained comparing the maximum likelihood estimates of every β_i to its estimated standard error (Hosmer & Lemeshow, 2000). A coefficient is significant if the tested null hypothesis that the estimated coefficient is 0 can be rejected at a 0.01 or 0.05 significance level. The output probability values ranging from 0 to 1 indicate respectively the 0 % to 100 % of probability of occurrence of a landslide.

The logistic regression function ($\pi(x)$) can be linearized with the following logit transformation. The logarithm of the odd, $log(\pi(x)/1 - \pi(x))$, called the logit, is linearly related with the independents variables:

$$\log\left(\frac{\pi(x)}{1 - \pi(x)}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n$$

4.4.2 Multicollinearity analysis

The logistic regression is sensitive to collinearities among the independent variables (Hosmer & Lemeshow, 2000). Two collinearity diagnostic parameters are widely used to identify multicollinearity: Tolerance (TOL) and the variance inflation factor (VIF) (O'brien, 2007; Wang *et al*., 2013; Bui *et al*., 2011).

Tolerance –TOL- is defined by: $1 - R_i^2$. The R_i^2 and used to represent the proportion of variance in the $_i$th independent variable that is associated with the other independent variables in the model. Tolerance for the $_i$th independent variable is 1 minus the proportion of variance it shares with the other independent variable in the analysis. This represents the proportion of variance in the *i*th independent variable that is not related to the other independent variables in the model. A small tolerance value indicates that the predicting variable under consideration is almost a perfect linear combination of the other independent variables already in the equation and that it should not be added to the regression equation (O'brien, 2007). TOL must be as large as possible.

The Variance Inflation Factor - VIF - is the reciprocal of tolerance defined as $1/(1 - R_i^2)$. It measures the impact of collinearity among the variables in a regression model. The VIF has an intuitive interpretation in terms of the effects of R_i^2 on the variance of the estimated regression coefficient for the $_i$th independent variable, it is always greater than or equal to 1 and it should be as low as possible (O'brien, 2007).

Multicollinearity diagnostic statistics was produced by linear regression. In this research variables with VIF > 2 and TOL <0.4 were excluded from the analysis (Van Den Eeckhaut, 2006; Dewitte *et al*., 2015).

4.4.3 Landslide Sampling

In order to use a model for prediction, the sampling is essential (Heckmann *et al*., 2014). The strategy for logistic regression is to select Y by collecting observations (randomly or all those available) for which Y = 1 (the "landslides") and a random selection of observations for which Y = 0 (the "non-landslides"), (King & Zeng, 2001).

According to King and Zeng (2001), the optimal trade-off between collecting more observations is application-specific, and so decisions will necessarily involve judgment calls and qualitative assessments. The trade-off when zeros and ones are equally easy to collect, and an unlimited number of each are available, an "equal shares sampling design" is optimal, if collecting zeros were costless, we should collect as many as we can get, since more data are always better. According to Heckmann *et al*. (2014), the literature suggests different ratios of non-event sample sizes which generally range from 1:1 to 1:10.

In this research when we speak of sample size, we always address the strategy of the sample of "non-events". To explore the sensitivity of model to the non-event sample size, the ratio between Y=1 and Y=0 will be 1:1 and 1:10. As two different sample sizes will be used, the Kolmogorov–Smirnov test (K–S test) will be applied to test whether samples are drawn from the same distribution data source (Guns & Vanacker, 2014).

The sampling zone to be used for calibration is usually selected considering the whole or a limited portion of study area (Ohlmacher & Davis, 2003; Dai & Lee, 2002). In this research the flat zone of study area are removed for the sampling.

The model is calibrated with the 2005 inventory. The points representing the absence of landslide were obtained using a spatially uniform sampling scheme. A zone of 30m for all

landslides were excluded to minimize errors related to boundaries and influence area of landslides.

4.4.4 Calibration

Logistic regression was performed using statistical software R, incorporating the predictor variables selected for their importance in the presence of landslide, the significance p-values for a variable to stay in the model was set at 0.05. The probability of landslide occurrence was calculated using the logistic regression coefficients in different landslides susceptibility maps with respect to different models. Probability ranges are between 0 and 1.

Logistic regression models will be calibrated using 80 % of the data in 2005 landslide inventory. To calibrate the best model in the three dataset and to achieve reduce the complexity of model, a variable selection procedure will be carried out, removing the parameters not significant for the prediction. The first model start with all predictor variables (without collinearity), in next models (two more) the variables with the higher rank of importance will be including.

4.5 Validation model

Model validation allows to establish the degree of confidence of the model, assessing its accuracy or predictive power. Studies in susceptibility modelling can use spatial or temporal cross validation (space or time partition) (Chung and Fabbri, 2003). In this research the dataset is split randomly containing 80 % for calibration and 20 % of dataset for validation for all models. Moreover temporal validation will be made with 2011 inventory.

Models performance can be estimated using the Receiver Operating Characteristic (ROC) curves (e.g. Van Asch *et al*., 1999; Peng & So, 2002; Beguería, 2006; Van Den Eeckhaut, 2006, 2010; Nefeslioglu *et al*., 2008; Rossi *et al*., 2010a; Pourghasemi *et al*., 2012; Wang *et al*., 2013; Pardeshi *et al*., 2013).

The ROC curve is a plot of sensitivity versus 1-specifivity, where sensitivity (y-axis) is defined as the proportion of observations correctly classified as an event, named true positive fraction (TP), over the total number of observed (real) landslide (TP + FN) and specificity is defined as the proportion of observations correctly classified as nonevent (TN), over the total number of observed (real) non landslides (FP + TN), therefore 1-specifivity (X-axis) is the proportions of the observation misclassified as an event named false positive fraction (Peng & So, 2002; Hosmer and Lemeshow, 2000). The area under the ROC curve (AUC) is a measure of discrimination that summarizes the information contained in the plot, an AUC value close to 1 indicates high accuracy of the prediction result (AUC=1 means perfect discrimination), whereas a value close to 0.5 indicates low accuracy, AUC=0.5 means no discrimination at all, or random forecast (Wang *et al*., 2013; Hosmer & Lemeshow, 2000).

The four-fold plot display of a 2 by 2 contingency table summarizing the number of true positive -TP- (pixels predicted landslide and observed landslide), true negatives -TN- (pixels predicted non-landslides and observed non-landslides, false positives -FP- (pixels predicted landslides but observed non-landslides), and false negatives -FN- (pixels predicted non-landslides but observer landslides). The confusion matrix plot is a visual representation to compares the predicted score and observed outcome. This evaluation method is useful if the susceptibility map is used for

planning purposes (Trigila *et al*., 2015). For this research, the decision probability threshold to classify a pixel as landslide and non-landslides is set at a typical value of 0.5 given the equal number of landslide and no landslide (Dewitte *et al*., 2015). These plots will be make for the best performance model in each dataset.

To visualize the model outputs, the probability values to landslides occurrences will be displayed with a map contained the continuous variable which is certainly more informative than a sorted categorical scale. Besides, as proposed by Iovine *et al*. (2013), for the sake of simplifying the probability map, the values of probabilities can be arranged into classes by adopting the ‘‘natural breaks (Jenks)’’, as form to variance-minimization classification.

5 RESULTS

5.1 Landslide inventory

For the 2005 event, tropical storm “Stan”, 766 landslides were identified and mapped, most of them, 464 (61 %) started as shallow landslides and 302 (39 %) evolved into debris flows. Moreover, a total of 99 landslides were mapped for the 2011 event, tropical depression “Twelve-E”, and classed as shallow landslides. Due to the rapid vegetation growth and the preparation of field for crops, the presence of flows that determine the debris mechanism was not clearly seen.

As the landslide types in the study area are not clearly and easily recognizable on the orthoimages, according to Cruden and Varnes (1996) in Guzetti *et al*. (2012), the classification are based on a quantitative criterion and on fieldtrip experience. We divided the landslides in two categories according to their area. In this study, landslides are considered shallow landslide when the area is less than or equal to 0.0215 Ha (215 m²), whereas they are categorized as debris flows when the area is more than 0.0215 Ha. Details about the sliding mechanism are given in Table 5 as well as details about the dimensions of the landslides (area and approximated length). Landslides were digitized using ArcGis to calculate their area. Their length was calculated using the bounding container toolbox in ArcGis. The length was estimated by the longest axis of the rectangle containing each digitized flow. With this method, the length is slightly underestimated for winding flows but this gives a satisfactory approximation of the magnitude of the landslides. Figure 22 shows the location of shallow landslides and debris flows which are illustrated with different examples in Figure 23 and Figure 24 respectively.

Table 5 — Number of Landslides and their dimensions for the study area.

Sliding mechanism	n	Area (Ha)				Length (m)			
		Mean	Standard deviation	Min	Max	Mean	Standard deviation	Min	Max
Shallow landslides	464	0.0082	0.0055	0.00055	0.0215	14.90	6.51	2.68	30.52
Debris flows	302	0.1051	0.1680	0.02254	1.9420	126.8	167.48	7.86	1042

The mean shallow landslides size is almost twelve times smaller than debris flows and the average length is eight times smaller.

As displayed on the map (Figure 22), the shallow landslides and debris flows are distributed all over the area, without a singular pattern among its, respect to the elevation, slope or land cover. Concerning debris flows, they are longer in the upper part of the area.

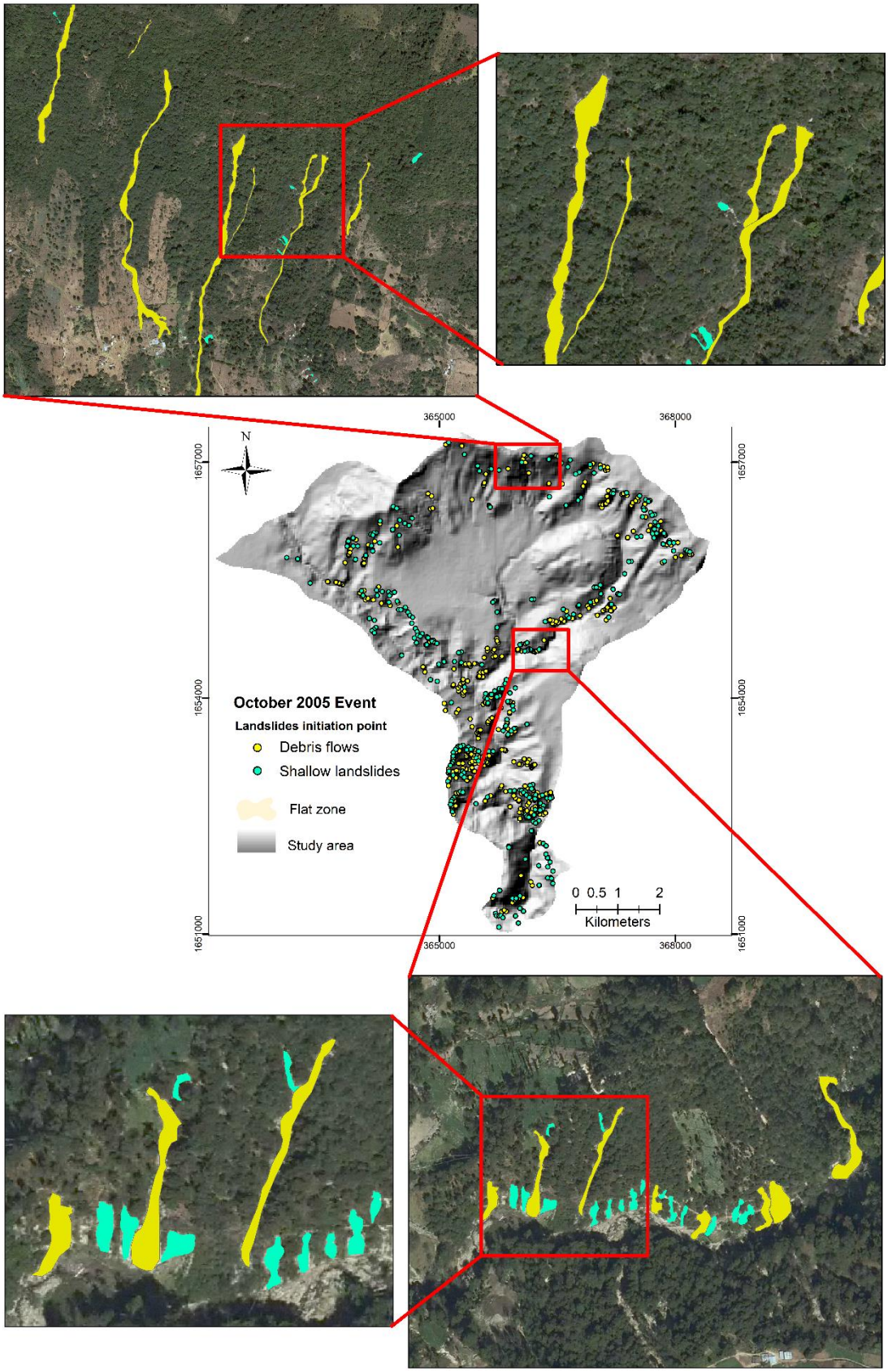


Figure 22 —Landslides inventory map for the study area.

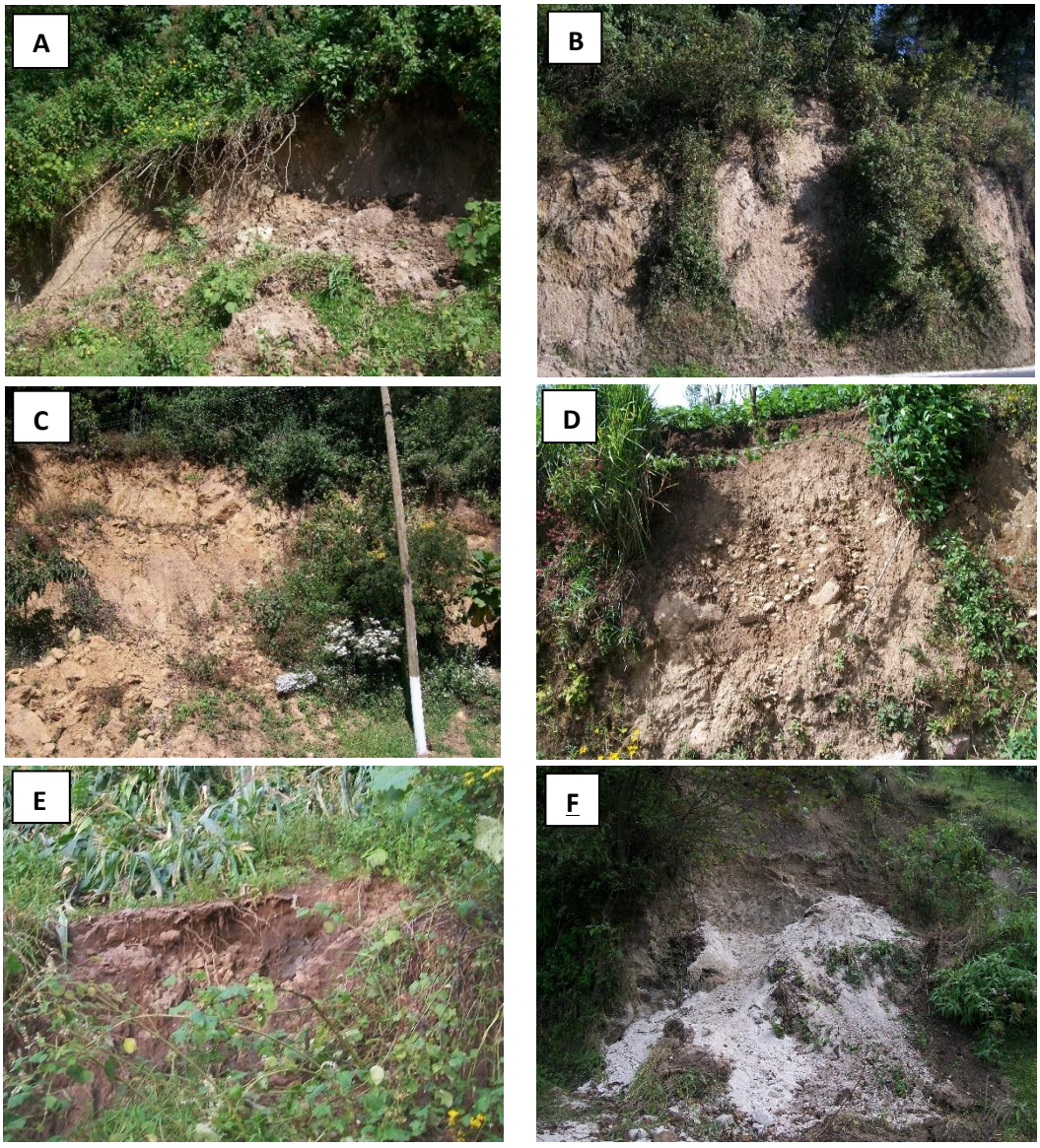


Figure 23 — A)-F) Pictures showing shallow landslides in the study area (© Estrada, 2011).

Figure 24 — A)-D) Pictures showing debris flows in the study area (© Estrada, 2013).

5.2 Frequency Ratio analysis

The frequency ratio values (Fri) for each class of the 12 predisposing factors are in Table 5. The relationship between landslides occurrence and slope percentage for three datasets (shallow landslides, debris flows, and all landslides) shows a positive correlation for three classes, the ratio is greater than 1 above 48 % slope, and ratio increase gradually with slope. For the first two classes between 0 % to less than 48 %, the Fri values are under 1. The slope distribution is displayed in Figure 25, comparing the landslides dataset, the range 48-72 and 71-102 have the major landslides frequency.

In the case of slope aspect, the ratio values are above 1 for South and South-West-facing slopes for the three datasets, and above 1 for South-East-facing slopes for two datasets all landslides and shallow landslides. For North, North-East, East, West and North-West-facing slopes the ratio is under 1. The rose diagram (Figure 26) shows clearly the occurrence of landslides in the slope facing South, South-East and South-West.

Concerning to elevation, Fri is above 1 for classes from 1800 – 2300 in the three datasets, the class 2700 – 2800 the Fri values are greater than 1 for debris flows. Shallow landslides are more related with the class from 2000 – 2200 (Fri > 2), and debris flows with the class 1800 – 1900. The classes between 2300 – 2600; 2800 – 3000 the Fri values are smaller than 1. The histogram (Figure 27) shows the occurrence of landslides in lowest class elevations.

The Fri values for contribute drainage area are greater than 1 in classes 0 – 2, 2 – 4 and 4 – 8 for all landslides and debris flows datasets. To shallow landslides classes 0 – 2, 8 – 16 and 32 – 64 and major than 128 Fri values are greater than 1. The occurrence of landslides is more frequently when there is less contributing drainage area, as showed in Figure 28.

In the case of distance to drains, the Fri values in three classes from 58 to 269 meters are greater than 1 for all landslides and shallow landslides datasets. The Fri values for debris flows are greater than 1 in two classes from 86 – 269 meters. The first and last classes 0 – 58 and 269 – 541 meters have Fri values smaller than 1. The Figure 29 show peaks of landslides occurrence in a major distance to drains until 269 meters, after that the frequency is lower, however the frequency of landslide in short distance is quite important.

For distance to roads, two classes from 58 to 304 meters Fri values are greater than 1 for shallow and all landslides. To debris flows two classes are greater than 1, 58 - 139 and 304 - 645 meters. The classes less than 58 meters the Fri values are smaller than 1. The occurrence of landslide tends to be more frequent when the distance increases (Figure 30).

Concerning lithology, for volcanic rocks, the Fri values are greater than 1 only for shallow landslides. Fr pumice ash landfills the Fri value is greater than 1 for debris flows and all landslides. The shallow landslides occurrence is higher in volcanic rocks in the study area (Figure 31).

In the case of land cover, for forests the Fri values are greater than 1 for the three datasets. Natural pasture class presents values greater than 1 for shallow landslides and al landslides. Roads are related only with debris flows. Urban and crops land cover have Fri values smaller

than 1. As showed in Figure 32 the occurrence of landslides is higher in forest, we can see that forest is the major land cover in study area.

Profile curvature is correlated with landslides occurrence in class upward convex with the three datasets, the Fri values are greater than 1. In classes straight and upward concave the Fri values are smaller than 1. The correlation between upward convex and landslides frequency is showed in Figure 33.

Concerning planform curvature, for the divergent class the Fri values are greater than 1 for all landslides and shallow landslides. The convergent class presents Fri values greater than 1 to debris flows and all landslides. Parallel class has Fri values smaller than 1 for the three datasets. The landslides frequency is showed in Figure 34.

Concerning morphometric regions, in class low part the Fri values are greater than 1 for the three datasets. For high part and plan the Fri values are smaller than 1. This variable confirms the values shown in elevation: landslides occurrence are most frequent at low elevation (Figure 35).

In the case of landscape fragmentation, only in class 1 patches has values greater than 1 for the three datasets. Classes 5 and 6 have Fri values greater than 1 for debris flows. The lower are the number of fragments, the less are the landslides occurrences, as we can see distribution on Figure 36.

Table 6 — Frequency Ratio (Fri) in the 12 predisposing factors.

Variable	Class	Class	Shallow landslides	Debris flows	All landslides	Class	Shallow landslides	Debris flows	All landslides	Shallow landslides	Debris flows	All landslides
		$N_{pix(N_i)}$	$N_{pix(L_i)}$	$N_{pix(L_i)}$	$N_{pix(L_i)}$	% $N_{pix(N_i)}$	% $N_{pix(L_i)}$	% $N_{pix(L_i)}$	% $N_{pix(L_i)}$	Fr_i	Fr_i	Fr_i
Slope (%)	0 - 25	12921	13	16	29	20.42	2.80	5.30	3.79	0.14	0.26	0.19
	25 - 48	17014	71	40	111	26.88	15.30	13.25	14.49	0.57	0.49	0.54
	48 - 72	18090	156	98	254	28.58	33.62	32.45	33.16	1.18	1.14	1.16
	72 - 102	11886	166	110	276	18.78	35.78	36.42	36.03	1.90	1.94	1.92
	102 -285	3379	58	38	96	5.34	12.50	12.58	12.53	2.34	2.36	2.35
Aspect	North	3642	8	7	15	5.75	1.72	2.32	1.96	0.30	0.40	0.34
	North - East	3726	20	13	33	5.89	4.31	4.30	4.31	0.73	0.73	0.73
	East	5912	34	20	54	9.34	7.33	6.62	7.05	0.78	0.71	0.75
	South - East	9189	110	41	151	14.52	23.71	13.58	19.71	1.63	0.94	1.36
	South	12993	137	106	243	20.53	29.53	35.10	31.72	1.44	1.71	1.55
	South - West	11920	105	67	172	18.83	22.63	22.19	22.45	1.20	1.18	1.19
	West	9626	35	38	73	15.21	7.54	12.58	9.53	0.50	0.83	0.63
	North - West	6282	15	10	25	9.93	3.23	3.31	3.26	0.33	0.33	0.33
Elevation	1800 - 1900	508	7	7	14	0.80	1.51	2.32	1.83	1.88	2.89	2.28
	1900 - 2000	1697	14	10	24	2.68	3.02	3.31	3.13	1.13	1.23	1.17
	2000 - 2100	4487	66	22	88	7.09	14.22	7.28	11.49	2.01	1.03	1.62
	2100 - 2200	6769	150	64	214	10.70	32.33	21.19	27.94	3.02	1.98	2.61
	2200 - 2300	12616	95	87	182	19.93	20.47	28.81	23.76	1.03	1.45	1.19
	2300 - 2400	10270	40	22	62	16.23	8.62	7.28	8.09	0.53	0.45	0.50
	2400 - 2500	8401	14	23	37	13.27	3.02	7.62	4.83	0.23	0.57	0.36
	2500 - 2600	8505	21	12	33	13.44	4.53	3.97	4.31	0.34	0.30	0.32
	2600 - 2700	4314	24	27	51	6.82	5.17	8.94	6.66	0.76	1.31	0.98
	2700 - 2800	3774	27	22	49	5.96	5.82	7.28	6.40	0.98	1.22	1.07
	2800 - 2900	1657	6	6	12	2.62	1.29	1.99	1.57	0.49	0.76	0.60
	2900 - 3000	292	0	0	0	0.46	0.00	0.00	0.00	0.00	0.00	0.00
CDA	0 - 2	15621	112	90	202	24.68	24.14	29.80	26.37	0.98	1.21	1.07
	2 - 4	8702	82	52	134	13.75	17.67	17.22	17.49	1.29	1.25	1.27
	4 - 8	12973	90	70	160	20.50	19.40	23.18	20.89	0.95	1.13	1.02
	8 - 16	12765	94	53	147	20.17	20.26	17.55	19.19	1.00	0.87	0.95
	16 - 32	7226	44	26	70	11.42	9.48	8.61	9.14	0.83	0.75	0.80
	32 - 64	2554	23	3	26	4.04	4.96	0.99	3.39	1.23	0.25	0.84
	64 - 128	1096	1	1	2	1.73	0.22	0.33	0.26	0.12	0.19	0.15
	> 128	2353	18	7	25	3.72	3.88	2.32	3.26	1.04	0.62	0.88
Distance to drains (m)	0- 58	16389	100	60	160	25.90	21.55	19.87	20.89	0.83	0.77	0.81
	58 - 86	7732	66	32	98	12.22	14.22	10.60	12.79	1.16	0.87	1.05
	86 - 144	14299	117	75	192	22.59	25.22	24.83	25.07	1.12	1.10	1.11
	144 - 269	19090	158	115	273	30.16	34.05	38.08	35.64	1.13	1.26	1.18
	269 - 541	5780	23	20	43	9.13	4.96	6.62	5.61	0.54	0.73	0.61

Variable	Class	Class	Shallow landslides	Debris flows	All landslides	Class	Shallow landslides	Debris flows	All landslides	Shallow landslides	Debris flows	All landslides
		$N_{pix(N_i)}$	$N_{pix(l_i)}$	$N_{pix(l_i)}$	$N_{pix(l_i)}$	% $N_{pix(N_i)}$	% $N_{pix(l_i)}$	% $N_{pix(l_i)}$	% $N_{pix(l_i)}$	Fr_i	Fr_i	Fr_i
Distance to roads (m)	0 - 19	7631	45	31	76	12.06	9.70	10.26	9.92	0.80	0.85	0.82
	19 - 58	12590	84	55	139	19.89	18.10	18.21	18.15	0.91	0.92	0.91
	58 - 139	18118	156	90	246	28.63	33.62	29.80	32.11	1.17	1.04	1.12
	139 - 304	18074	146	85	231	28.56	31.47	28.15	30.16	1.10	0.99	1.06
	304 - 645	6877	33	41	74	10.87	7.11	13.58	9.66	0.65	1.25	0.89
Lithology	Volcanic rocks	42798	324	171	495	67.62	69.83	56.62	64.62	1.03	0.84	0.96
	Pumice ash landfills	20492	140	131	271	32.38	30.17	43.38	35.38	0.93	1.34	1.09
Land cover	Forest	37463	347	251	598	59.19	74.78	83.11	78.07	1.26	1.40	1.32
	Urban	2392	1	0	1	3.78	0.22	0.00	0.13	0.06	0.00	0.03
	Crops	20030	90	40	130	31.65	19.40	13.25	16.97	0.61	0.42	0.54
	Natural pastures	2447	25	6	31	3.87	5.39	1.99	4.05	1.39	0.51	1.05
	Roads	958	1	5	6	1.51	0.22	1.66	0.78	0.14	1.09	0.52
Profile curvature	Upward convex	30309	267	207	474	47.89	57.54	68.54	61.88	1.20	1.43	1.29
	Straight ®	2829	9	5	14	4.47	1.94	1.66	1.83	0.43	0.37	0.41
	Upward Concave	30152	188	90	278	47.64	40.52	29.80	36.29	0.85	0.63	0.76
Plan curvature	Divergent (convex)	28013	219	116	343	44.26	47.20	38.41	44.78	1.07	0.87	1.01
	Parallel ®	4012	18	4	22	6.34	3.88	1.32	2.87	0.61	0.21	0.45
	Convergent (concave)	31265	227	182	401	49.40	48.92	60.26	52.35	0.99	1.22	1.06
Region	High part	30611	106	96	202	48.37	22.84	31.79	26.37	0.47	0.66	0.55
	Plan	117	0	0	0	0.18	0.00	0.00	0.00	0.00	0.00	0.00
	Low part	32562	358	206	564	51.45	77.16	68.21	73.63	1.50	1.33	1.43
Fragmentation	1	35908	294	198	492	56.74	63.36	65.56	64.23	1.12	1.16	1.13
	2	18988	127	75	202	30.00	27.37	24.83	26.37	0.91	0.83	0.88
	3	4746	29	15	44	7.50	6.25	4.97	5.74	0.83	0.66	0.77
	4	1520	6	4	10	2.40	1.29	1.32	1.31	0.54	0.55	0.54
	5	1180	8	6	14	1.86	1.72	1.99	1.83	0.92	1.07	0.98
	6	794	0	4	4	1.25	0.00	1.32	0.52	0.00	1.06	0.42
	7	106	0	0	0	0.17	0.00	0.00	0.00	0.00	0.00	0.00
	8	24	0	0	0	0.04	0.00	0.00	0.00	0.00	0.00	0.00
	9	12	0	0	0	0.02	0.00	0.00	0.00	0.00	0.00	0.00
	10	8	0	0	0	0.01	0.00	0.00	0.00	0.00	0.00	0.00
	11	4	0	0	0	0.01	0.00	0.00	0.00	0.00	0.00	0.00

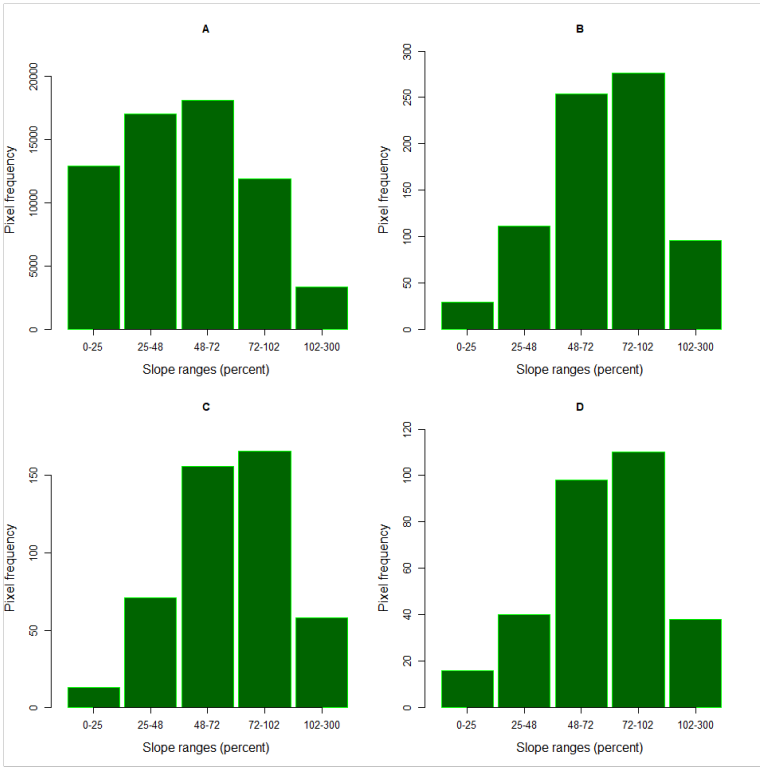


Figure 25 — Frequency distribution of slope for study area and landslides datasets. A) All research surface. B) All landslides. C) Shallow landslides. D) Debris flows.

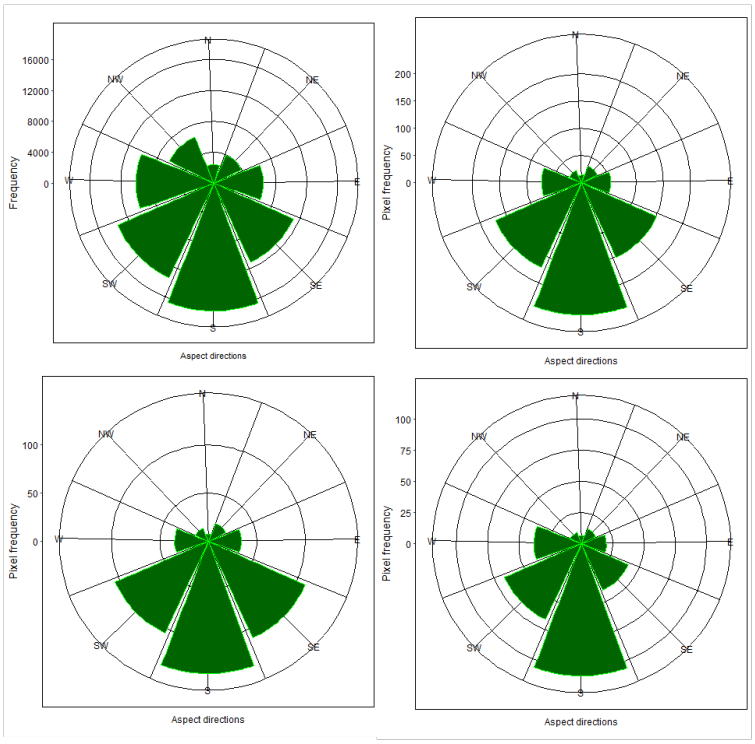


Figure 26 — Frequency distribution of slope aspect for study area and landslides datasets. A) All research surface. B) All landslides. C) Shallow landslides. D) Debris flows.

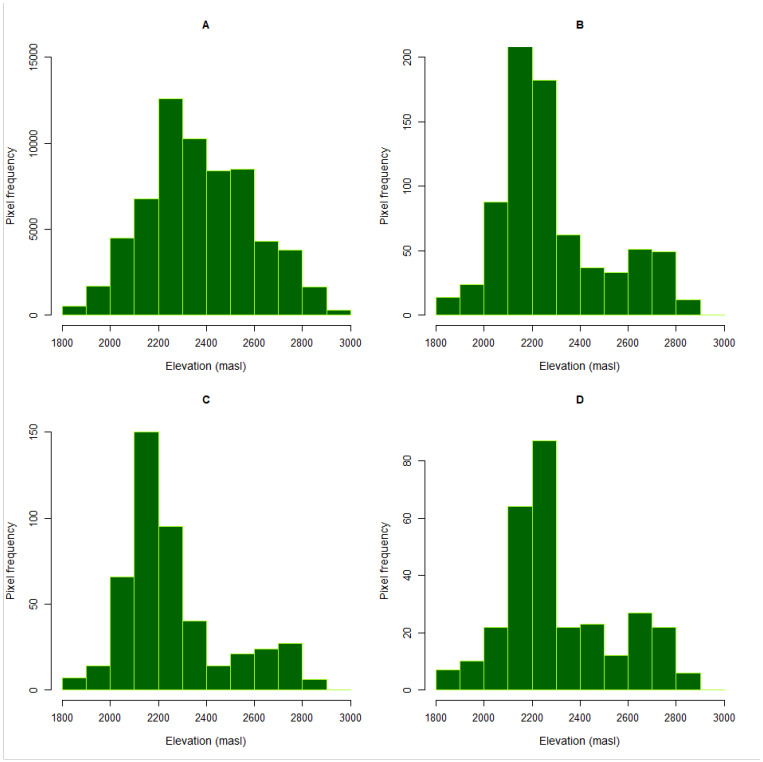


Figure 27 — Frequency distribution of elevation for study area and landslides datasets. A) All research surface. B) All landslides. C) Shallow landslides. D) Debris flows.

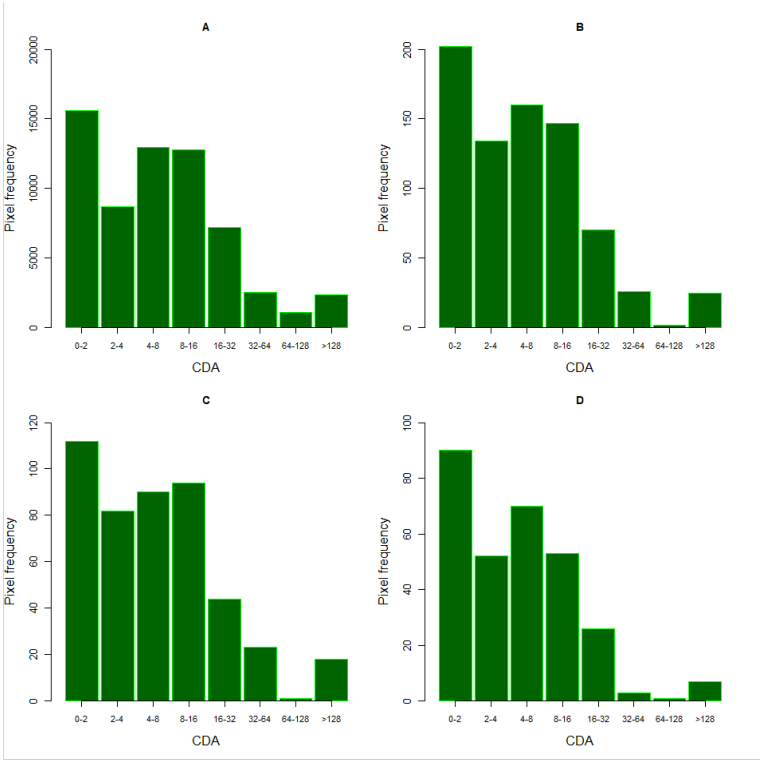


Figure 28 — Frequency distribution of contributing drainage area for study area and landslides datasets. A) All research surface. B) All landslides. C) Shallow landslides. D) Debris flows.

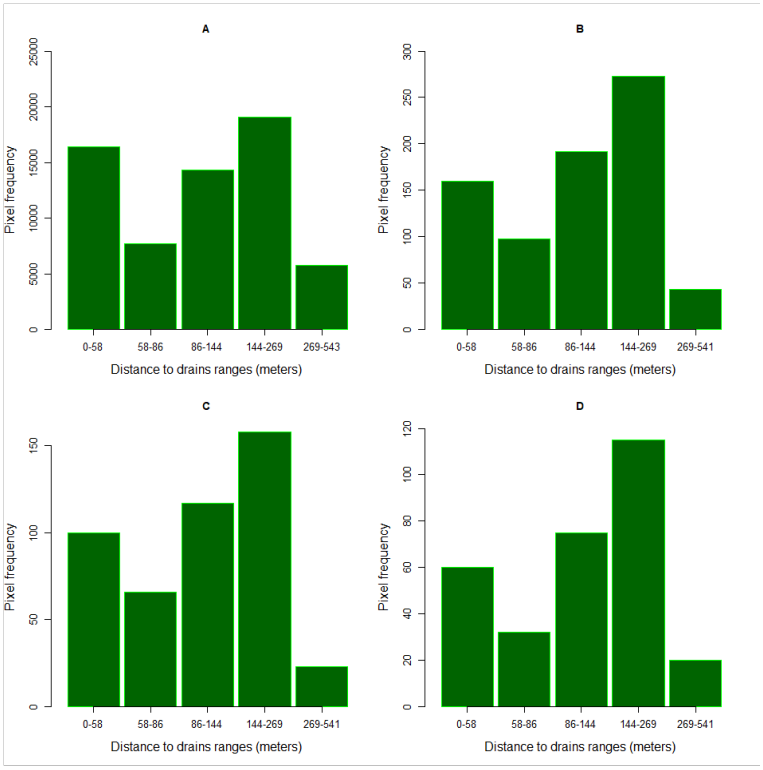


Figure 29 — Frequency distribution of Euclidian distance to drains for study area and landslides datasets. A) All research surface. B) All landslides. C) Shallow landslides. D) Debris flows.

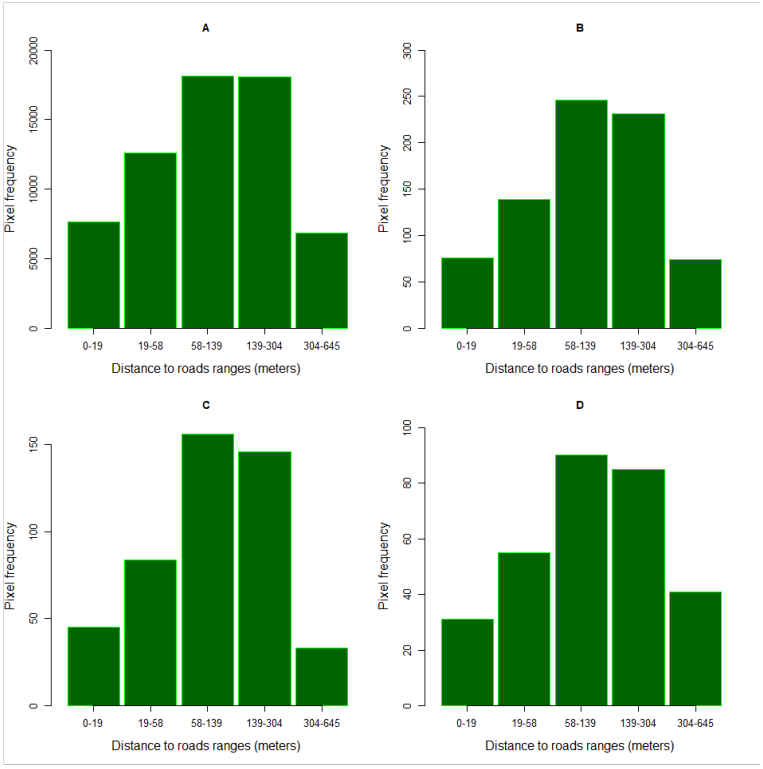


Figure 30 — Frequency distribution of Euclidian distance to roads for study area and landslides datasets. A) All research surface. B) All landslides. C) Shallow landslides. D) Debris flows.

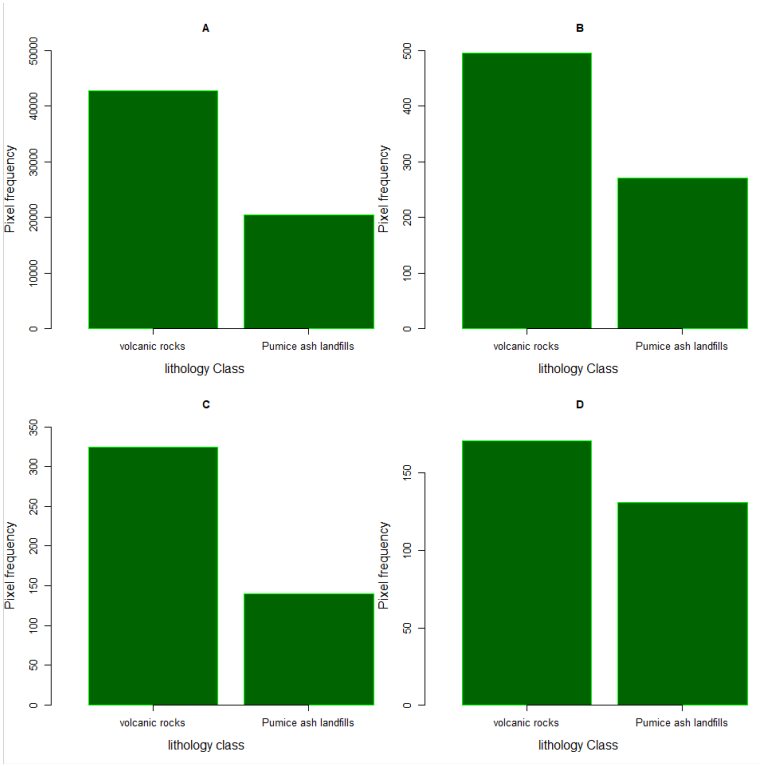


Figure 31 — Frequency distribution of lithology for study area and landslides datasets. A) All research surface. B) All landslides. C) Shallow landslides. D) Debris flows.

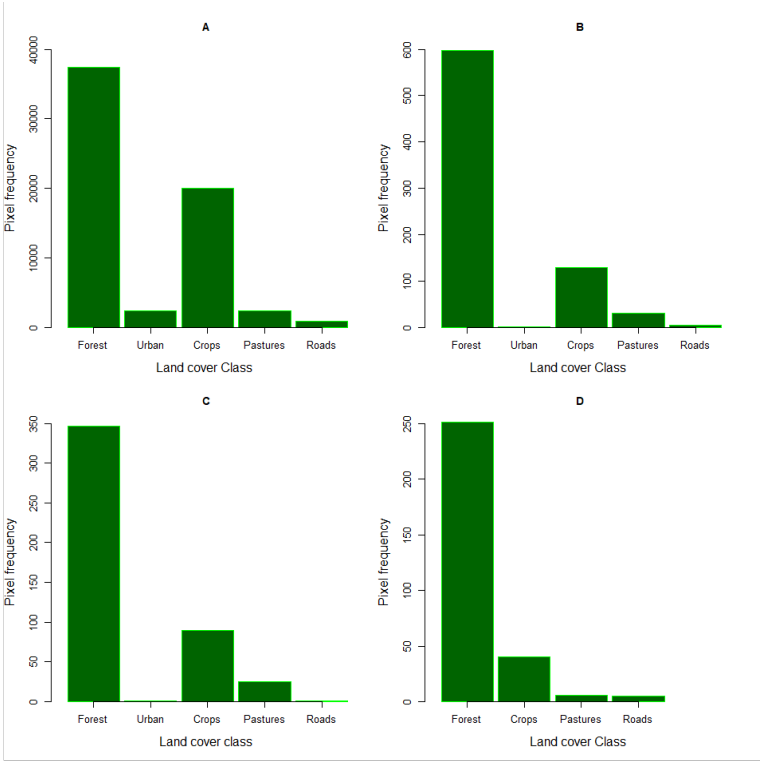


Figure 32 — Frequency distribution of land cover for study area and landslides datasets. A) All research surface. B) All landslides. C) Shallow landslides. D) Debris flows.

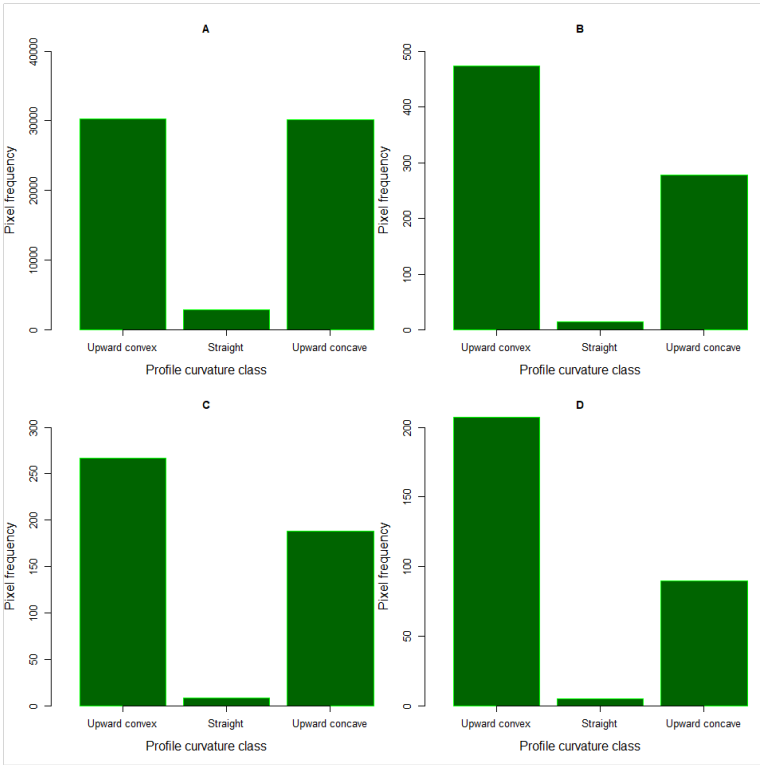


Figure 33 — Frequency distribution of profile curvature for study area and landslides datasets. A) All research surface. B) All landslides. C) Shallow landslides. D) Debris flows.

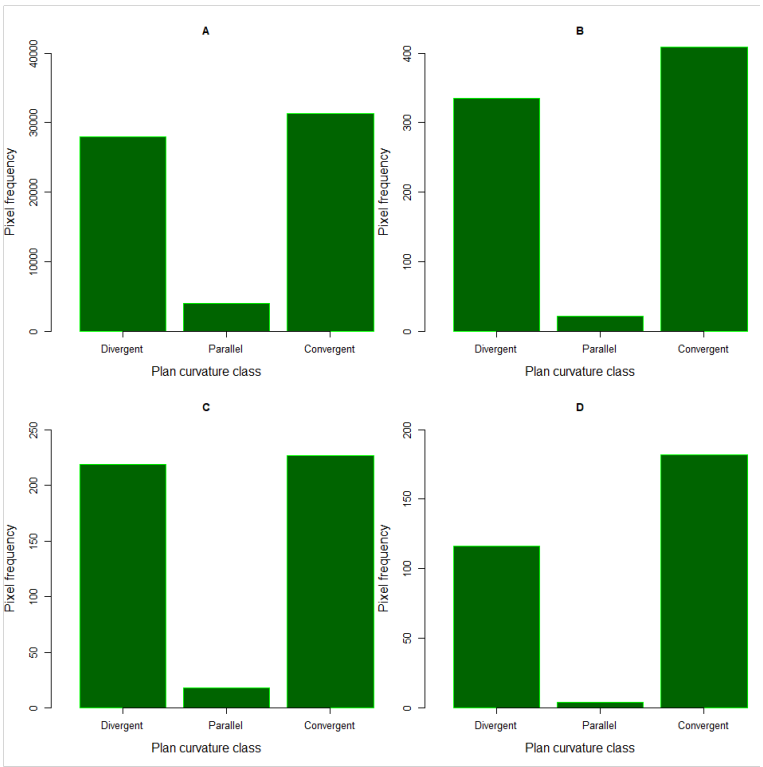


Figure 34 — Frequency distribution of planform curvature for study area and landslides datasets. A) All research surface. B) All landslides. C) Shallow landslides. D) Debris flows.

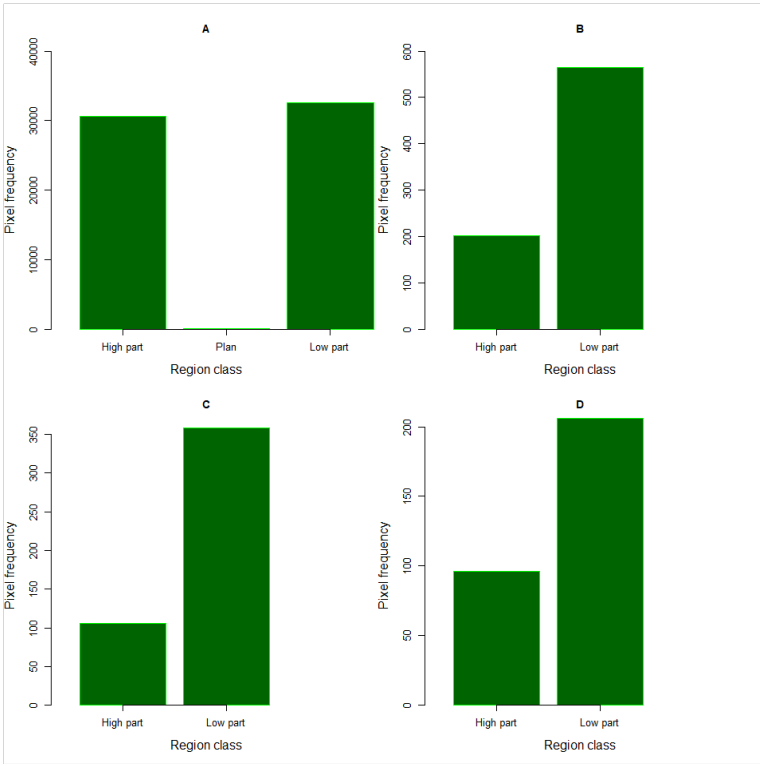


Figure 35 — Frequency distribution of morphometric region for study area and landslides datasets. A) All research surface. B) All landslides. C) Shallow landslides. D) Debris flows.

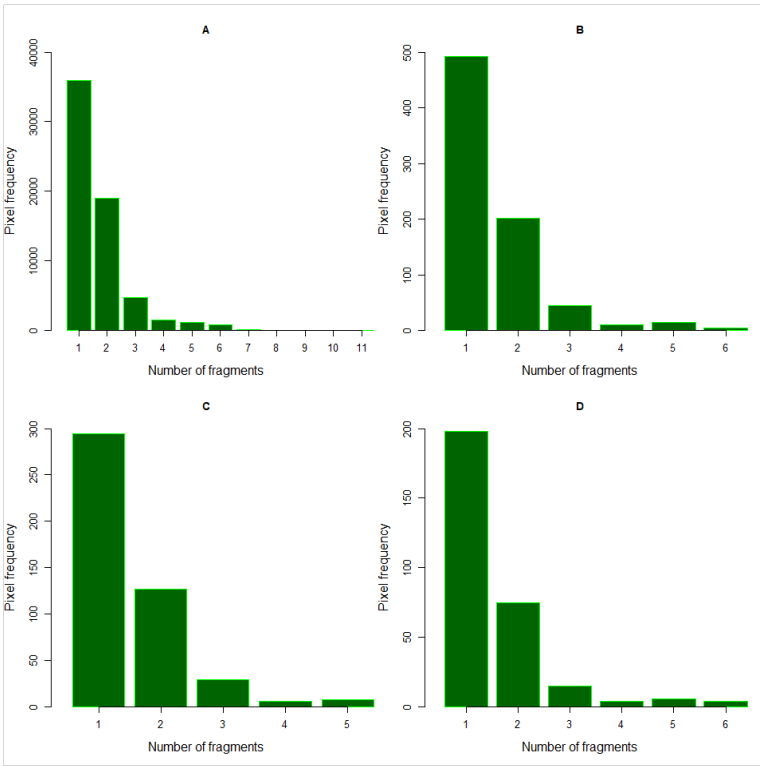


Figure 36 — Frequency distribution of landscape fragmentation for study area and landslides datasets. A) All research surface. B) All landslides. C) Shallow landslides. D) Debris flows.

5.3 Logistic regression results

The landslide inventory was used to calibrate logistic regression models. Shallow landslides, debris flows and the union of both datasets were treated differently. The resultant beta (β) coefficients, Z value and significance p-value in the logistic regression for the independent variable are given for each dataset.

5.3.1 Multicollinearity analysis

To avoid the multicollinearity, the analysis was applied for the predictor variables (Table 7). It revealed that Elevation and Region have VIF >2 and TOL <0.4 even <0.2. When the region variable has been taken out of multicollinearity analysis, the values VIF and TOL to Elevation are 1.183 and 0.846 respectively. For this research the elevation variable has been taken for calibration models.

Table 7 — Multicollinearity analysis, indexes for 12 variables.

No.	Variable	Tolerance	VIF
1	Slope percent	0.859	1.164
2	Aspect	0.941	1.062
3	Elevation*	0.181	5.510
4	Contributing drainage area	0.976	1.024
5	Distance to drains	0.645	1.549
6	Distance to roads	0.788	1.268
7	Lithology	0.782	1.279
8	Land cover	0.795	1.257
9	Profile curvature	0.742	1.348
10	Plan curvature	0.841	1.189
11	Region*	0.199	5.025
12	Fragmentation	0.813	1.230

*variables with collinearity

5.3.2 Sample size analysis

In the case of shallow landslides two different sample size strategies are used, 1 to 1 and 1 to 10 landslides and non-landslide events. To analyze if the samples have the same distribution as landslides in the study area, the cumulative distribution functions among non-landslide events and the distribution of the factors (elevation, slope, profile, planform, distance to drains and distance to roads) were compared statistically using Kolmogorov-Smirnov's test. The results values of "D" and significant p-value are showed in Table 8. The empirical cumulative distribution function was plotted for each variable (Figure 37).

Table 8 — Kolmogorov-Smirnov test for two sample size and 6 predisposing factors compared to the distribution on the study area.

No.	Variable	Sample size 1:1		Sample size 1:10	
		D	*P*-value	D	*P*-value
1	Slope percent	0.0309	0.7725	0.0210	0.0461*
2	Profile curvature	0.0321	0.7286	0.0108	0.7036
3	Planform curvature	0.0349	0.6265	0.0136	0.4112
4	Elevation	0.0402	0.4484	0.0310	0.0005*
5	Distance to drains	0.0419	0.3952	0.0176	0.1416
6	Distance to roads	0.0475	0.2491	0.0171	0.1655

Signif. codes: '*' 0.05

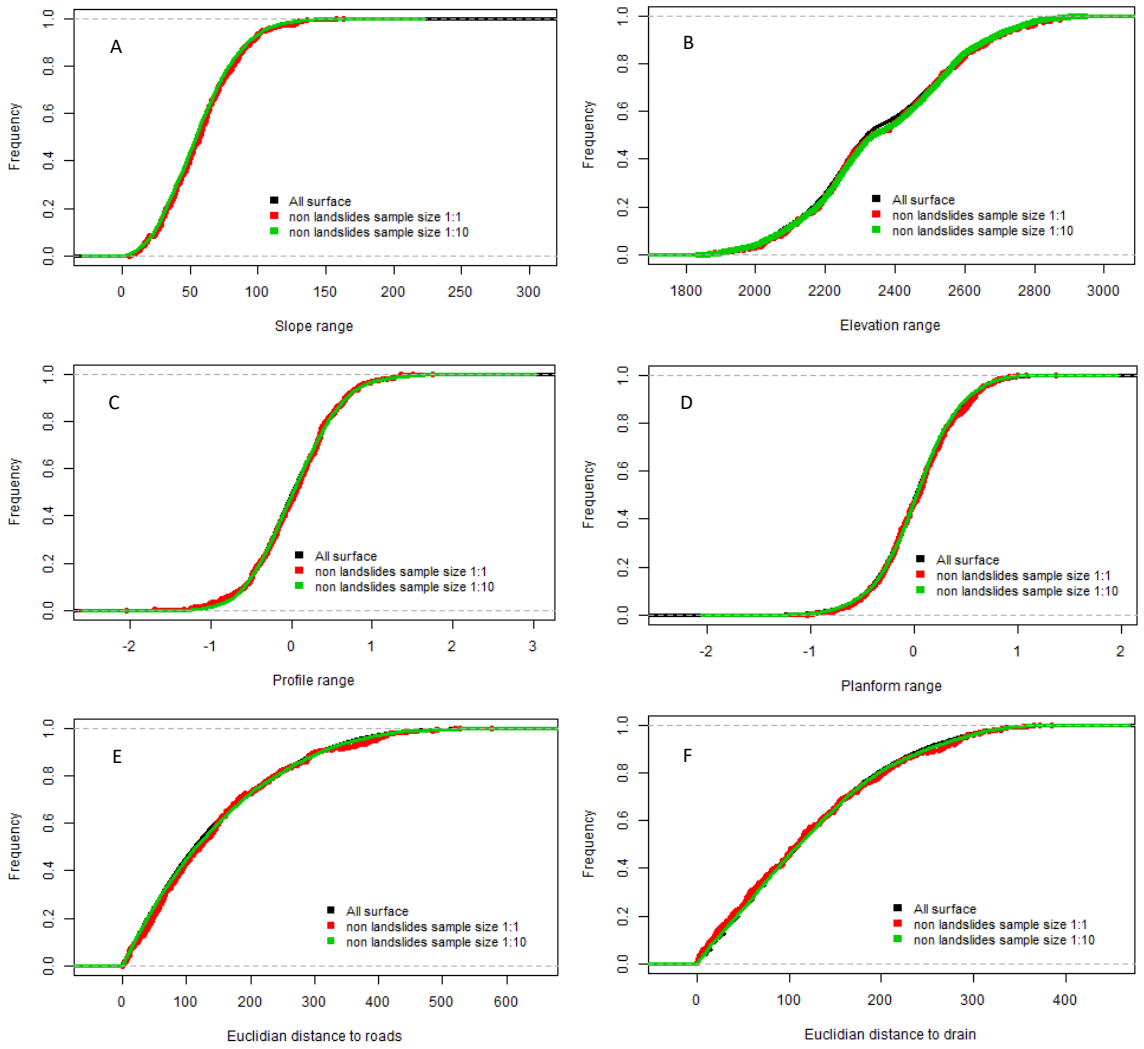


Figure 37 — Empirical cumulative distribution function for all research surface vs non-shallows landslides, sample size 1:1 and 1:10, 6 varaibles included. A) Slope. B) Elevation. C) Profile. D) Planform. E) Distance to roads. F) Distance to drain.

In the case of non-landslide events 1:1 and the variable distribution in the study area, the *P*-values reveals that there is no significant difference between the distributions of the variables, it is reasonable to assume that they come from the same distribution.

In the case of non-landslide events 1:10, the variables slope and elevation, *P*-values refers that there is significant difference between the distributions. For others variables, there is no significant difference between both of them.

As non-landslides sample size 1:1 is not biased from the distribution study area with respect to all variables we started the calibration models with this dataset. Nevertheless the non-landslide events 1:10 dataset is analyzed distinctively to detect potential performance variations.

5.3.3 Shallow landslides results (sample size 1:1)

Three models with different numbers of input variables (11, 6 and 5) have been calibrated. The results for this dataset are given in Table 9. The susceptibility maps derived from the model results "6 variables" are displayed in Figure 38.

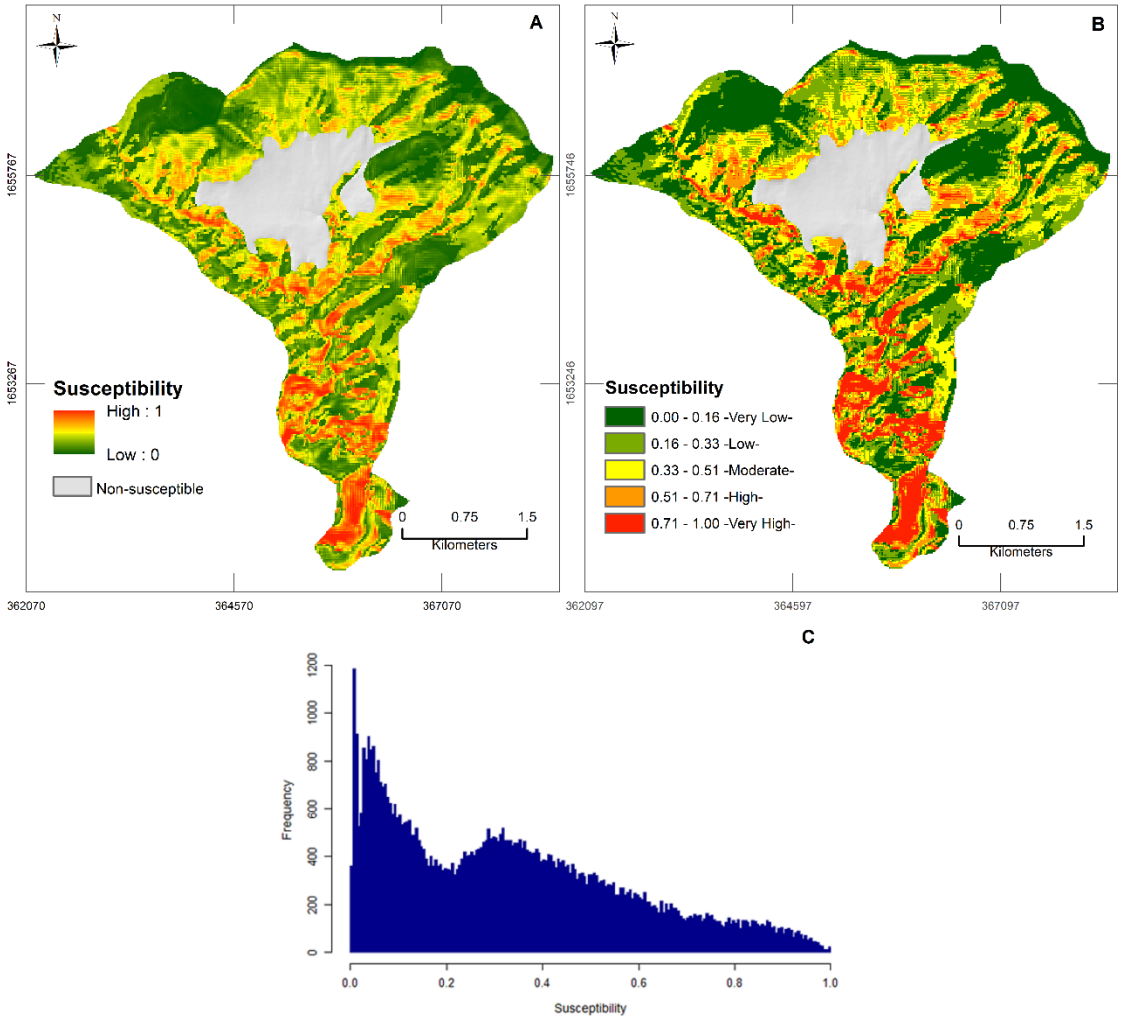


Figure 38 — Susceptibility maps Shallow landslides sample size 1:1. Comparison of probabilities visualization. A) Continuous variable. B) Naturals breaks (Jenks). C) Distribution of probabilities pixels values in study area.

The area under the ROC curve (calculated with the 20 % subset) for model “11 variables” is equal to 0.809, for model “6 variables is equal to 0.804, for model “5 variables” is equal to 0.806 and for temporal validation “2011” is equal to 0.707. The ROC curve plot for validation models is presented in Figure 39.

As the calibration models with 11 and 6 variables present a slightly difference in the area under the ROC curve, we use the “6 variables” model with an AUC=0.804 for the prediction, due to the importance incorporation of the distance to roads in this model.

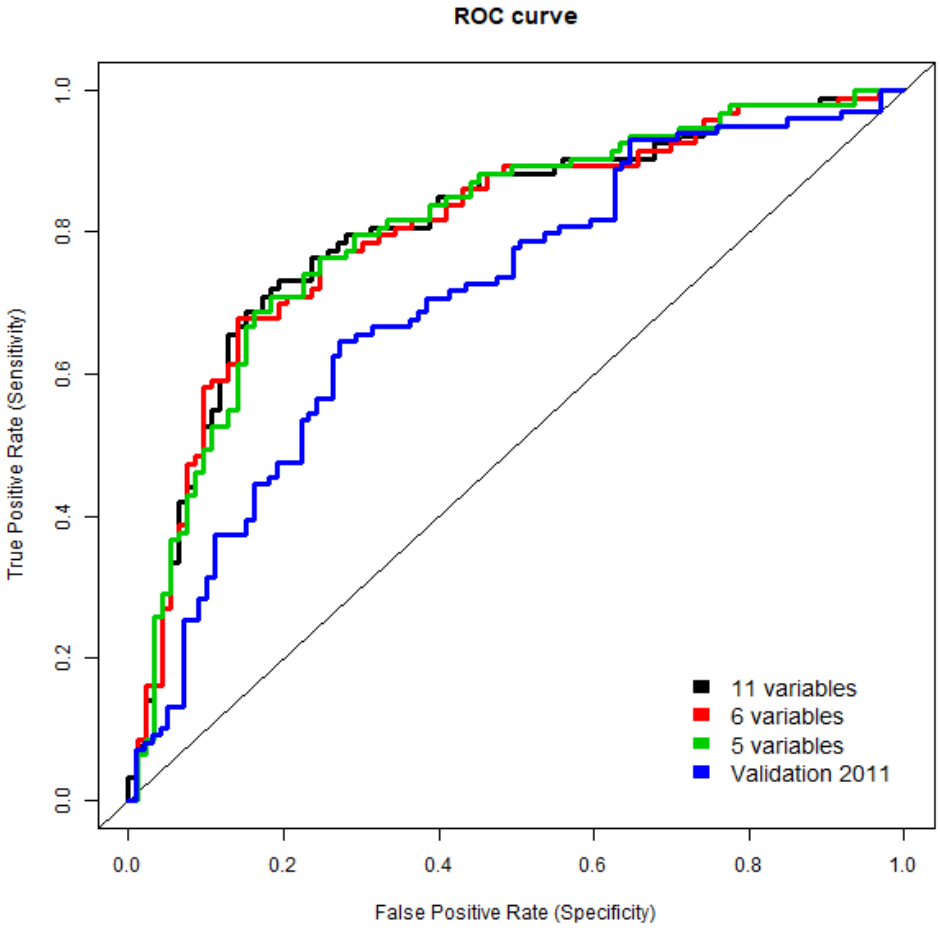


Figure 39 — ROC Curves comparison with 11, 6 and 5 input variables in calibrated models and temporal validation 2011, shallow landslides sample size 1:1.

Table 9 — Summary of logistic regression results for shallow landslides (sample size 1:1) for 3 models with different numbers of input variables.

Variables	11 variables			6 variables			5 variables		
	β	z value	Pr(>|z|)	β	z value	Pr(>|z|)	β	z value	Pr(>|z|)
Intercept	4.897	3.75	0.000177 ***	4.483	3.54	0.000401 ***	4.214	3.383	0.000717 ***
Slope	0.028	6.938	3.98e-12 ***	0.029	7.869	3.57e-15 ***	0.029	7.857	3.93e-15 ***
Aspect									
North	Reference category								
North - East	1.119	1.24	0.215068	1.063	1.192	0.233428	1.019	1.157	0.247183
East	1.806	2.037	0.041643 *	1.709	1.965	0.049469 *	1.725	2.008	0.044621 *
South - East	2.41	2.806	0.005017 **	2.401	2.824	0.004744 **	2.414	2.876	0.004030 **
South	2.578	2.994	0.002753 **	2.597	3.044	0.002331 **	2.571	3.054	0.002256 **
South - West	2.531	2.939	0.003297 **	2.524	2.949	0.003188 **	2.451	2.906	0.003663 **
West	1.337	1.527	0.126814	1.368	1.573	0.115754	1.416	1.65	0.099016 .
North - West	0.796	0.872	0.383076	0.811	0.893	0.372088	0.756	0.842	0.399642
Profile	-0.536	-2.801	0.005090 **	-0.618	-3.676	0.000237 ***	-0.529	-3.224	0.001266 **
Planform	-0.68	-3.044	0.002334 **	-0.646	-3.018	0.002548 **			no
Fragmentation	-0.059	-0.531	0.595089			no			no
Lithology						no			no
Volcanic rocks	Reference category								
Pumice ash landfills	-0.044	-0.223	0.823833						
Land cover						no			no
Forest	Reference category								
Urban	-14.8	-0.027	0.97885						
Crops	-0.203	-0.905	0.365721						
Natural pastures	-0.623	-1.359	0.174049						
Roads	-14.42	-0.01	0.992093						
Elevation	-0.0035	-7.828	4.98e-15 ***	-0.0034	-8.103	5.38e-16 ***	-0.0033	-7.956	1.78e-15 ***
CDA	-8.79E-05	-0.759	0.447698			no			no
Distance to drains	4.18E-04	0.317	0.750923			no			no
Distance to roads	-0.0035	-3.79	0.000151 ***	-0.0032	-3.639	0.000274 ***	-0.003	-3.431	0.000601 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

No: Variable not included in the initial dataset for calibrated model.

5.3.4 Shallow landslides dataset results (sample size 1:10)

The models that have been calibrated in logistic regression included 11, 6 and 5 input variables. The results for this dataset are given in Table 10. The susceptibility maps derived from the model results “6 variables” are displayed in Figure 40.

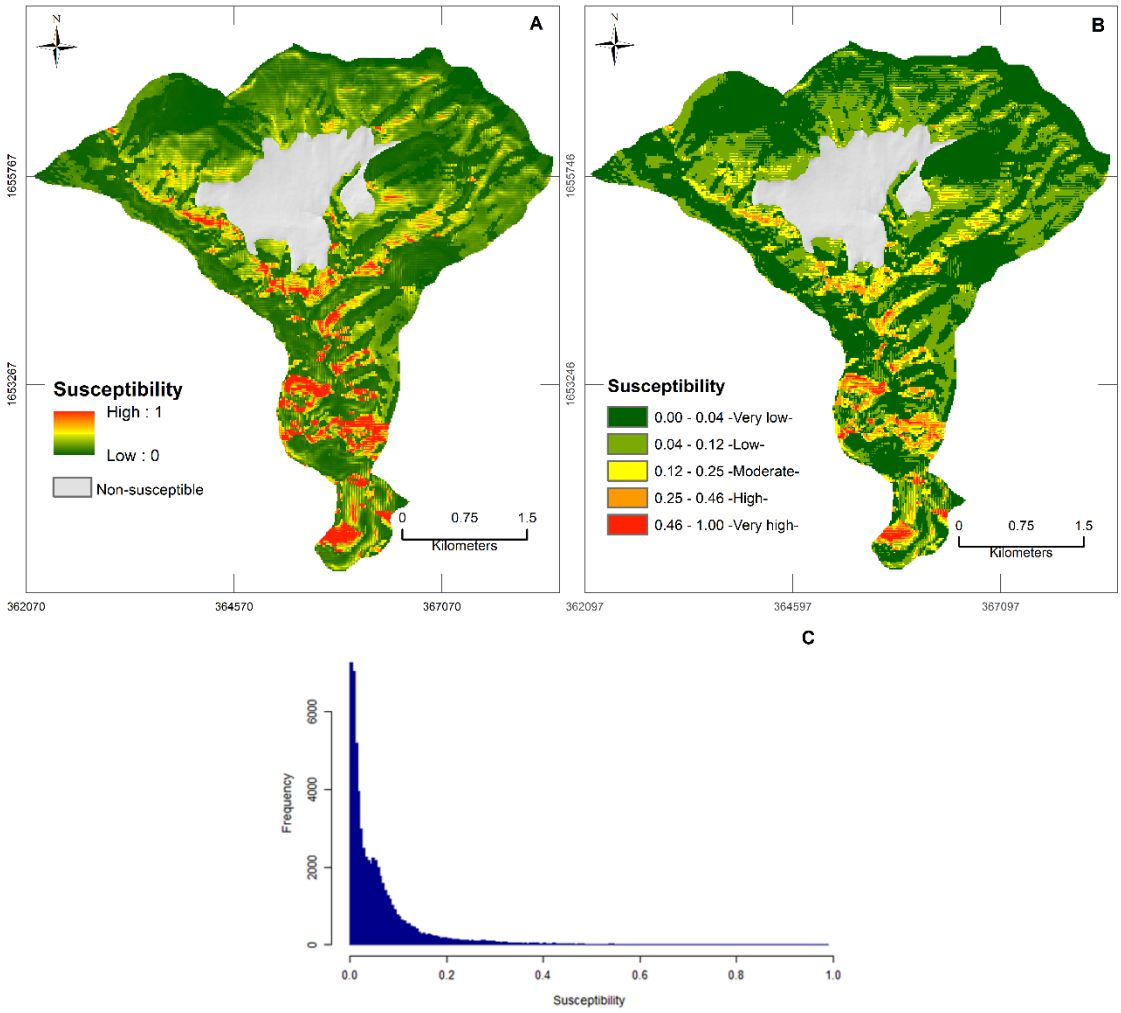


Figure 40 — Susceptibility map Shallow landslides sample size 1:10. . Comparison of probabilities visualization. A) Continuous variable. B) Naturals breaks (Jenks). C) Distribution of probabilities pixels values in study area.

The area under the ROC curve (calculated with the 20 % subset) for 11 variables model is equal to 0.827, for 6 variables model is equal to 0.821, for 5 variables model is equal to 0.818 and for temporal validation "2011" is equal to 0.679. The ROC curve plot for validation models is presented in Figure 41.

Because the calibration models with 11 and 6 variables are similar to sample size 1:1 and present almost the same performance, we take "6 variables" model (AUC= 0.821) for prediction.

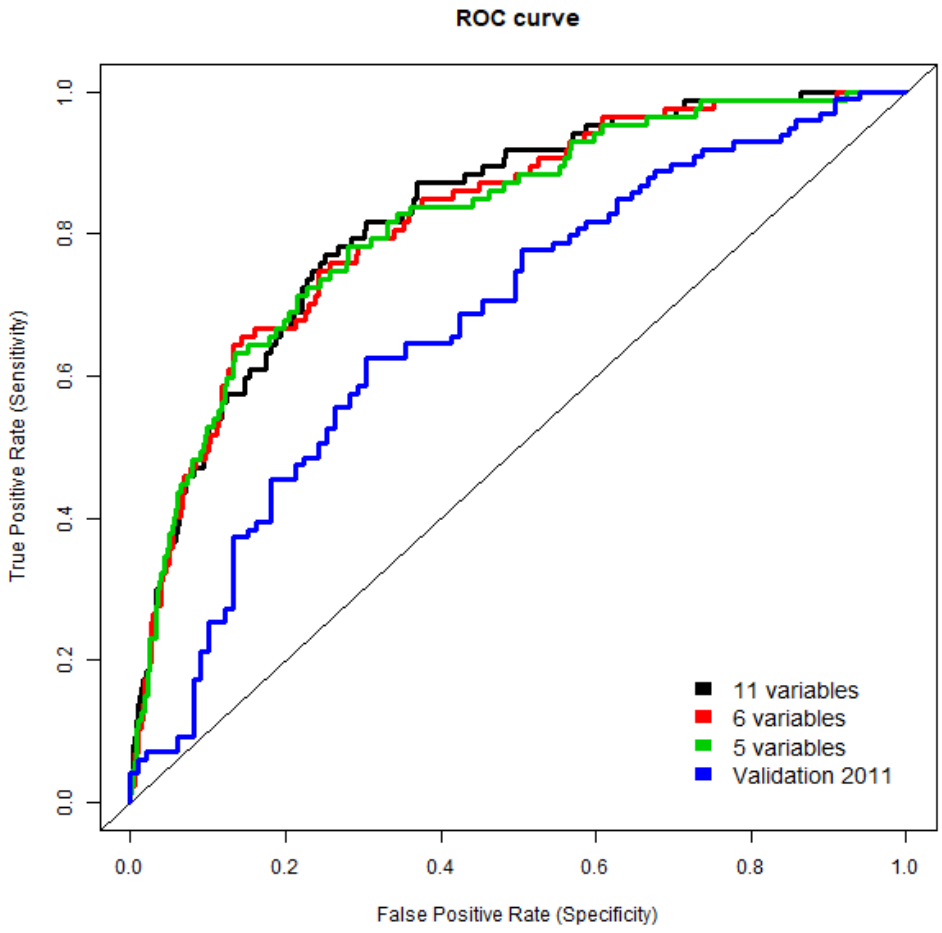


Figure 41 — ROC Curves comparison with 11, 6 and 5 input variables in calibrated models and temporal validation 2011, Shallow landslides sample size 1:10.

Table 10 — Summary of logistic regression results shallow landslides (sample size 1:10) for 3 models with different numbers of input variables.

Variables	11 variables			6 variables			5 variables		
	β	z value	Pr(>|z|)	β	z value	Pr(>|z|)	β	z value	Pr(>|z|)
Intercept	-11.5	-0.03	0.976	2.363	2.899	0.00374 **	2.224	2.737	0.0062 **
Slope	0.021	9.522	< 2e-16 ***	0.023	10.887	< 2e-16 ***	0.023	11.029	< 2e-16 ***
Aspect									
North	Reference category								
North - East	0.713	1.285	0.199	0.683	1.236	0.21651	0.668	1.208	0.227
East	0.749	1.468	0.142	0.665	1.309	0.19049	0.637	1.254	0.21
South - East	2.45	5.17	2.34e-07 ***	2.321	4.926	8.40e-07 ***	2.286	4.853	1.22e-06 ***
South	2.567	5.42	5.96e-08 ***	2.449	5.193	2.07e-07 ***	2.434	5.158	2.50e-07 ***
South - West	2.595	5.468	4.56e-08 ***	2.439	5.163	2.43e-07 ***	2.457	5.198	2.01e-07 ***
West	1.104	2.226	0.026 *	0.949	1.92	0.05492 .	0.967	1.956	0.0504 .
North - West	0.193	0.348	0.728	0.115	0.208	0.83513	0.0944	0.171	0.8644
Profile	-0.761	-6.255	3.97e-10 ***	-0.749	-7.096	1.29e-12 ***	-0.757	-7.163	7.92e-13 ***
Planform	-0.179	-1.203	0.229			no			no
Fragmentation	-0.078	-0.947	0.343			no			no
Lithology						no			no
Volcanic rocks	Reference category								
Pumice ash landfills	0.17	1.264	0.206	0.218	1.643	0.10042			no
Land cover						no			no
Forest	Reference category								
Urban	0.608	0.001	0.999						
Crops	13.74	0.036	0.971						
Natural pastures	14.424	0.037	0.97						
Roads	13.43	0.035	0.972						
Elevation	-0.0032	-10.427	< 2e-16 ***	-0.0033	-10.874	< 2e-16 ***	-0.0032	-10.714	< 2e-16 ***
CDA	-2.01E-04	-0.769	0.442			no			no
Distance to drains	-4.13E-04	-0.46	0.646			no			no
Distance to roads	-0.00392	-5.971	2.35e-09 ***	-0.0034	-5.334	9.61e-08 ***	-0.0034	-5.35	8.81e-08 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

No: Variable not included in the initial dataset for calibrated model.

5.3.5 Debris flows dataset results

Three models with 11, 6 and 5 input variables have been calibrated in logistic regression. The results for this dataset are given in table 11. The susceptibility maps derived from the model results “6 variables” are displayed in Figure 42.

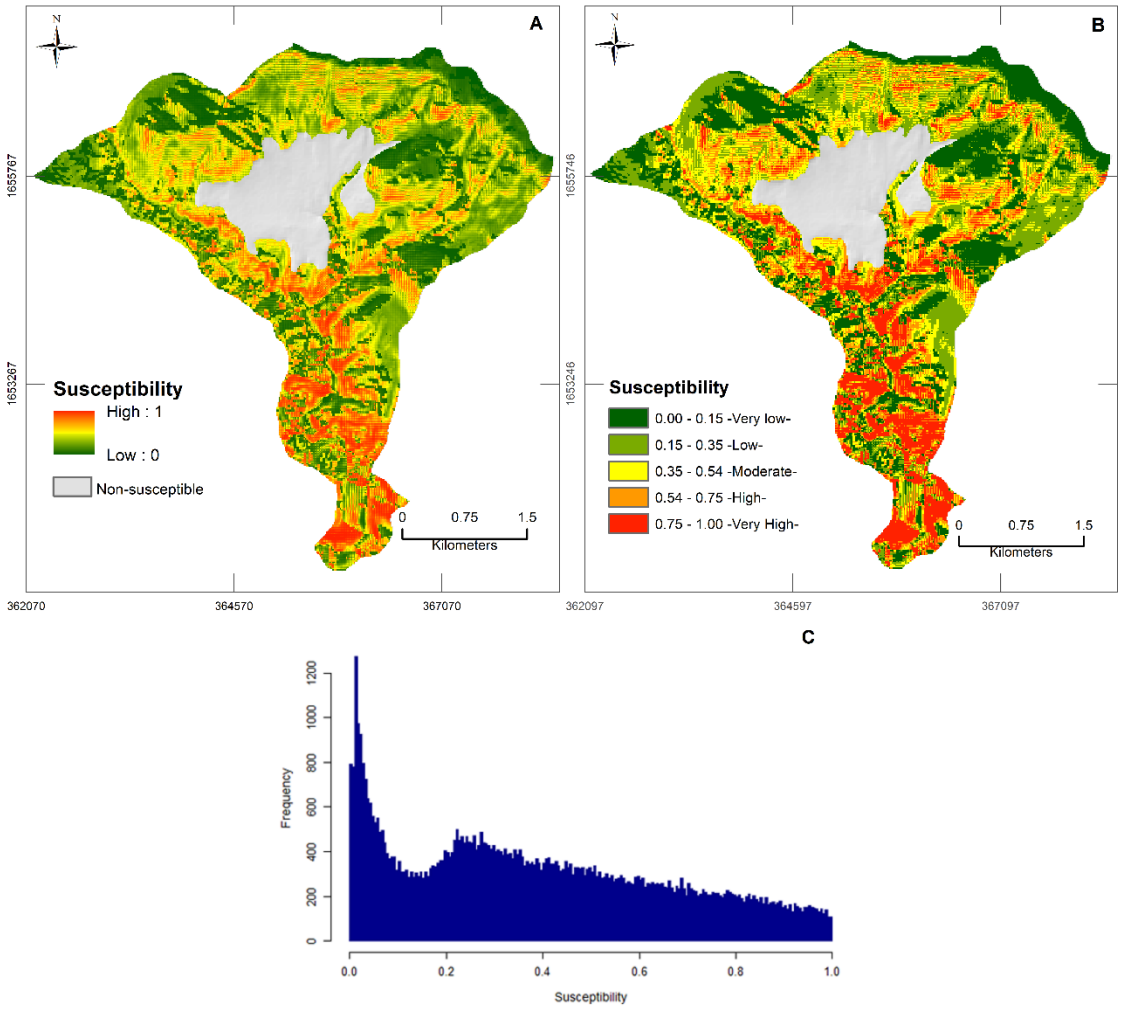


Figure 42 — Susceptibility maps, debris flows dataset. Comparison of probabilities visualization. A) Continuous variable. B) Naturals breaks (Jenks). C) Distribution of probabilities pixels values in study area.

The area under the ROC curve (calculated with the 20 % subset) for model “11 variables” is equal to 0.783, for “6 variables” is equal to 0.777, and for “5 variables” is equal to 0.777. The ROC curve plot for validation models is presented in Figure 43.

The performance is evaluated in the same way as for shallow landslide, to avoid the complexity of the model and comparing the AUC that are similar between the three models, the calibration model with 5 variables all of them topographic, is used to predict.

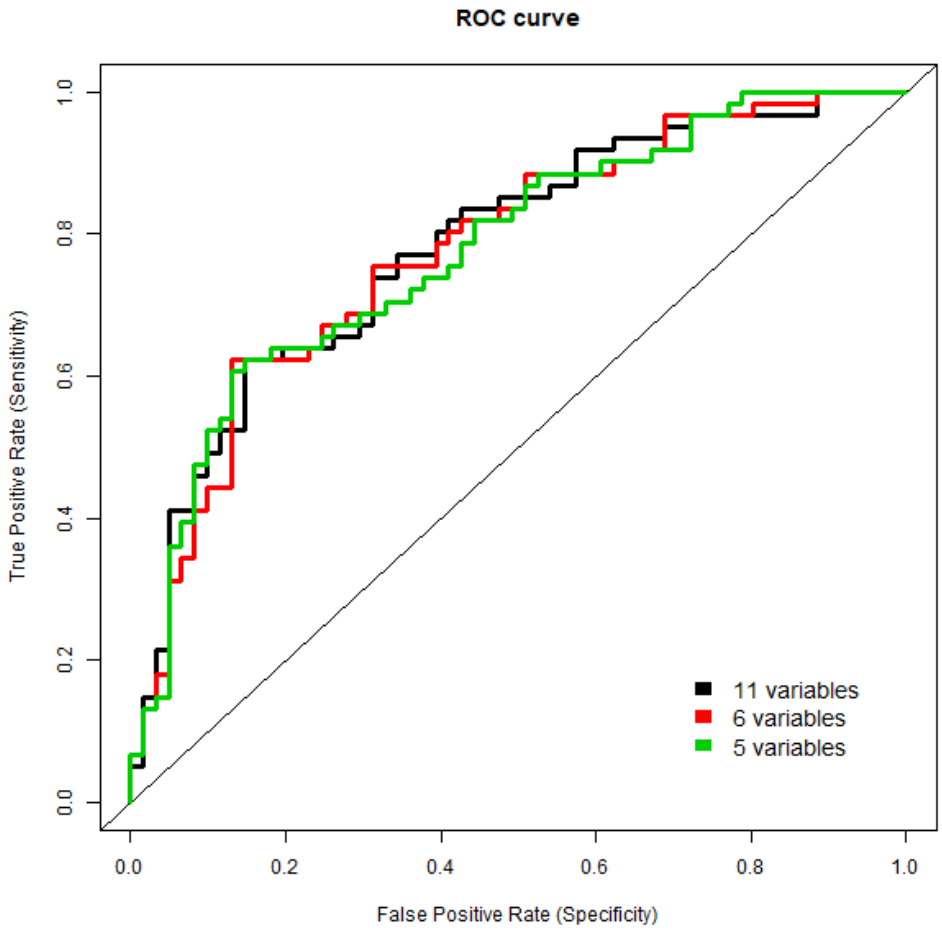


Figure 43 — ROC Curves comparison with 11, 6 and 5 input variables in calibrated models, debris flows dataset.

Table 11 — Summary of logistic regression results debris flows dataset for 3 models with different numbers of input variables.

Variables	11 variables			6 variables			5 variables		
	β	z value	Pr(>|z|)	β	z value	Pr(>|z|)	β	z value	Pr(>|z|)
Intercept	4.862	3.223	0.001266 **	5.373	4.006	6.17e-05 ***	4.782	3.667	0.000246 ***
Slope	0.027	5.14	2.75e-07 ***	2.785	5.326	1.00e-07 ***	0.031	6.251	4.09e-10 ***
Aspect									
North	Reference category								
North - East	2.525	2.913	0.003582 **	2.509	3.027	0.002470 **	2.625	3.223	0.001268 **
East	1.893	2.207	0.027301 *	1.633	1.99	0.046586 *	1.5	1.848	0.064575 .
South - East	3.779	4.675	2.93e-06 ***	3.579	4.674	2.95e-06 ***	3.531	4.611	4.00e-06 ***
South	3.378	4.268	1.97e-05 ***	3.202	4.271	1.94e-05 ***	3.33	4.446	8.74e-06 ***
South - West	2.912	3.681	0.000232 ***	2.836	3.776	0.000159 ***	3.01	4.02	5.83e-05 ***
West	2.761	3.396	0.000684 ***	2.826	3.618	0.000296 ***	3.037	3.902	9.55e-05 ***
North - West	0.889	1.006	0.314196	0.691	0.825	0.409145	0.735	0.882	0.377915
Profile	-1.071	-4.25	2.13e-05 ***	-1.09	-4.568	4.91e-06 ***	-1.092	-4.613	3.96e-06 ***
Planform	0.824	2.688	0.007182 **	0.669	2.487	0.012878 *	0.657	2.48	0.013141 *
Fragmentation	0.338	2.26	0.023829 *			no			no
Lithology						no			no
Volcanic rocks	Reference category								
Pumice ash landfills	0.209	0.796	0.425966						
Land cover									no
Forest	Reference category								
Urban	-14.91	-0.024	0.980947	-13.71	-0.022	0.98248			
Crops	-0.655	-2.142	0.032219 *	-0.584	-2.019	0.043445 *			
Natural pastures	-1.665	-2.153	0.031347 *	-1.4	-1.849	0.064505 .			
Roads	0.1615	0.127	0.898955	1.118	0.97	0.332216			
Elevation	-0.0043	-6.985	2.84e-12 ***	-0.0042	-7.34	2.14e-13 ***	-0.004	-8.066	7.29e-16 ***
CDA	4.54E-04	0.956	0.339292			no			no
Distance to drains	3.17E-04	0.192	0.847708			no			no
Distance to roads	0.0014	1.339	0.180712			no			no

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

No: Variable not included in the initial dataset for calibrated model.

5.3.6 All landslides dataset results

The models with 11, 6 and 5 input variables have been calibrated in logistic regression. The results for this dataset are given in Table 12. The susceptibility maps derived from the model results “6 variables” are displayed in Figure 44.

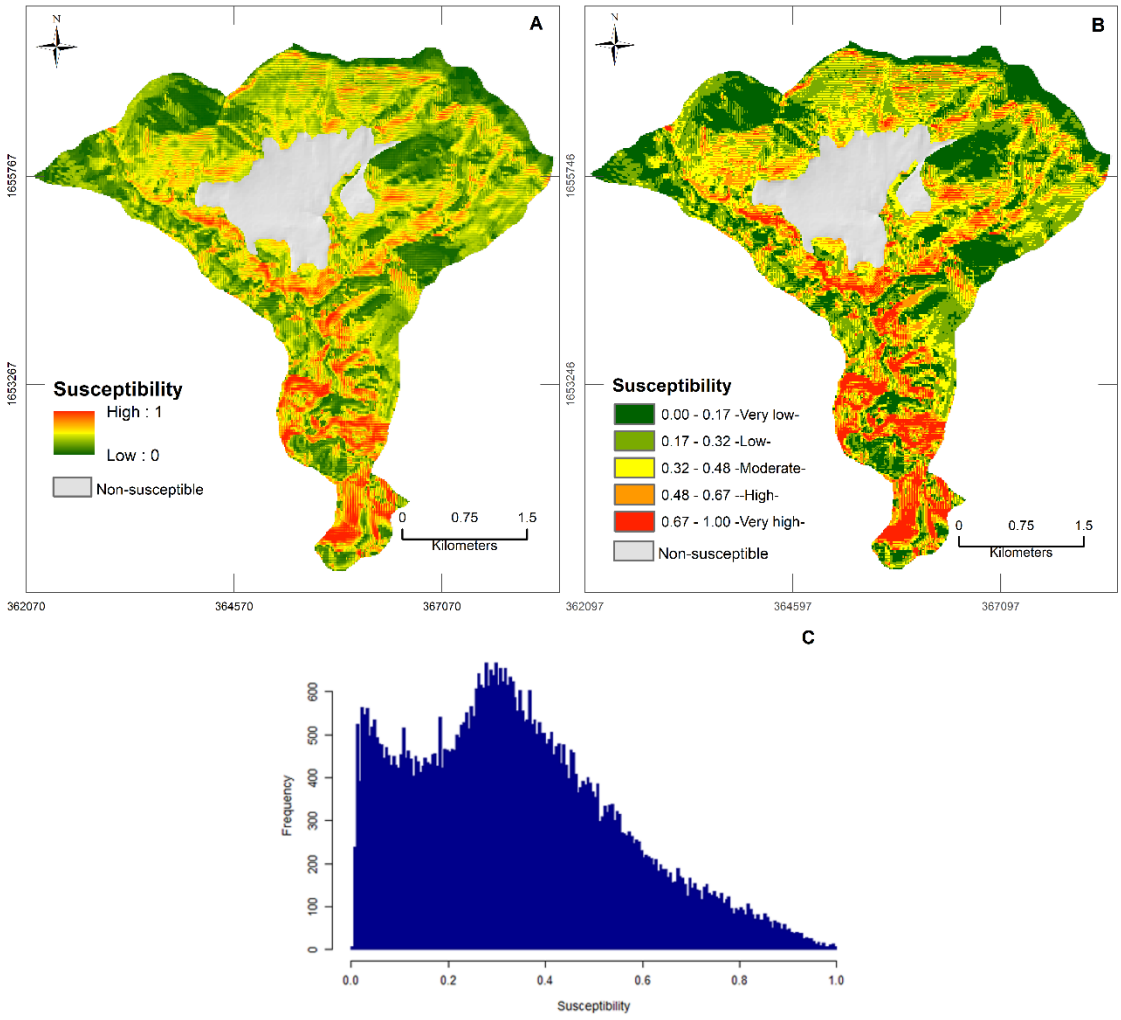


Figure 44 — Susceptibility maps, all landslides dataset. Comparison of probabilities visualization. A) Continuous variable. B) Optimal cutoff value associated to cost. C) Naturals breaks (Jenks).

The Area under the ROC curve (calculated with the 20 % subset) for model “11 variables” is equal to 0.834, for “6 variables is equal to 0.827, and for “5 variables” is equal to 0.828. The ROC curve plot for validation models is presented in Figure 45.

As the AUC between the three models are similar, and to avoid the complexity, the model with the less number of input variables (5 variables) was used to predict.

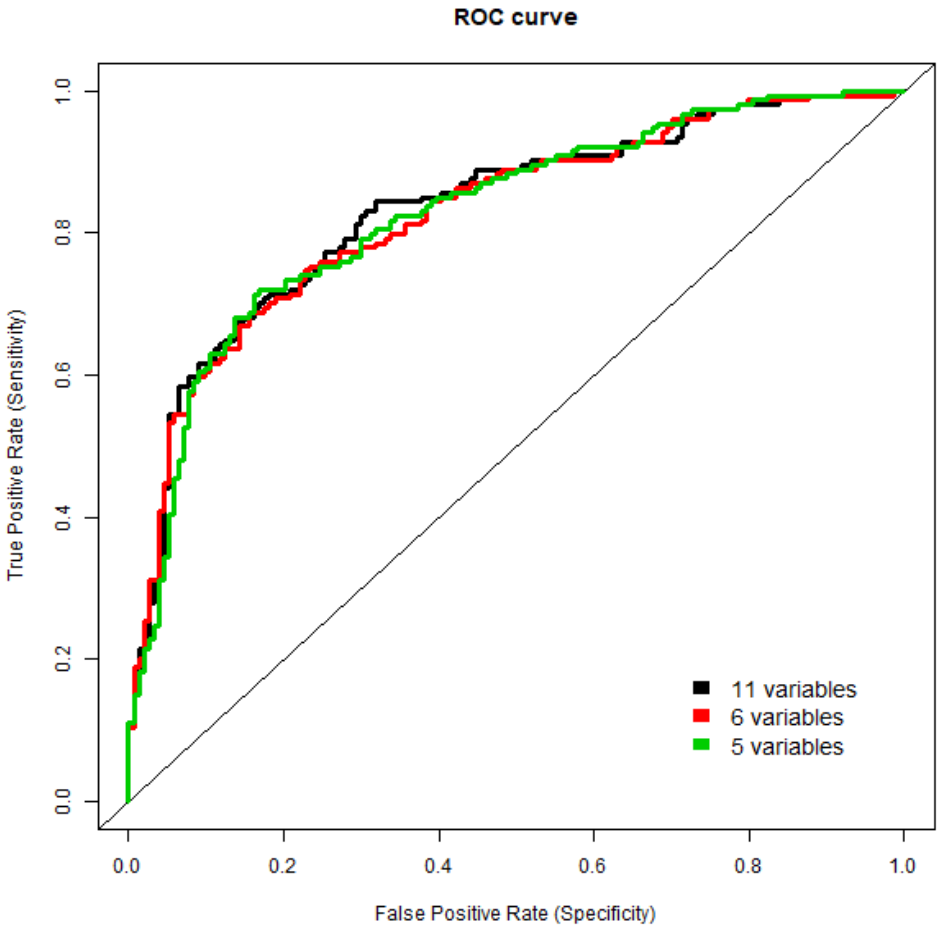


Figure 45 — ROC Curves comparison with 11, 6 and 5 input variables in calibrated models, all landslides dataset.

Table 12 — Summary of logistic regression results all landslides dataset for 3 models with different numbers of input variables.

Variables	11 variables			6 variables			5 variables		
	β	z value	Pr(>|z|)	β	z value	Pr(>|z|)	β	z value	Pr(>|z|)
Intercept	3.122	3.673	0.000239 ***	2.864	3.488	0.000487 ***	2.237	2.803	0.00507 **
Slope	0.0216	7.845	4.33e-15 ***	0.022	8.047	8.47e-16 ***	0.023	8.932	< 2e-16 ***
Aspect									
North	Reference category								
North - East	1.014	1.976	0.048209 *	1.087	2.11	0.034851 *	1.114	2.186	0.02884 *
East	1.537	3.073	0.002119 **	1.557	3.099	0.001944 **	1.499	3.018	0.00255 **
South - East	2.563	5.475	4.38e-08 ***	2.59	5.505	3.70e-08 ***	2.577	5.538	3.06e-08 ***
South	2.718	5.836	5.34e-09 ***	2.762	5.902	3.60e-09 ***	2.777	5.983	2.19e-09 ***
South - West	2.363	5.033	4.84e-07 ***	2.447	5.208	1.90e-07 ***	2.483	5.328	9.92e-08 ***
West	1.7	3.536	0.000407 ***	1.799	3.736	0.000187 ***	1.893	3.963	7.41e-05 ***
North - West	0.398	0.751	0.452394	0.437	0.823	0.410595	0.509	0.967	0.33368
Profile	-0.852	-5.792	6.94e-09 ***	-0.928	-7.204	5.87e-13 ***	-0.939	-7.401	1.35e-13 ***
Planform	0.105	0.642	0.520983			no			no
Fragmentation	0.048	0.62	0.535223			no			no
Lithology						no			no
Volcanic rocks	Reference category								
Pumice ash landfills	0.303	2.008	0.044671 *						
Land cover									no
Forest	Reference category								
Urban	-11.53	-0.035	0.971688	-11.31	-0.035	0.972221			
Crops	-0.297	-1.764	0.077740 .	-0.341	-2.088	0.036803 *			
Natural pastures	-1.295	-3.082	0.002053 **	-1.318	-3.155	0.001606 **			
Roads	0.447	0.564	0.572866	0.505	0.665	0.50613			
Elevation	-0.0028	-8.459	< 2e-16 ***	-0.0027	-8.449	< 2e-16 ***	-0.0025	-8.066	7.29e-16 ***
CDA	-4.92E-05	-0.91	0.36302			no			no
Distance to drains	2.86E-04	0.278	0.781095			no			no
Distance to roads	-0.0018	-2.76	0.005788 **	-0.0019	-3.061	0.002203 **	-0.0018	-3.03	0.00244 **

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

No: Variable not included in the initial dataset for calibrated model.

6 GENERAL DISCUSSION

Guatemala is a country at risk from multiple hazards, most especially from landslides triggered by tropical cyclones. However, for a country where landslides are common, assessments of landslide susceptibility are not yet developed. The identification and analysis of landslide predisposing factors is crucial, particularly in populated mountainous areas.

The issues in this research are multiple in terms of the complexity of inventory, quality of data, accuracy of assessment, and the practical uses of the results for various stakeholders.

To reach the objectives, a landslide inventory was made based on orthoimages interpretation and on fieldworks with local guides who helped in the recognition of areas that experienced landslides. Then, “landslide types were analyzed, including their possible predisposing factors. Moreover, to calibrate the predictive models, the susceptibility analysis was made using a multivariate probabilistic approach (logistic regression). To validate the performance of the model results, ROC curve, four-fold and confusion matrix plots were used. A susceptibility map was also generated to display the results of models in terms of probability values.

This research highlighted the six most important variables that impact the susceptibility predicted model. They are by order of importance: elevation, slope, aspect, profile curvature, planform curvature, and distance to roads. Landslide occurrences decrease with elevation; increase with the slope gradient; increase for slopes oriented to the south; are lower for concave slopes; and increase with the distance to roads. Planform curvature influence depends on the landslide types: for shallow landslides convergent curvature decreases the occurrence, whereas for debris flows convergent curvature increases the occurrence.

In this section, the different choices made in this study are explained. First, the technical choices related to the quality of data will be discussed, followed by the analysis of the impacts of the geomorphological variables. Finally, propose and discuss an operational use for the susceptibility maps.

6.1 Technical choices

Given all the characteristics of the studied landslide phenomena in the Guatemalan region, some choices have been made to make these complex processes more understandable especially among the non-experts.

The most important data to predict landslide is the variable related to topography which are derived from DEM, provided in this study by the Guatemalan Geographical National Institute. The mean size of shallow landslides and debris flows are 82 m^2 and 1050 m^2, respectively. However, the pixel size of the DEM is 15x15 m (225 m^2) which means that it could be smaller or bigger than the landslide size according with its type. Therefore, the size of the landslides is not well represented in the model. Only the highest point was taken to represent the landslide and each one takes the elevation of the pixel of the DEM. This influences the accuracy of the modelization process of landslide susceptibility and the final results for the susceptibility maps. However, this choice allows to compare the influence of the different predisposing factors in the landslide starting zone (Simon *et al*., 2017).

The DEM was used to derive the topographic characteristics such as slope, elevation, planform and profile curvature, contributing drainage area, and aspect. As a consequence, DEM errors affect the accuracy of the variables, especially the first (slope) and second (curvature) derivatives. For example, as we can see on Figure 13, pattern errors of profile curvature exist everywhere on the map but especially on the top portion. In the study area, the slope gradient was underestimated due to the DEM errors and the pixel size. Rather than an instant slope gradient, a mean slope gradient every 15 m was obtained. This error was also observed in the study of gullies prediction in a watershed in Northern Algeria by Dewitte *et al*. (2015) and in the study of Zhu *et al*. (2013). Both studies discussed the accuracy impacts caused by DEM errors concerning elevation and slope variables.

The landslides identified from the 2005 inventory were classified in two different types of slide mechanism (shallow landslides and debris flows) according to the mean area. Due to the inventory containing two different landslide types, each had to be modelized separately. The process is complex to make a classification with only this characteristic but the field data (and knowledge) justified the classification process.

It is important to emphasize that debris flows have more impacts in the valleys as they follow the thalwegs. The length and direction of flow may vary considerably and can occupy many pixels from the initial point (head) to the final impact. This adds weight to the usefulness of representing landslides by one single point. In the susceptibility maps, the areas classified as highly susceptible show only the unstable starting area but in the case of debris flows it could form a very large downstream flow causing also high impacts, and damage to the population. As we can see in the Figure 22, the flows can travel distances from a few meters up to kilometers.

In the case of the 2011 landslide inventory, the coordinates of the landslide heads were measured with a GPS. This was used only to validate the performance of the predictive model. Due to the fast growing vegetation (tropical climate) and to the preparation of field for crops, the presence of flows that characterize the slide mechanism was not easy to notice. Therefore, it could be possible that some of landslide cases were classified incorrectly. Nevertheless, as no further impacts in the valleys were observed, and as validated by the local field guides, it can be assumed that the flows were not big enough to be classified as debris flows.

As explained above, two different landslides types are included in the inventory. Models have been created for shallow landslides, debris flows, and a union of both data. The last one was meant at exploring and comparing the performance of the models, given that only the landslide initial point was taken into account in the calibration. The models were constructed with logistic regression method considering twelve predictor variables. In order to assess the level of correlation between the location of the landslides in the study area and the predisposing factor selected, the frequency ratio analysis was carried out previously to the logistic regression analysis. The results have shown that variables such as slope and elevation have the highest ratio (Table 5). For the other variables, at least one discretized class correlates with the occurrence of landslides. To calibrate the models in the logistic regression, 80 % of each dataset was used, while the other 20 % was used to evaluate the performance. In the case of shallow landslides dataset, cross validation and temporal validation with the shallow landslides events triggered in 2011 by tropical depression “Twelve-E” (99 events) was carried out. As logistic regression is

susceptible to multicollinearity, an analysis was applied to 12 variables, allowing to identify that elevation and morphometric region variables were associated. For this research, the one derived from DEM (elevation) is more suitable to be included in the logistic regression to calibrate models. Morphometric region was not included in the models. For each landslide data type, the first logistic regression started with 11 variables, and for investigating the model performances with a smaller set of variables, a second and third model including the 6 and 5 most significant variables were performed.

Two different sample sizes strategies have been used to calibrate the shallow landslide model 1:1 and 1:10. The result of the analyses showed differences among the sample sizes selected. The models selected to predict the susceptibility in the watershed are the ones containing 6 variables, with an AUC equal to 0.804 for 1:1 and 0.821 for 1:10. The model for the sample size 1:10 was slightly better if the AUCs were compared (Figure 39 and Figure 41). The performance difference between the models were due to the sample size., In the case of 1:10, the fact that the non-landslide events are ten times more frequent than the landslides cases, it could better represent the spatial variability and the combination of the predisposing factor. The performance of logistic regression models with 6 input variables evaluated with the ROC curve, shows an AUC of 0.804, and it classified correctly 74 % (of which 38 % of true positive –TP - and 36 % of true negative - TN) in sample size 1:1. The model for sample size 1:10 shows an AUC of 0.821 and classified correctly the 91 % (TP=0.60 % and TN=90.8 %). As expected it can be observed that the ratio of TN is bigger. The influence of this ratio in the probabilities distribution result can be observed in the susceptibility map that displays more values close to 0 (Figure 40). The performance of the model with sample size 1:10 was not significantly better than the sample size 1:1 model. The distribution of probabilities in the last were less unbalanced. The model with the major proportion of non-landslides event (1:10) skewed the distribution of probabilities pixels values in the susceptibility map. The frequency of pixels in the study area was clearly close to 0 as show in Figure 40C. In the case of sample size 1:1, the frequency has picks close to 0 and 0.5 but pixels values distributed along the probabilities 0 to 1 were also observed. Despite the distribution, the susceptibility map categorized to sample size 1:10 by natural breaks (Jenks) set the threshold at lower values compared with the sample size 1:1 (Figure 38B and Figure 40B).

For this research, it can be assumed that the sample size 1:1 represented well the features of the study area to calibrate and validate the model for landslides susceptibility assessment. The performance was good when it was compared to other rainfall events such as the tropical depression Twelve-E (AUC=0.707) which occurred 6 years after tropical storm Stan.

The sample size strategy influenced the susceptibility map result. However, the performance of the model did not show significant differences between the selected sample size 1:1 and 1:10 for shallow landslides. These observations were also confirmed in the study of *Heckmann et al*. (2014) who concluded that the sample size does not influence the "average predictive power". The calibration models for debris flows and the union of both was thus made with the sample size of 1:1.

The performance of debris flows dataset revealed AUC from 0.777 to 0.783. The differences were not significant (Figure 43) and can be considered as an acceptable discrimination. The model with 5 input variables reduced the complexity of the model; this model classified correctly

68 % (TP=35 % and TN=32 %). However, the distribution of predicted probability values in the whole surface tends to classify more TP (Figure 42C). The susceptibility maps displayed bigger surface of the watershed as landslide prone areas compared with shallow landslides models (Figure 42A and Figure 42B).

The union of shallow and debris flows landslides was used in logistic regression and the results showed marginal differences between the AUC's from 0.827 to 0.834. The model with 5 input variables correctly classified the 77.6 % (TP=36 and TN=41.6).

Finally, the different performances of the shallow landslides models among the cross validation (2005) and the temporal validation (2011) is discuss. The prediction was significantly less accurate for the 2011 event (AUC ~0.7) than for the cross validation with the 2005 intra-event (AUC ~0.8). First, the nature of the two events influenced different slide mechanisms. In the case of 2005, 61 % of the landslides in the inventory were related to shallow landslides and 39 % to debris flows. For 2011, all data were classified as shallow landslides which confirmed that the conditions of the events both triggered shallow landslides. The landslides mapped in 2005 were related to a tropical storm and the landslides mapped in 2011 were related to a tropical depression. These two events did not follow the same trajectory and did not induce the same amount of precipitations. The two last points also influenced the aspect variable that impact the landslide occurrences. Moreover, looking at the results of the landslide inventory, it is possible that the classification type procedure had some issues related to classification errors during the field survey that affecting the validation accuracy.

As has been discussed before above, the AUC between the different datasets were quite similar and can be considered as an excellent discrimination following the general rule (AUC= 0.8-0.9: excellent; AUC=0.7-0.8: acceptable) (Hosmer & Lemeshow, 2000). The validation with the temporal dataset AUC = 0.707 can be taken as an acceptable discrimination. The results revealed the good predictive capability of the models in terms of spatial and temporal prediction of landslides phenomena.

The predicted landslides susceptibility was presented in two different susceptibility maps. The first map contain the continuous values of probability (0 to 1) (Figure 38A, Figure 40A, Figure 42A and Figure 44A) that is certainly more precise (Trigila *et al*., 2015). The second map applied natural breaks (Jenks) (Iovine *et al*., 2013) with five classes aiming at reducing the variance within classes and maximizing the variance between the classes. This discretization allowed to establish five classes of susceptibility for a better understanding of the probability values (Figure 38B, Figure 40B, Figure 42B and Figure 44B). The distribution of probabilities to landslide occurrences in the study area can be seen next to the maps.

6.2 Geomorphological discussion

Three different numbers of input variables were used to calibrate the models. The validation process showed that ROC and AUC had small variations in the three datasets, as said in previous section. However, the logistic regression in shallow landslides with 6 input variables (elevation, slope, aspect, profile curvature, planform curvature, and distance to roads) discriminated better the susceptibility areas. This model demonstrated the importance of the distance to roads as a factor of slope failure, as observed during the fieldwork. In the study, it was the single

anthropogenic influence on shallow landslide occurrences. As Vega and Hidalgo (2016) and Highland and Bobrowsky (2008) wrote in their papers, common human-induced factors may initiate landslides. They mentioned particularly the oversteepening of slopes and the improper excavating or grading on slopes.

In the case of debris flows, the selected model counts 5 input variables (elevation, slope, aspect, profile curvature and planform curvature). It had the same performance as the model with 6 variables, similar to the performance of the model that included all variables, although only the topographic factors were included.

For the union of shallow landslides and debris flows, the model with 5 input variables (elevation, slope, aspect, profile curvature and distance to roads) which had the same performance as the model with 6 variables and similar to the model with the all variables, In this case, the distances to roads were taken into account, similar as to how shallows landslides were treated.

Even if the models with 5 variables for debris flows and the union of the data types were selected, it is important to point out that the same results can be obtained if the coefficients are taken for the model with 6 variables. The 5-variable model was chosen to simplify the results in order to create the susceptibility maps. It is worth mentioning that the model with 6 variables for all the datasets was more effective in discriminating the distribution of susceptibility areas as it included one more variable.

It is important to model separately the different sliding processes: the difference of the susceptibly zones in the map are explained for this reason. Moreover, the maps showed the starting zones of landslides. Taking into consideration the final local users, the map representing the union of both shallow landslides and debris flows proved to be more concise and comprehensible".

An analysis comparing the relative contributions of the explanatory variables and their geomorphological significance is presented below. Topographic variables are significant, in the case of land use, while it was expected to be an important explanatory variable, the model did not consider it as significant.

As shown in the bivariate analysis, landslide occurrence was more frequent on the steepest slopes, for lower elevations, for slopes facing to South, South-East and South-West, for convex profile curvature and for convergent planform curvature. As expected, these variables had high influence on the model results. The logistic regression result revealed that slope, elevation, aspect and profile curvature have been included systematically in all models. Distance to roads and planform curvature were included for different datasets.

The positive coefficients for slope highlight that the probability of landslides occurrence increased with steeper slopes. For elevation, the negative coefficients meant that the occurrence of landslides increased at lower elevations. Geologically, the highest points tend to be older and, therefore, more resistant. It was for this reason that the morphometric region was created based on features through different elevations.

In the study area, the amount of rainfall was the same for all studied territory but it triggered more landslides in the recent relief located in the lower part of the watershed.

The coefficients were positive for aspect dummy variables and more significant for slopes facing to South-East, to South, and to South-West. This variable had influence on the stability of the slope because moisture retention and vegetation were reflected by the slope aspect, as was also shown in the study of Ohlmacher and Davis (2003). The profile curvature coefficient was negative which meant that the upward (concave) profile decreased the probability of landslides. This is important because, as was previously mentioned in the methodology section with Ohlmacher *et al*. (2007), the profile curvature affects the acceleration or deceleration of flows along the maximum slope direction on the surface. In the case of shallow landslides particularly, the distance to roads is significant and the coefficient is negative which means that the landslides decrease with increased distance. On field, traces of landslides (triggered) near road networks were observed. The planform curvature for shallow landslides was negative which means that negative convergent (concave) planform decreased the landslide occurrences. Concerning debris flows, the planform curvature was positive which meant that positive convergent (concave) planform increased landslide occurrences. This is important because, as discussed in the methodology section with Ohlmacher *et al*. (2007), planform curvature influences the convergence or divergence of water during downhill flows. As reviewed in literature about the importance of predisposing factor for this research, in line with Van Westen *et al*. (2008) and Corominas *et al*. (2014), the factors related to topography were crucial in the statistical analysis.

In the bivariate analysis, the landslide occurrences were frequent in the forest but in the models, this land cover was not significant. However as Persichillo *et al.* (2017) highlighted, "shallow landslides was influenced by the combined effect of land use changes and the geological conditions". In the study area, sites with lower slopes are easily accessible and could be easily deforested and converted to field crops. As a consequence, forests are now only located on the steepest slopes. In this case, slope variable was more significant than land cover for landslide occurrence. From a geomorphological perspective, it can be assumed that in the study area, the amount of precipitation is high that it always triggers landslides in any land cover land cover. In this study the specific land cover change was not part of the susceptibility analysis even if the spatial environments of pixels and their temporal trajectory could influence the model results. Instead, the land cover fragmentation variable to approach the assessment of land cover change variable and to test a possible important incidence on the landslide occurrence was created. It appeared that this variable had no influences because the only class of this variable in the bivariate analysis that showed some importance was the landscape fragmentation with one fragment. This type of fragment appeared in the forest land cover as seen when comparing the Figure 17 and the Figure 18.

6.3 Displayed susceptibility maps in a practical form

The probabilistic method to quantify the landslide susceptibility showed satisfactory results. However, the results should be presented in a manner that is easily understood by local stakeholders. This could be achieved through the use of susceptibility maps. According to Beguería (2006), sometimes, final users such as land use planners, decision makers and engineers request for legends with labels as opposed to the continuous variables normally presented in maps. One approach to understand phenomena and to select the areas with the

highest and the lowest susceptibility was by choosing the cutoff associated with the cost discriminating the predicted values.

The choice of the cutoff value with the ROC curve associated with cost, allows to establish the threshold where FP (False Positive) are bigger than FN (False Negative), assuming that the cost to have FN is more dangerous for society. According to Trigila *et al*. (2015) and Corominas *et al*. (2014), False Positive misclassification errors determine a limitation in the use of the terrain due to land-use restriction and, therefore, a loss of their economic value. False Negative misclassification errors determine higher and not socially acceptable costs in case of causalities and damage to exposed elements. For example, if the model predicted a pixel as landslide (unsafe) and prevention measures costing $100 were recommended but landslide did not occur, investment is lost. . On the other hand, suppose that the model predicted a pixel as non-landslide (safe) and no prevention measures were recommended but a landslide occurs damaging infrastructure with a higher repair cost (>$100) or even causing human losses. What value should be considered for the cost? Due to the lack of socioeconomic information about the cost of committing these types of errors, a cost value according to the sample size should be empirically set. In the case of the sample with equal proportion of landslides and non-landslide events for a FP we assumed a cost of $100 and by a FN a cost of $150, thus increasing the cost to commit this error. In the case of the sample with unequal proportion of non-landslides, ten times higher than landslides events, the cost must be adjusted based on the disproportion to avoid the bias produced by the large number of non-landslides. A cost of $100 for FP and by the FN a cost of $1000 was assumed.

The Cost Plot associated with the ROC curve provided a minimum zone cost. Further, the analysis allowed to establish a minimum cost value. However, the zone with the minimum costs was used to represent an intermediate zone where the cutoff values can be chosen to define the lowest and highest landslide susceptibility (Figure 46, Figure 49, Figure 52 and Figure 55). The predicted score and observed outcome are displayed in the four-fold and confusion matrix plots (Figure 47, Figure 50, Figure 53 and Figure 56).

The maps (Figure 48, Figure 51, Figure 54 and Figure 57) contain the threshold with the cutoff values assigned to define the susceptibilities areas. Among the maps with different sample size 1:1 and 1:10, the frequency of pixels that represent landslide prone areas were quite different. More areas were classified as non-landslide susceptibility in 1:10 than in 1:1, due to the fact that more non-events were selected to calibrate model, thus considerably reducing the landslide susceptible areas.

Relevant discussions integrate the analysis of the different maps of land cover, and of susceptibility and the cost analyses. The zone displayed with high susceptibility was more frequent in the lower part of the study area. This particular area is being used for intensive agriculture, especially the vegetables such as potatoes, carrots, and Brussels sprouts which need sandy soils. Because of the soil plowing, this area is particularly susceptible to rainfall triggered landslides. The zone displayed with low susceptibility was more frequent in the higher part of the study area. In this area, the soils are more clayey under forest cover. I Susceptible places were also present but were less frequent than in the lower part of the study area.

Current trends in the growth and unplanned occupation of the villages in the study area, particularly in San Marcos department, made evident the need to regulate the construction of houses, roads or changes in land cover especially in landslide prone areas. Due to the lack of reliable information to define areas susceptible to landslides occurrences, the local governments cannot design and implement measures to reduce the vulnerability to landslides triggered by rainfall that causes considerable economic and human losses. Susceptibility analysis based on probabilistic methods as presented in this research contribute to local governments and stakeholders in urban planning.

Cutoff values associated with cost, shallow landslides (sample size 1:1)

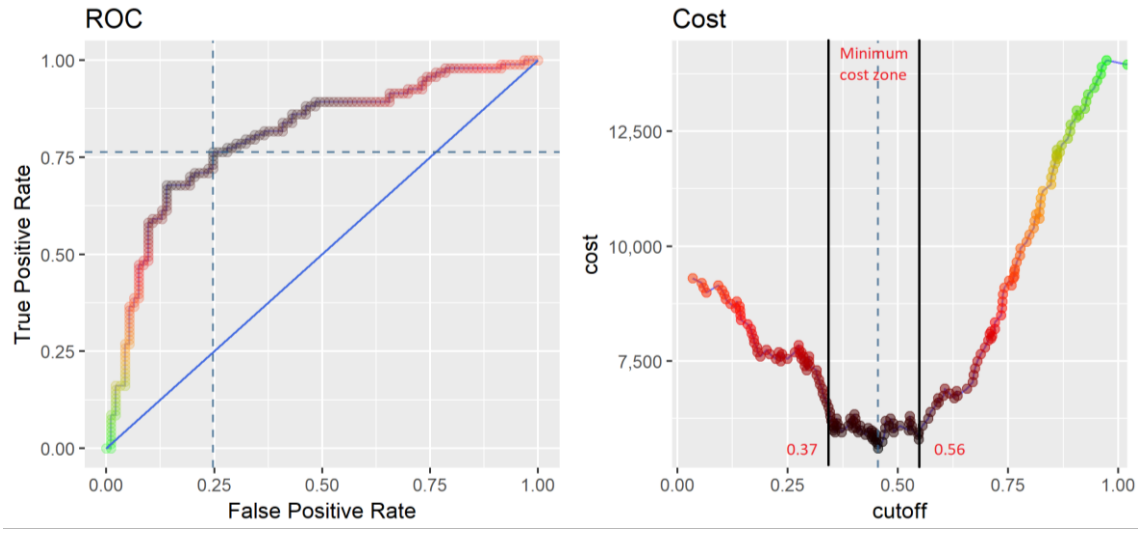


Figure 46 —ROC curve and Cost plot showing the cutoff values in the minimum cost zone, for shallow landslides sample size 1:1. Cyan color dotted line denotes where that optimal point lies and black lines denotes the zone with minimum cost.

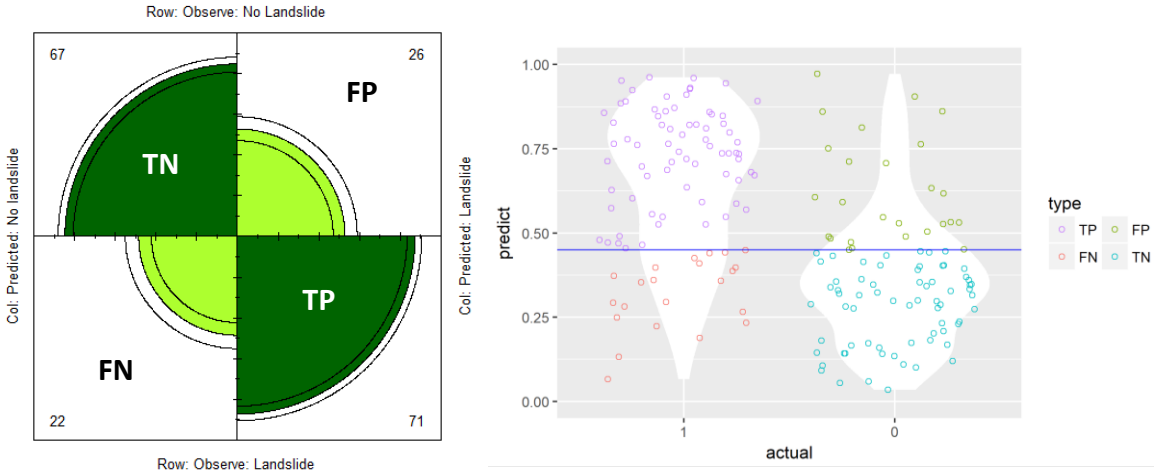


Figure 47 Four-fold and Confusion matrix plots (using minimum cutoff value = 0.45) for "6 variables" model, shallow landslides sample size 1:1.

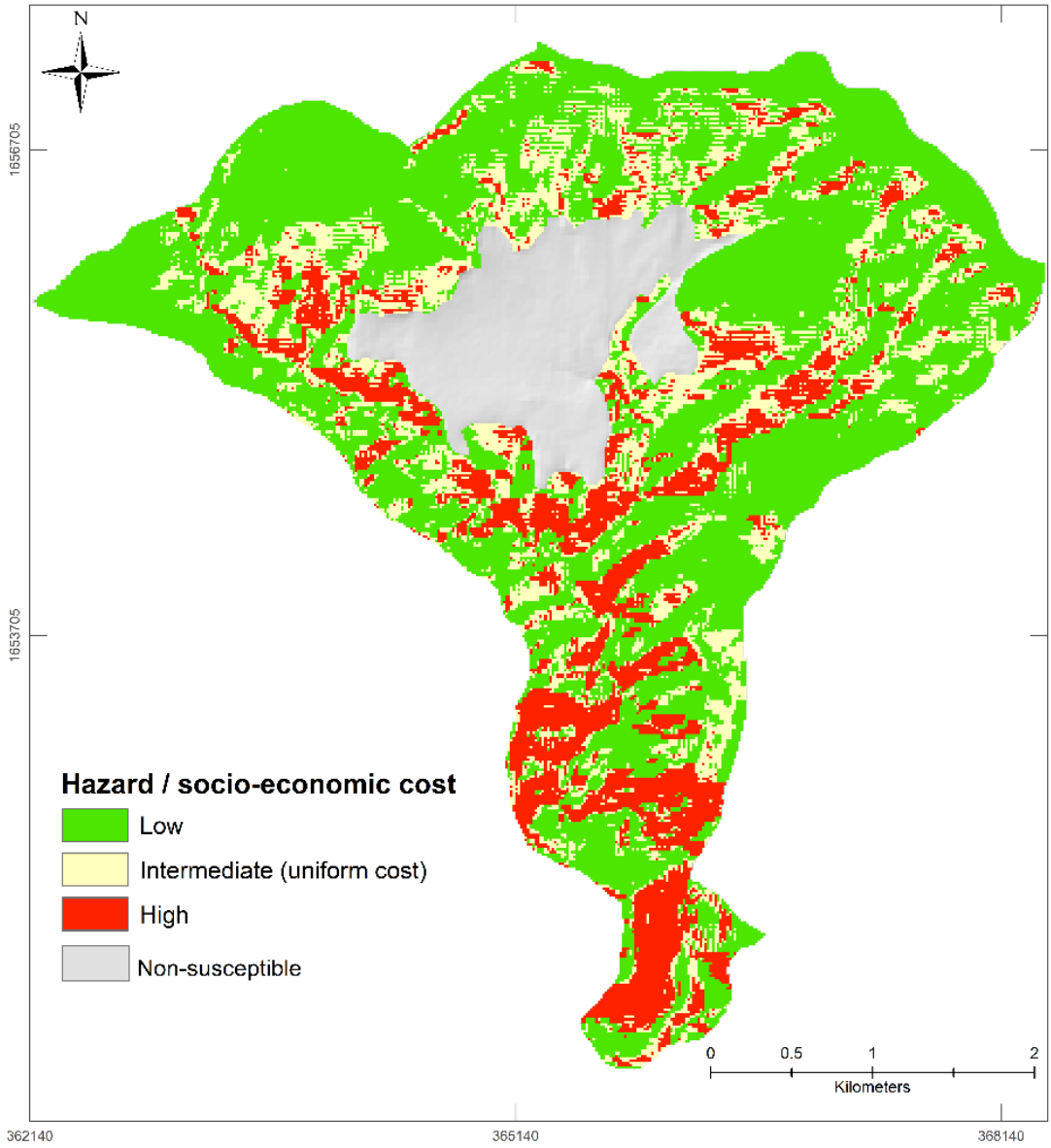


Figure 48 — Susceptibility map associated with cost, shallow landslides sample size 1:1.

Cutoff values associated with cost, shallow landslides (sample size 1:10)

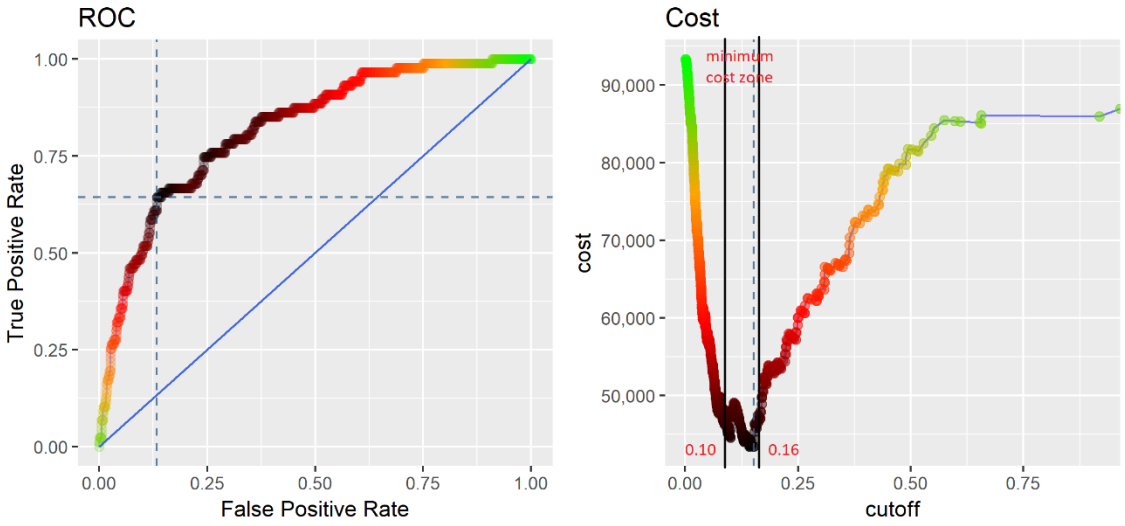


Figure 49 — ROC curve and Cost plot showing the cutoff values in the minimum cost zone, for shallow landslides sample size 1:10. Cyan color dotted line denotes where that optimal point lies and black lines denotes the zone with minimum cost.

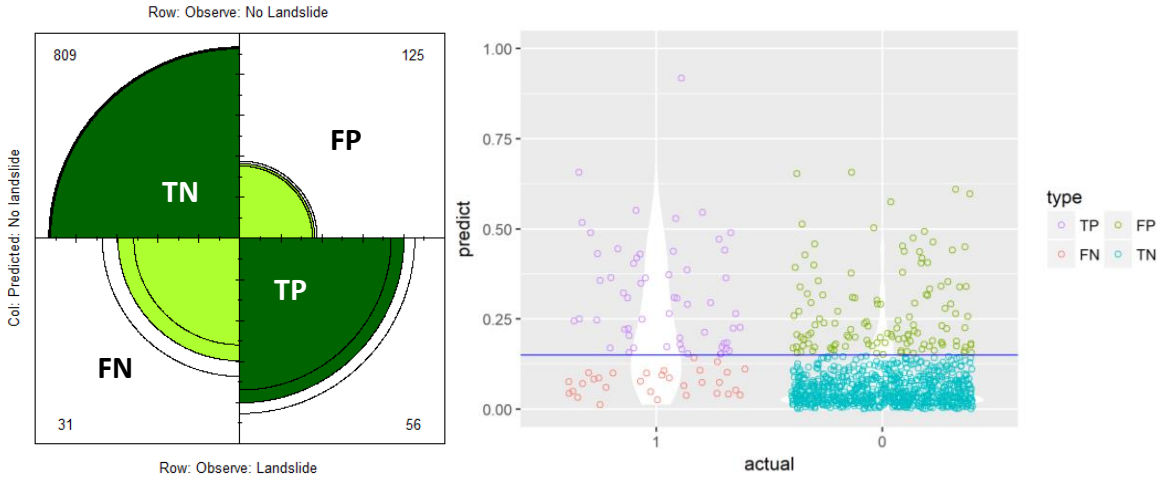


Figure 50 Four-fold and Confusion matrix plots (using minimum cutoff value = 0.15) for "6 variables" model, shallow landslides sample size 1:10.

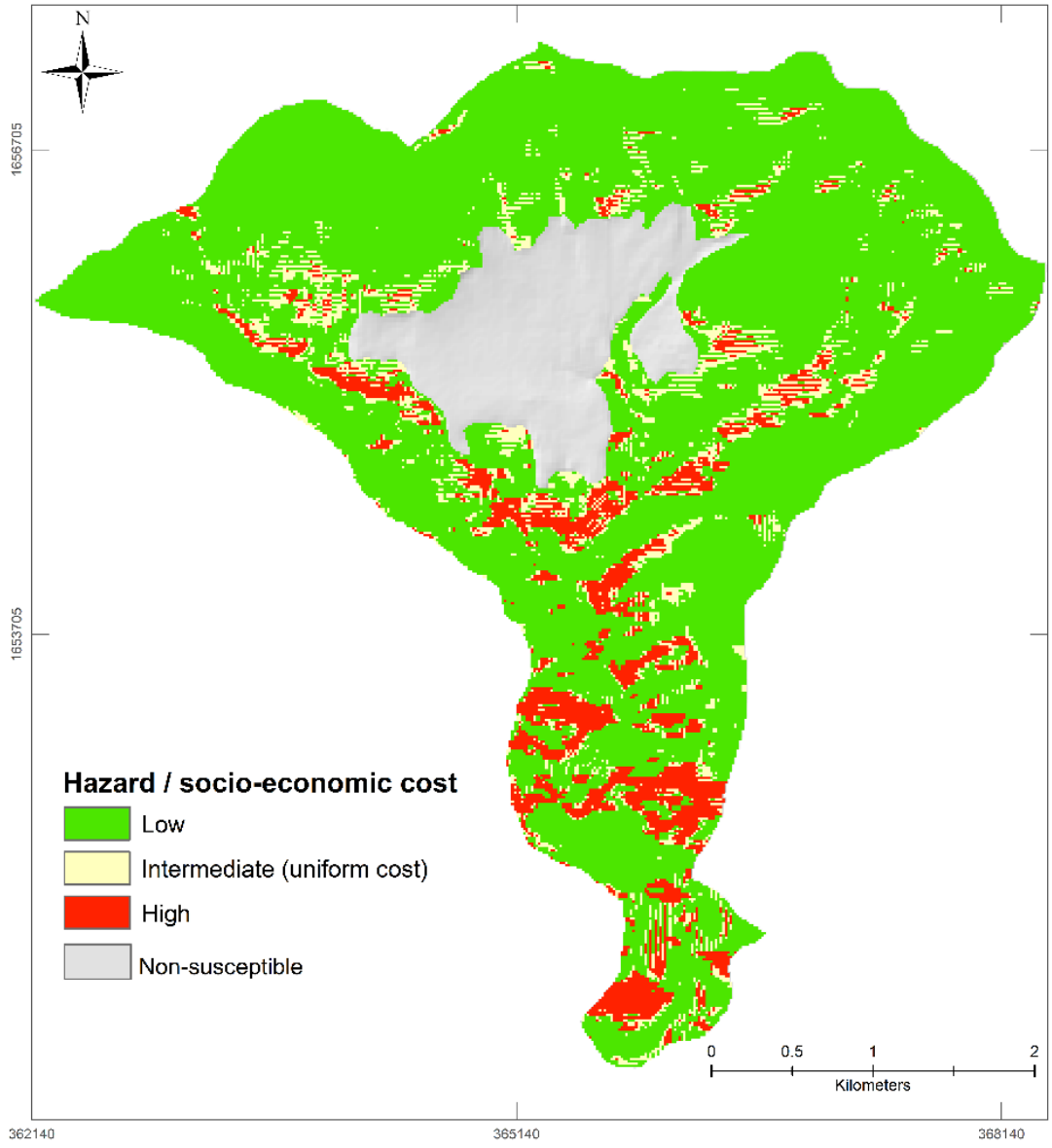


Figure 51— Susceptibility map associated with cost, shallow landslides sample size 1:10.

Cutoff values associated with cost, debris flows

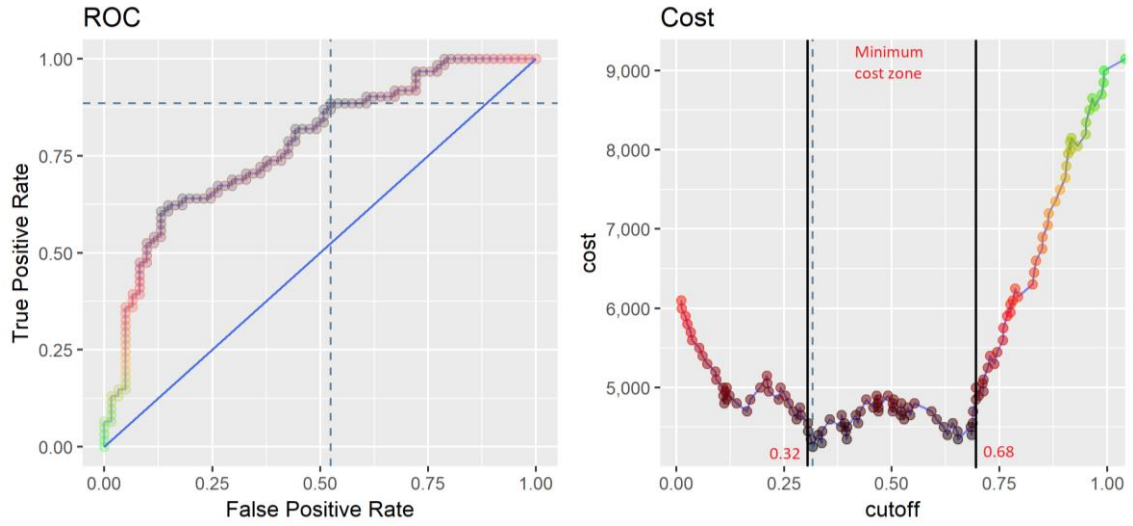


Figure 52 — ROC curve and Cost plot showing the cutoff values in the minimum cost zone, for debris flows. Cyan color dotted line denotes where that optimal point lies and black lines denotes the zone with minimum cost.

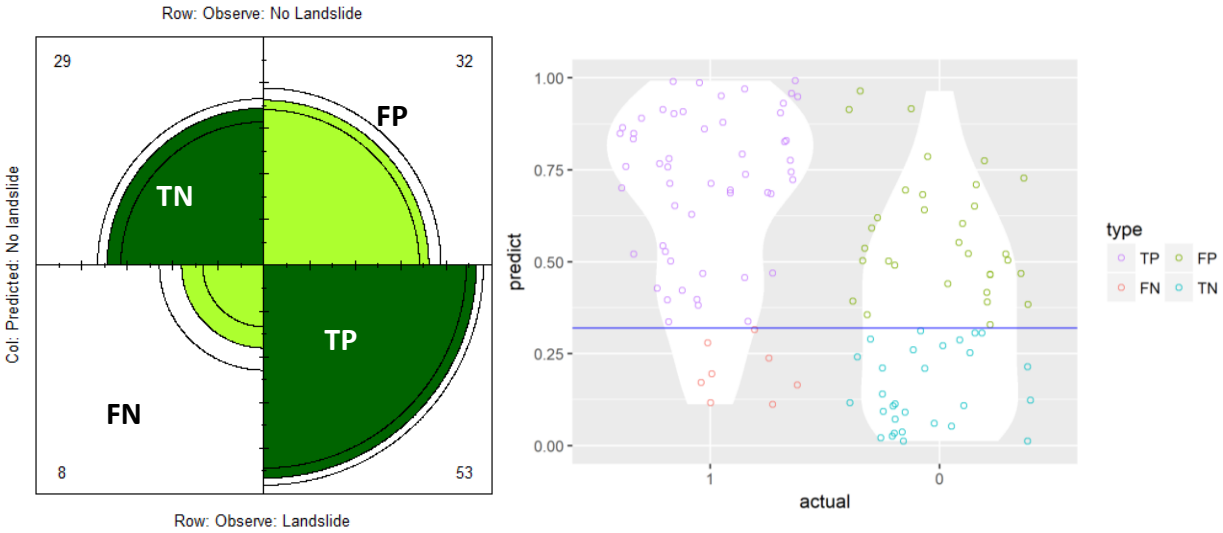


Figure 53 Four-fold and Confusion matrix plots (using minimum cutoff value = 0.32) for "5 variables" model, Debris flows dataset.

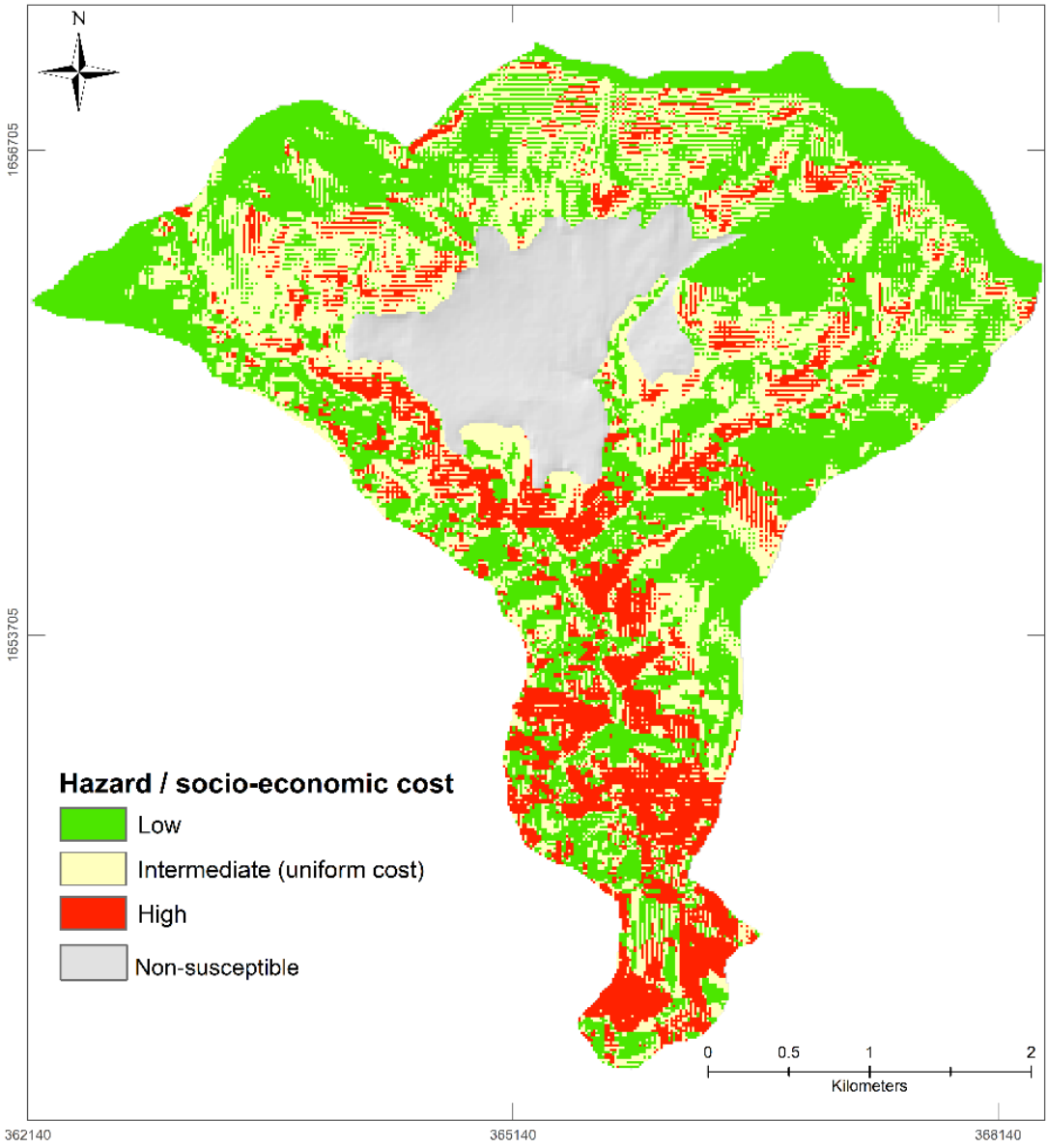


Figure 54— Susceptibility map associated with cost, debris flows.

Cutoff values associated with cost, all landslides

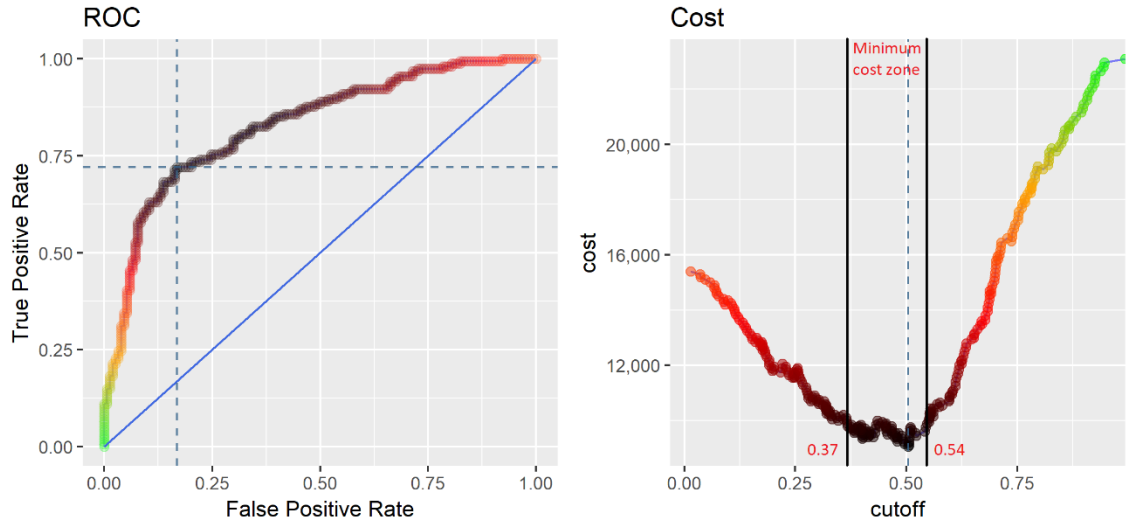


Figure 55 — ROC curve and Cost plot showing the cutoff values in the minimum cost zone, for all landslides. Cyan color dotted line denotes where that optimal point lies and black lines denotes the zone with minimum cost.

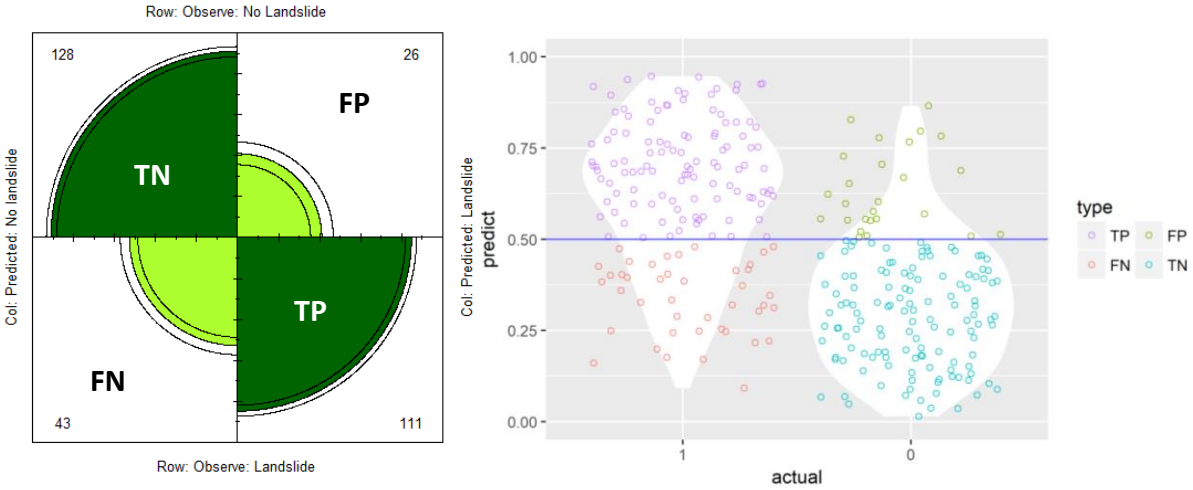


Figure 56 Four-fold and Confusion matrix plots (using minimum cutoff value = 0.5) for "5 variables" model, all landslides dataset.

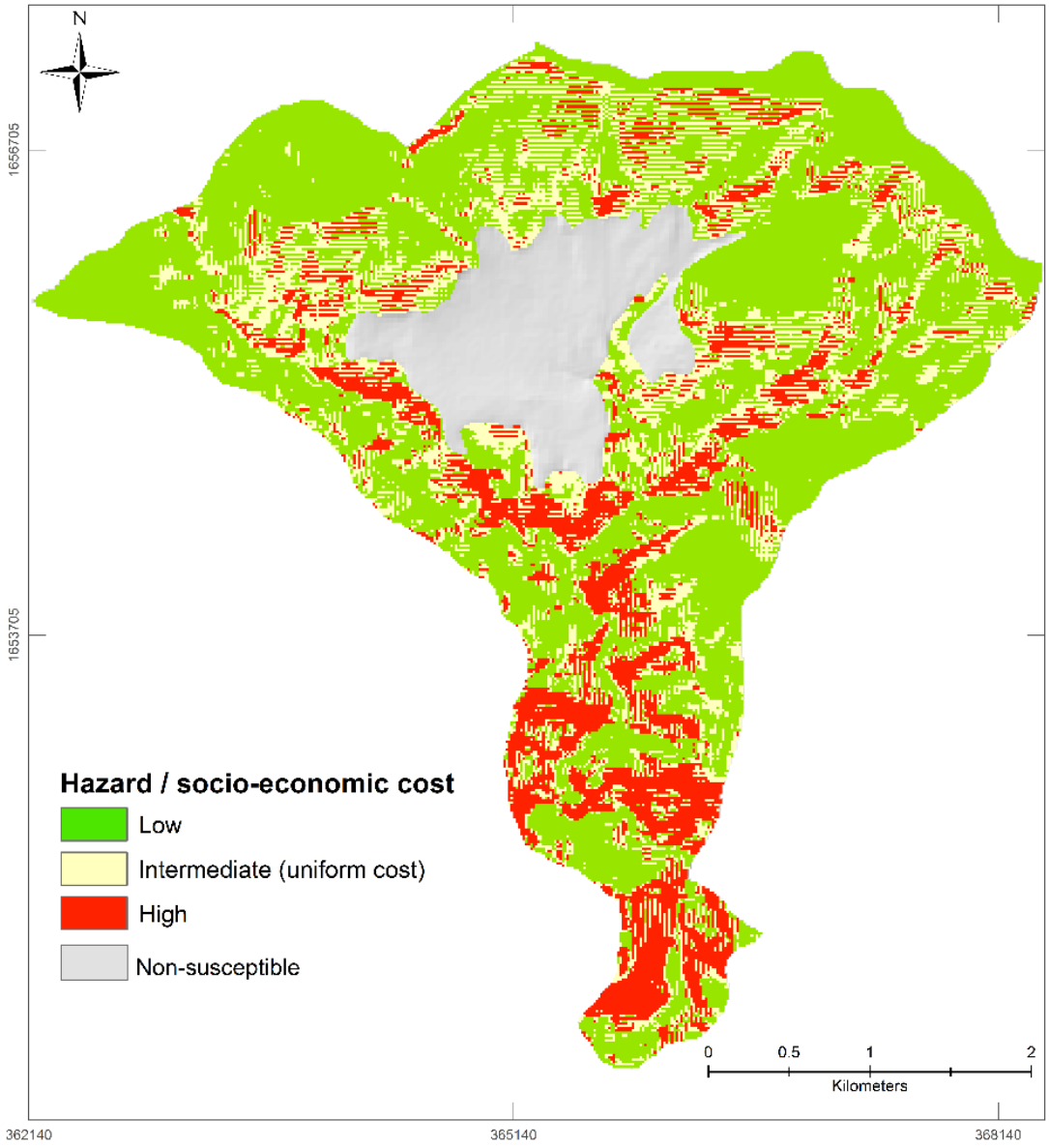


Figure 57 — Susceptibility map associated with cost, all landslides.

7 CONCLUSION AND PERSPECTIVES

The present research take place in an area where landslides are common. Many landslide events occurred in the past almost everywhere in Guatemala. The aims of this research were to create a landslide inventory to analyze the spatial distribution, and to understand the factors that explain the spatial occurrence of landslides triggered by tropical rainfall events in populated mountain areas.

This study, in the Naranjo basin, highlighted that landslides were often triggered by rainfall. With global warming, climate change could result in more frequent and intense rainfall, as what happened in 2005. The possible increase in the frequency of rainfall could therefore increase landslide hazards and, as a consequence, landslide risks. It is important to continue studying this topic and its issues to improve the analysis of general and local factors.

The landslide inventory in the study area allowed for the focus on the susceptibility analysis of shallow landslide occurrences as the main data available in the area. The spatial landslides distribution was related to 12 predisposing factors - slope, aspect, elevation, profile curvature, planform curvature, CDA, distance to drains, distance to roads, lithology, land cover, morphometric region, and landscape fragmentation. The logistic regression method, despite the low quality of data collected to calibrate the model, was efficient to assess landslide susceptibility. Regarding the importance of predisposing factors in shallow landslide susceptibility analysis, this probabilistic approach was able to identify the explanatory variables that are significantly associated with the landslide occurrence. The six most significant variables are slope, elevation, aspect, profile curvature, planform curvature and Euclidean distance to roads.

From a technical perspective, it is useful to explore modeling not only with the starting point of the landslides but with the whole landslides itself. It is also useful to test and validate the Samia *et al*. (2017) methods to Guatemala: *do landslide follow landslides* in Guatemala? It is possible that almost each single square meter of Guatemala had been affected once in history by landslides, thus questioning the interest of focus on small specific area. However, at the scale of one human life, this issue of assessing landslide susceptibility remains, by far, relevant. Samia *et al*. (2017) underlined in their study in Italy that landslides may follow landslides. They also insist that this probability of consecutive landslides at a same place decrease with the time. The interest of mapping landslide sensitivity in specific places is important, especially in Guatemala where a large part is susceptible to landslides. Susceptibility lead to hazard, hazard lead to vulnerability, and vulnerability lead to risk.

The landslide susceptibility assessment leading to an optimal Cutoff value discussion provided a threshold that determines the most landslide prone areas by differentiating the cost of errors type (False Positives and False Negatives).

It is the interest of this research to bridge the gap between scientific maps and practical application of scientific information by creating a map of landslide susceptibility that are accessible to the general public and other stakeholders. Such map is useful to land use planners,

engineers, and decision-makers in choosing suitable locations to implement development programs in Guatemala, including the application of landslide protection measures.

Guatemala is an ideal country to explore landslides in details. Local people have many practical experiences about landslides: when, why, and how do they occur? However, there is a lack of scientific knowledge about landslides in the country, as in many other countries from the "Global South" (Maes *et al*., 2015). The literature review highlighted the few studies about landslides in Guatemala. Paper written in Spanish about this topic dealt more on heuristic and qualitative methods.

The research perspective employed in this study could also be applied in other watersheds in Guatemala. Using the same variables, although not exactly the same coefficients, testing and validating the models in other drainage basins would help to generalize and validate the models at the national scale.

Another relevant perspective would be more related on how people live, feel, know, and perceive landslide in their own living environment. This perspective open bridges between physical and human geography. Indeed, as said above, landslides are more than a natural phenomenon but a natural hazard (Ocampo & Schmitz, 2015). Although it is very important to know the processes that trigger landslides, it is also crucial to have knowledge on how local people deal with its phenomenon.

8 REFERENCES

Alcántara-Ayala, I. (2004). Hazard assessment of rainfall-induced landsliding in Mexico. *Geomorphology, 61*, 19-40.

Alcántara-Ayala, I. (2009). Disasters in Mexico and Central America: A Little Bit More than a Century of Natural Hazards. In: Latrubesse, E.M. (Ed.). *Developments in Earth Surface Processes*, Amsterdam: Elsevier, 75-97.

Alcántara-Ayala, I., & Oliver-Smith, A. (2014). ICL Latin-American Network: on the road to landslide reduction capacity building. *Landslides, 11*, 315-318.

Alcántara-Ayala, I., Esteban-Chávez, O., & Parrot, J. F. (2006). Landsliding related to land-cover change: A diachronic analysis of hillslope instability distribution in the Sierra Norte Puebla, Mexico. *CATENA, 65*, 152-165.

Álvarez-Gómez, J. A. (2009). *Tectónica Activa y Geodinámica en el Norte de Centroamérica*. PhD Thesis, Facultad de Ciencias Geológicas, Madrid: Universidad Complutense de Madrid, 241.

Ardizzone, F., Cardinali, M., Galli, M., Guzzetti, F., & Reichenbach, P. (2007). Identification and mapping of recent rainfall-induced landslides using elevation data collected by airborne Lidar. *Natural Hazards and Earth System Sciences, 7*, 637-650.

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

Baldi, G., Guerschman, J.P., & Paruelo, J.M. (2006). Characterizing fragmentation in temperate South America grasslands. *Agriculture, Ecosystems & Environment, 116*, 197-208.

Beguería, S. (2006). Validation and Evaluation of Predictive Models in Hazard Assessment and Risk Management. *Natural Hazards, 37*, 315-329.

Bommer, J.J., & Rodríguez, C.E. (2002). Earthquake-induced landslides in Central America. *Engineering Geology, 63*, 189-220.

Bonis, S.B., & Williams, H. (2012). *Mapa geológico general de Guatemala*. Scale 1:250000, Guatemala city: Dirección General de Cartografía, revised map from 1964.

Brabb, E.E. (1984). Innovative approaches to landslide hazard mapping. In: *Proceedings 4th International Symposium on Landslides (Toronto),* Vancouver: BiTech Publishers, 307-324.

Brenning, A. (2005). Spatial prediction models for landslide hazards: review, comparison and evaluation. *Natural Hazards and Earth System Sciences, 5*, 853-862.

Broothaerts, N., Kissi, E., Poesen, J., Van Rompaey, A., Getahun, K., Van Ranst, E., & Diels, J. (2012). Spatial patterns, causes and consequences of landslides in the Gilgel Gibe catchment, SW Ethiopia. *CATENA, 97*, 127-136.

Bucknam, R.C., Coe, J.A., Chavarría, M.M., Godt, J.W., Tarr, A.C., Bradley, L., Rafferty, S.A., Hancock, D., Dart, R.L., & Johnson, M.L. (2001). *Landslides triggered by Hurricane Mitch in Guatemala Inventory and Discussion*. U.S. Geological Survey Open File Report 01-443, 39.

Bui, D.T., Lofman, O., Revhaug, I., & Dick, O. (2011). Landslide susceptibility analysis in the Hoa

Binh province of Vietnam using statistical index and logistic regression. *Natural Hazards, 59*, 1413-1444.

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*, 427-445.

Cascini, L., Cuomo, S., & Della Sala, M. (2011). Spatial and temporal occurrence of rainfall-induced shallow landslides of flow type: A case of Sarno-Quindici, Italy. *Geomorphology, 126*, 148-158.

Catani, F., Lagomarsino, D., Segoni, S., & Tofani, V. (2013). Landslide susceptibility estimation by random forests technique: sensitivity and scaling issues. *Natural Hazards and Earth System Sciences, 13*, 2815-2831.

Cavallo, E.A., & Noy, I. (2010). *The Economics of Natural Disasters: A Survey*. Washington D.C.: Inter-American Development Bank. Retrieved on from http://EconPapers.repec.org/RePEc:idb:wpaper:4649, May 2017.

CEPAL - Comisión Económica para América Latina. (2005). *Efectos en Guatemala de las lluvias torrenciales y la tormenta tropical Stan, octubre 2005*. Informe anual 2005, LC/MEX/R.895, Guatemala City: CEPAL, 122.

Chang, K.-T., Chiang, S.-H., & Hsu, M.-L. (2007). Modeling typhoon - and earthquake - induced landslides in a mountainous watershed using logistic regression. *Geomorphology, 89*, 335-347.

Chung, C.-J., & Fabbri, A. (2003). Validation of Spatial Prediction Models for Landslide Hazard Mapping. *Natural Hazards, 30*, 451-472.

Chung, J.-W., Rogers, J.D., & Watkins, C.M. (2014). Estimating severity of seismically induced landslides and lateral spreads using threshold water levels. *Geomorphology, 204*, 31-41.

Coe, J.A., Godt, J.W., Baum, R.L., Bucknam, R.C., & Michael, A. J. (2004). Landslide susceptibility from topography in Guatemala. *London: Taylor & Francis Group,* 69-78.

Corominas, J., van Westen, C.J., Frattini, P., Cascini, L., Malet, J.-P., Fotopoulou, S., Catani, F., van den Eeckhaut, M., Mavrouli, O., Agliardi, F., Pitilakis, K., Winter, M.G., Pastor, M., Ferlisi, S., Tofani, V., Hervas, J., & Smith, J.T. (2014). Recommendations for the quantitative analysis of landslide risk. *Bulletin of Engineering Geology and the Environment, 73*, 209-263.

Costanzo, D., Chacón, J., Conoscenti, C., Irigaray, C., & Rotigliano, E. (2013). Forward logistic regression for earth-flow landslide susceptibility assessment in the Platani river basin (southern Sicily, Italy). *Landslides*, *4*, 1-15.

Crozier, M.J., & Glade, T. (2005). Landslides hazard and risk: Issues, Concepts, and approach. In: Glade, T., Anderson, M., Crozier, M.J. (Eds), *Landslides Hazard and Risk*. Chichester: John Wiley and Sons, 1-39.

Cruden, D.M., & Varnes, D.J. (1996). Landslide Types and Processes. In: Turner, A.K. & Shuster, R.L. (Eds.), *Landslides: Investigation and Mitigation*, Washington D.C.: Transportation Research Board, Special Report No. 247, 36-75.

Dahal, R.K., & Hasegawa, S. (2008). Representative rainfall thresholds for landslides in the Nepal

Himalaya. *Geomorphology, 100*, 429-443.

Dahal, R.K., & Hasegawa, S. (2008). Representative rainfall thresholds for landslides in the Nepal Himalaya. *Geomorphology, 100*, 429-443.

Dai, F.C., & Lee, C.F. (2002). Landslide characteristics and slope instability modeling using GIS, Lantau Island, Hong Kong. *Geomorphology, 42*, 213-228.

de Smith, M.J., Goodchild, M.F., & Longley, P.A. (2007). *Geospatial Analysis: A Comprehensive Guide to Principles, Techniques and Software Tools*. Winchelsea: Troubador Publishing, 748.

Dewitte, O., Daoudi, M., Bosco, C., & Van Den Eeckhaut, M. (2015). Predicting the susceptibility to gully initiation in data-poor regions. *Geomorphology, 228*, 101-115.

Dietrich, W.E., Reiss, R., Hsu, M.-L., & Montgomery, D.R. (1995). A process-based model for colluvial soil depth and shallow landsliding using digital elevation data. *Hydrological Processes, 9*, 383-400.

Dislich, C., & Huth, A. (2012). Modelling the impact of shallow landslides on forest structure in tropical montane forests. *Ecological Modelling, 239*, 40-53.

EERI - Earthquake Engineering Research Institute. (2013). *The November 7, 2012 M7.4 Guatemala Earthquake EERI-AGIES_Guatemala_Report*. Earthquake Engineering Research Institute.

Efremidis, G., Avlonitis, M., Konstantinidis, A., & Aifantis, E.C. (2017). A statistical study of precursor activity in earthquake-induced landslides. *Computers and Geotechnics, 81*, 137-142.

Ercanoglu, M., Gokceoglu, C., & Van Asch, Th.W.J. (2004). Landslide Susceptibility Zoning North of Yenice (NW Turkey) by Multivariate Statistical Techniques. *Natural Hazards*, 32, 1-23.

Fan, L., Lehmann, P., McArdell, B., & Or, D. (2017). Linking rainfall-induced landslides with debris flows runout patterns towards catchment scale hazard assessment. *Geomorphology, 280*, 1-15.

Fu, W., Liu, S., Degloria, S.D., Dong, S., & Beazley, R. (2010). Characterizing the “fragmentation–barrier” effect of road networks on landscape connectivity: A case study in Xishuangbanna, Southwest China. *Landscape and Urban Planning, 95*, 122-129.

Galli, M., Ardizzone, F., Cardinali, M., Guzzetti, F., & Reichenbach, P. (2008). Comparing landslide inventory maps. *Geomorphology, 94*, 268-289.

Gill, J.C., & Malamud, B.D. (2017). Anthropogenic processes, natural hazards, and interactions in a multi-hazard framework. *Earth-Science Reviews, 166*, 246-269.

Girvetz, E.H., Thorne, J.H., Berry, A.M., & Jaeger, J.A.G. (2008). Integration of landscape fragmentation analysis into regional planning: A statewide multi-scale case study from California, USA. *Landscape and Urban Planning, 86*, 205-218.

Glade, T. (2003). Landslide occurrence as a response to land use change: a review of evidence from New Zealand. *CATENA, 51*, 297-314.

Gu, W., Heikkilä, R., & Hanski, I. (2002). Estimating the consequences of habitat fragmentation on extinction risk in dynamic landscapes. *Landscape Ecology, 17*, 699-710.

Guns, M. & V. Vanacker (2014). Shifts in landslide frequency-area distribution after forest conversion in the tropical Andes. *Anthropocene*, *6*, 75-85.

Guzzetti, F., Aleotti, P., Malamud, B. D., Turcotte, D. L. (2003). Comparison of three landslide event inventories in central and northern Italy using frequency-area statistics. In Jansa, A., & Romero, R., (Eds.), *Proceedings of the 4th Plinius Conference on Mediterranean Storms*: Unknown Publisher.

Guzzetti, F., Ardizzone, F., Cardinali, M., Galli, M., Reichenbach, P., & Rossi, M. (2008b). Distribution of landslides in the Upper Tiber River basin, central Italy. *Geomorphology, 96*, 105-122.

Guzzetti, F., Cardinali, M., Reichenbach, P., & Carrara, A. (2000). Comparing Landslide Maps: A Case Study in the Upper Tiber River Basin, Central Italy. *Environmental Management, 25*, 247-263.

Guzzetti, F., Carrara, A., Cardinali, M., & Reichenbach, P. (1999). Landslide hazard evaluation: a review of current techniques and their application in a multi-scale study, Central Italy. *Geomorphology, 31*, 181-216.

Guzzetti, F., Mondini, A.C., Cardinali, M., Fiorucci, F., Santangelo, M., & Chang, K.-T. (2012). Landslide inventory maps: New tools for an old problem. *Earth-Science Reviews, 112*, 42-66.

Guzzetti, F., Peruccacci, S., Rossi, M., & Stark, C.P. (2008a). The rainfall intensity–duration control of shallow landslides and debris flows: an update. *Landslides, 5*, 3-17.

Guzzetti, F., Reichenbach, P., Ardizonne, F., Cardinali, M., & Galli, M. (2009). Landslide hazard assessment, vulnerability estimation and risk evaluation: An example from the Colaazzone area (Central Umbria, Italy. *Geografia Fisica E Dinamica Quaternaria, 32*, 183-192.

Guzzetti, F., Reichenbach, P., Cardinali, M., Galli, M., & Ardizzone, F. (2005). Probabilistic landslide hazard assessment at the basin scale. *Geomorphology, 72*, 272-299.

Harp, E.L, Wieczorek, G.F, & Wilson, R.C. (1978). *Earthquake - induced landslides from the February 4, 1976 Guatemala earthquake and their implications for landslide hazard reduction.* Guatemala: Instituto de Fomento de Hipotecas Aseguradas (FHA), Centro de Estudios Mesoamericanos sobre Tecnología Apropiada (CEMAT).

Heckmann, T., Gegg, K., Gegg, A., & Becht, M. (2014). Sample size matters: investigating the effect of sample size on a logistic regression susceptibility model for debris flows. *Natural Hazards and Earth System Sciences, 14*, 259-278.

Hernandez, G., Leon, R., Salinas, L., & Dimnet, E. (2012). A fragmentation model with neighborhood interaction. *Applied Mathematical Modelling, 36*, 1694-1702.

Highland, L.M., & Bobrowsky, P. (2008). *The landslide handbook—A guide to understanding landslides.* Reston, Virginia, U.S.: Geological Survey Circular, 1325, 129.

Hobbs, N.T., Galvin, K.A., Stokes, C.J., Lackett, J.M., Ash, A.J., Boone, R.B., Reid, R.S., Thornton, P.K. (2008). Fragmentation of rangelands: Implications for humans, animals, and landscapes. *Global Environmental Change, 18*, 776-785.

Holm, A.McR, Cridland, S.W., & Roderick, M.L. (2003). The use of time-integrated NOAA NDVI data and rainfall to assess landscape degradation in the arid shrubland of Western Australia.

Remote Sensing of Environment, 85, 145-158.

Horton, P., Jaboyedoff, M., & Bardou, E. (2008). Debris flow susceptibility mapping at a regional scale. In: Locat J., Perret D., Turmel D., Demers D., Leroueil S. (Eds). *Proceedings of the 4th Canadian Conference on Geohazards – From Causes to management.* Québec: Presse de l'Université Laval, 8.

Hosmer, D.W., & Lemeshow, S. (2000). *Applied Logistic Regression* (2nd ed.). New York: Wiley, 369

Hungr, O. (2005). Classification and terminology. In: Jakob, M. & Hungr, O. (Eds.). *Debris-flow Hazards and Related Phenomena*. Berlin Heidelberg: Springer-Verlag, 9-23.

Hungr, O., Leroueil, S., & Picarelli, L. (2014). The Varnes classification of landslide types, an update. *Landslides, 11*, 167-194.

Husid, R. (1978). The february 4, 1976 Earthquake in Guatemala City and Vicinity. *Engineering Field Report, Guatemala.*

Hussin, H.Y., Zumpano, V., Reichenbach, P., Sterlacchini, S., Micu, M., van Westen, C., & Bălteanu, D. (2016). Different landslide sampling strategies in a grid-based bi-variate statistical susceptibility model. *Geomorphology, 253*, 508-523.

Hutchinson, J.N. (1968). Mass Movement. In: Fairbridge, RW. (Ed.), *The Encyclopedia of Geomorphology*, New York: Reinhold Book Corp, 688-696.

Hutchinson, J.N. (1988). Morphological and geotechnical parameters of landslides in relation to geology and hydrogeology, state-of-the-art report. In: Bonnard, C. (Ed*.), Proceedings of the Fifth International Symposium on Landslides*. Rotterdam: Balkema, 1, 3-35.

Hutchinson, J.N. (1995). Keynote paper: landslide hazard assessment. In: Bell, DH. (Ed.), *Landslides*. Rotterdam: Balkema, 1805-1841.

Hutchinson, J.N., & Chandler, M.P. (1991). A preliminary landslide hazard zonation of the Undercliff of the Isle of Wight. In: Chandler, R.J. (Ed.), *Slope Stability Engineering*. London: Thomas Telford, 197-205.

IFRC - International Federation of Red Cross and Red Crescent Societies. (2012). *Emergency appeal final report Guatemala: Tropical Storm Agatha*. International Federation of Red Cross and Red Crescent Societies -IFRC-, 16.

IFRC - International Federation of Red Cross and Red Crescent Societies. (2013). *Emergency appeal operation update Guatemala Tropical Depression 12-E*. International Federation of Red Cross and Red Crescent Societies -IFRC-, 10.

INSIVUMEH - Instituto Nacional de Sismología Vulcanología Meteorología e Hidrología. (2005). *Resumen del impacto asociado al huracan "Stan" en Guatemala*. Guatemala: INSIVUMEH, 5.

INSIVUMEH - Instituto Nacional de Sismología Vulcanología Meteorología e Hidrología. (2006). *Distribución Espacial de Eventos Sísmicos en Guatemala Período de registro: 1984-2005*. Guatemala: INSIVUMEH, 4.

INSIVUMEH - Instituto Nacional de Sismología Vulcanología Meteorología e Hidrología. (2010). *Resumen del Impacto Meteorológico tormenta tropical Aghata mayo 2010,* Guatemala:

INSIVUMEH, 9.

INSIVUMEH - Instituto Nacional de Sismología Vulcanología Meteorología e Hidrología. (2011). *Report Impacto Meteorológico Temporal de lluvias Octubre 2011*, Guatemala. INSIVUMEH, 8.

INSIVUMEH - Instituto Nacional de Sismología Vulcanología Meteorología e Hidrología. (2016). *Weather Database of Guatemala*.

Iovine, G.R., Greco, R., Gariano, S., Iaquinta, P., Pellegrino, A., & Terranova, O. (2013). Shallow-Landslide Susceptibility in the Costa Viola Mountain Ridge (Italia). In: Margottini, C., Canuti, P., & Sassa, K. (Eds.), *Landslide Science and Practice,* Berlin Heidelberg: Springer, 81-87.

Jaboyedoff, M., Michoud, M., Derron, M.-H., Voumard, J., Leibundgut, G., Sudmeier-Rieux, K., Nadim, F., & Leroi, E. (2016). Human-Induced Landslides: Toward the analysis of anthropogenic changes of the slope environment. In: Avresa, S., Cascini, L., Picarelli, L., & Scavia, C. (Eds.), *Landslides and Engineering Slopes – Experiences, Theory and practices*. Leiden: CRC Press, 217-232.

Jaeger, J. (2000). Landscape division, splitting index, and effective mesh size: new measures of landscape fragmentation. *Landscape Ecology, 15*, 115-130.

Kamusoko, C., & Aniya, M. (2007). Land use/cover change and landscape fragmentation analysis in the Bindura District, Zimbabwe. *Land Degradation & Development, 18*, 221-233.

Kannan, M., Saranathan, E., & Anabalagan, R. (2013). Landslide vulnerability mapping using frequency ratio model: a geospatial approach in Bodi-Bodimettu Ghat section, Theni district, Tamil Nadu, India. *Arabian Journal of Geosciences, 6*, 2901-2913.

Keefer, D. (1984). Landslides caused by earthquakes. *Geological Society of America Bulletin, 95*, 406-421.

Keefer, D. (2002). Investigating Landslides Caused by Earthquakes – A Historical Review. *Surveys in Geophysics, 23*, 473-510.

Keefer, D. (2013). 5.11 Landslides Generated by Earthquakes: Immediate and Long-Term Effects. In: Shroder, J.F. (Ed.), *Treatise on Geomorphology,* San Diego: Academic Press, 250-266.

King, G., & Zeng, L. (2001). Logistic Regression in Rare Events Data. *Political Analysis, 9*, 137-163.

Kirschbaum, D., Adler, R., Hong, Y., Hill, S., & Lerner-Lam, A. (2010). A global landslide catalog for hazard applications: method, results, and limitations. *Natural Hazards, 52*, 561-575.

Kirschbaum, D., Stanley, T., & Zhou, Y. (2015). Spatial and temporal analysis of a global landslide catalog. *Geomorphology, 249*, 4-15.

Korup, O., & Stolle, A. (2014). Landslide prediction from machine learning. *Geology Today, 30*, 26-33.

Korup, O., Densmore, A.L., & Schlunegger, F. (2010). The role of landslides in mountain range evolution. *Geomorphology, 120*, 77-90.

Lee, C.F., Tham, L.G., Shum, W.L., Dai, F.C., & Ng, K. . (2004). Logistic regression modelling of storm-induced shallow landsliding in time and space on natural terrain of Lantau Island, Hong Kong. *Bulletin of Engineering Geology and the Environment, 63*, 315.

Lee, S., & Talib, J.A. (2005). Probabilistic landslide susceptibility and factor effect analysis. *Environmental Geology, 47*, 982-990.

Lei, C.I. (2012). Earthquake-Triggered Landslides. In: *1st Civil and Environmental Engineering Student Conference,* 25-26 June 2012, London: Imperial College.

Ligorría, J.P., Lindholm, H., Bungum, H., & Dahle, A. (1995). Seismic Hazard for Guatemala. *NORSAR Technical Report*, N°2-21, 47.

Lin, C.W., Chang, W.S., Liu, S.H., Tsai, T.T., Lee, S.P., Tsang, Y.C., Shieh, C.L., & Tseng, C.M. (2011). Landslides triggered by the 7 August 2009 Typhoon Morakot in southern Taiwan. *Engineering Geology, 123*, 3-12.

Lin, C.-W., Liu, S.-H., Lee, S.-Y., & Liu, C.-C. (2006). Impacts of the Chi-Chi earthquake on subsequent rainfall-induced landslides in central Taiwan. *Engineering Geology, 86*, 87-101.

Lin, Y.-P., Chu, H.-J., & Wu, C.-F. (2010). Spatial pattern analysis of landslide using landscape metrics and logistic regression: a case study in Central Taiwan, *Hydrology and Earth System Sciences Discussions*, *7*, 3423-3451.

Maes, J., Kervyn, M., Vranken, L., Dewitte, O., Vanmaercke, M., Mertens, K., Jacobs, L., Poesen, J. (2015). Landslide risk reduction strategies: an inventory for the Global South. *Geophysical Research Abstracts*, 17, EGU2015-5988.

Malamud, B., Turcotte, D., Guzzetti, F., & Reichenbach, P. (2004). Landslide inventories and their statistical properties. *Earth Surface Processes and Landforms, 29*, 687-711.

Matumoto, T., & Latham, G. (1976). Aftershocks of the Guatemalan Earthquake of February 4, 1976. *Geophysical Research Letters, 3*, 599-602.

Melchiorre, C., Matteucci, M., Azzoni, A., & Zanchi, A. (2008). Artificial neural networks and cluster analysis in landslide susceptibility zonation. *Geomorphology, 94*, 379-400.

Mikoš, M., Četina, M., & Brilly, M. (2004). Hydrologic conditions responsible for triggering the Stože landslide, Slovenia. *Engineering Geology, 73*, 193-213.

Ming, G., Wenbing, Y., Mingguo, M., & Xin, L. (2008). Study on the oasis landscape fragmentation in northwestern China by using remote sensing data and GIS: a case study of Jinta oasis. *Environmental Geology, 54*, 629-636.

Mondini, A.C., Guzzetti, F., Reichenbach, P., Rossi, M., Cardinali, M., & Ardizzone, F. (2011). Semi-automatic recognition and mapping of rainfall induced shallow landslides using optical satellite images. *Remote Sensing of Environment, 115*, 1743-1757.

Monterroso, D. (2003). *Statistical Seismology Studies in Central America, b-value, seismic hazard and seismic quiescence*. Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology, Upsala: Acta Universitatis Upsaliensis, *897*, 27.

Moosavi, V., Talebi, A., & Shirmohammadi, B. (2014). Producing a landslide inventory map using pixel-based and object-oriented approaches optimized by Taguchi method. *Geomorphology, 204*, 646-656.

Nagendra, H., Munroe, D.K., & Southworth, J. (2004). From pattern to process: landscape fragmentation and the analysis of land use/land cover change. *Agriculture, Ecosystems &*

Environment, 101, 111-115.

National Census. (2002). *Caracteristicas de la poblacion y de los locales de habitacion censados*, Guatemala City: National Institute of statistic, 278.

Nefeslioglu, H.A., Gokceoglu, C., & Sonmez, H. (2008). An assessment on the use of logistic regression and artificial neural networks with different sampling strategies for the preparation of landslide susceptibility maps. *Engineering Geology, 97*, 171-191.

NOAA - National Oceanic and Atmospheric Administration. (2016). *National Geophysical Data Center / World Data Service (NGDC/WDS): Significant Earthquake Database*. Retrieved on https://www.ngdc.noaa.gov/nndc/struts/form?t=101650&s=1&d=1, Septembre 2016.

O'brien, R. (2007). A Caution Regarding Rules of Thumb for Variance Inflation Factors. *Quality & Quantity, 41*, 673-690.

Ocampo, L.A., & Schmitz, S. (2015). Making "Place" count in Risk perception studies. In: Firmino, A. (Ed.), *Smart Answers for a smiling future: 23rd Annual Colloquium on Sustainable Rural Systems*, Lisbon: Universidade Nova de Lisboa.

OCHA - United Nations Office for the Coordination of Humanitarian Affairs. (2012). *Guatemala: Earthquake November 2012*.

Ohlmacher, G. (2007). Plan curvature and landslide probability in regions dominated by earth flows and earth slides. *Engineering Geology, 91*, 117-134.

Ohlmacher, G., & Davis, J. (2003). Using multiple logistic regression and GIS technology to predict landslide hazard in northeast Kansas, USA. *Engineering Geology, 69*, 331-343.

Orozco, E. (2007). *Hydrological study High part Watershed Naranjo river, Guatemala*. Guatemala: Fundación Solar, 57.

Ozdemir, A. (2011). GIS-based groundwater spring potential mapping in the Sultan Mountains (Konya, Turkey) using frequency ratio, weights of evidence and logistic regression methods and their comparison. *Journal of Hydrology, 411*, 290-308.

Ozdemir, A., & Altural, T. (2013). A comparative study of frequency ratio, weights of evidence and logistic regression methods for landslide susceptibility mapping: Sultan Mountains, SW Turkey. *Journal of Asian Earth Sciences, 64*, 180-197.

Paolini, L., Villalba, R., & Ricardo Grau, H. (2005). Precipitation variability and landslide occurrence in a subtropical mountain ecosystem of NW Argentina. *Dendrochronologia, 22*, 175-180.

Pardeshi, S., Autade, S., & Pardeshi, S. (2013). Landslide hazard assessment: recent trends and techniques. *Springer Plus, 2*, 523.

Park, D.W., Nikhil, N.V., & Lee, S.R. (2013). Landslide and debris flow susceptibility zonation using TRIGRS for the 2011 Seoul landslide event. *Natural Hazards and Earth System Sciences, 13*, 2833-2849.

Parker, R.N., Hales, T.C., Mudd, S.M., Grieve, S.D., & Constantine, J.A. (2016). Colluvium supply in humid regions limits the frequency of storm-triggered landslides. *Scientific Reports, 6*, 34438.

Peel, M.C., Finlayson, B.L., & McMahon, T.A. (2007). Updated world map of the Köppen-Geiger climate classification. *Hydrology and Earth System Sciences, 11*, 1633-1644.

Peng, C.-Y.J., & So, T.-S.H. (2002). Logistic Regression Analysis and Reporting: A Primer. *Understanding Statistics, 1*, 31-70.

Persichillo, M.G., Bordoni, M., & Meisina, C. (2017). The role of land use changes in the distribution of shallow landslides. *Science of the Total Environment, 574*, 924-937.

Picard, M. (2007). *Legal Issues from the International Response to Tropical Storm Stan in Guatemala*. *A case study commissioned by the International Federation of Red Cross and Red Crescent Societies' International Disaster Response Laws, Rules and Principles (IDRL) Programme.* International Federation of Red Cross and Red Crescent Societies, 42.

Porfido, S., Esposito, E., Spiga, E., Sacchi, M., Molisso, F., & Mazzola, S. (2015). Impact of Ground Effects for an Appropriate Mitigation Strategy in Seismic Area: The Example of Guatemala 1976 Earthquake. In: Lollino, G. *et al*. (Eds.), *Engineering Geology for Society and Territory (Volume 2),* Basel: *S*pringer, 703-708.

Pourghasemi, H.R., Moradi, H.R., Mohammadi, M., Pradhan, B., Mostafazadeh, R., & Jirandeh, A. Goli. (2012). Landslide Hazard Assessment Using Remote Sensing Data, GIS and Weights-of-evidence Model (South of Golestan Province, Iran). In: Lee, W. (Ed.), *Asia Pacific Conference on Environmental Science and Technology (APEST 2012*) - *Advances in Biomedical Engineering*, February 1-2 2012, Kuala Lumpur: APEST, 30-36.

Rossi, M., Guzzetti, F., Reichenbach, P., Mondini, A.C., & Peruccacci, S. (2010a). Optimal landslide susceptibility zonation based on multiple forecasts. *Geomorphology, 114*, 129-142.

Rossi, M., Witt, A., Guzzetti, F., Malamud, B.D., & Peruccacci, S. (2010b). Analysis of historical landslide time series in the Emilia-Romagna region, northern Italy. *Earth Surface Processes and Landforms, 35*, 1123-1137.

Salciarini, D., Godt, J., Savage, W., Conversini, P., Baum, R., & Michael, J. (2006). Modeling regional initiation of rainfall-induced shallow landslides in the eastern Umbria Region of central Italy. *Landslides, 3*, 181-194.

Samia, J., Temme, A., Bregt, A., Wallinga, J., Guzzetti, F., Ardizzone, F., & Rossi, M. (2017). Do landslides follow landslides? Insights in path dependency from a multi-temporal landslide inventory. *Landslides*, *14*, 547-558.

SEGEPLAN - Secretaría de Planificación y Programación de la Presidencia. (2010a). *Municipal development plan, San Pedro Sacatepéquez, San Marcos, Guatemala*. Guatemala: SEGEPLAN, 137.

SEGEPLAN - Secretaría de Planificación y Programación de la Presidencia. (2010b). *Municipal development plan, San Antonio Sacatepéquez, San Marcos, Guatemala*. Guatemala: SEGEPLAN, 119.

Sepúlveda, S.A., & Petley, D.N. (2015). Regional trends and controlling factors of fatal landslides in Latin America and the Caribbean. *Natural Hazards and Earth System Sciences, 15*, 1821-1833.

Serrano, M., Sanz, L., Puig, J., & Pons, J. (2002). Landscape fragmentation caused by the transport

network in Navarra (Spain): Two-scale analysis and landscape integration assessment. *Landscape and Urban Planning, 58*, 113-123.

Shabanzadeh, K., Suroor, J., & Mohammadi, T. (2011). Zoning Landslide by Use of Frequency Ratio Method (Case Study: Deylaman Region). *Middle-East Journal of Scientific Research, 9*, 578-583.

Sidle, R.C., & Bogaard, T.A. (2016). Dynamic earth system and ecological controls of rainfall-initiated landslides. *Earth-Science Reviews, 159*, 275-291.

Sidle, R.C., Ziegler, A.D., Negishi, J., Nik, A.R., Siew, R., & Turkelboom, F. (2006). Erosion processes in steep terrain—Truths, myths, and uncertainties related to forest management in Southeast Asia. *Forest Ecology and Management, 224*, 199-225.

Silverman, J. (2011). *"The Need Obliged Us": Culture as Capacity during the Hurricane Stan Emergency Response. A Case Study from Tectitán, Huehuetenango, Guatemala*. Michigan Technological University: ProQuest LLC, 99.

Simon N., de Róiste, M., Crozier, M., & Rafek, A.G. (2017). Representing Landslides as Polygon (Areal) or Points? How Different Data Types Influence the Accuracy of Landslide Susceptibility Maps. *Sains Malaysiana, 46*, 27-34.

Soeters, R., & Van Westen, C.J. (1996). Slope instability recognition analysis and zonation. In: Turner, K.T., Schuster, R.L. (Eds.), *Landslides: Investigation and Mitigation*, Special Report, Washington, DC.: Transportation Research Board National Research Council, 247, 129-177.

Solaimani, K., Mousavi, S.Z., & Kavian, A. (2013). Landslide susceptibility mapping based on frequency ratio and logistic regression models. *Arabian Journal of Geosciences, 6*, 2557-2569.

Southworth, J., Munroe, D., & Nagendra, H. (2004). Land cover change and landscape fragmentation--comparing the utility of continuous and discrete analyses for a western Honduras region. *Agriculture, Ecosystems & Environment, 101*, 185-205.

Süzen, M.L., & Doyuran, V. (2004). Data driven bivariate landslide susceptibility assessment using geographical information systems: a method and application to Asarsuyu catchment, Turkey. *Engineering Geology, 71*, 303-321.

Tomaselli, V., Tenerelli, P., & Sciandrello, S. (2012). Mapping and quantifying habitat fragmentation in small coastal areas: a case study of three protected wetlands in Apulia (Italy). *Environmental Monitoring and Assessment, 184*, 693-713.

Trigila, A., Iadanza, C., Esposito, C., & Scarascia-Mugnozza, G. (2015). Comparison of Logistic Regression and Random Forests techniques for shallow landslide susceptibility assessment in Giampilieri (NE Sicily, Italy). *Geomorphology, 249*, 119-136.

Turner, A.K., & Schuster, R.L. (1996). *Landslides: investigation and mitigation*. Washington D.C: National Academy Press, 673.

Turner, T.R., Duke, S.D., Fransen, B.R., Reiter, M.L., Kroll, A.J., Ward, J.W., Bach, J.L., Justice, T.E., & Bilby, R.E. (2010). Landslide densities associated with rainfall, stand age, and topography on forested landscapes, southwestern Washington, USA. *Forest Ecology and Management, 259*, 2233-2247.

UNISDR - United Nations Office for Disaster Risk Reduction. (2009). *Global assessment report on*

disaster risk reduction: Risk and poverty in a changing climate. Genève: UNISDR, 207.

UNISDR - United Nations Office for Disaster Risk Reduction. (2011). *Global assessment report on disaster risk reduction: Revealing risk, redefining development*. Genèves: UNISDR, 178.

UNISDR - United Nations Office for Disaster Risk Reduction. (2013). *Global assessment report on disaster risk reduction: From shared risk to shared value, the business case for disaster risk reduction*. Genève: UNISDR, 288.

UNISDR - United Nations Office for Disaster Risk Reduction. (2015). *Global assessment report on disaster risk reduction: Making development sustainable, the future of disaster risk management*. Genève: UNISDR, 316.

Vanacker, V., Vanderschaeghe, M., Govers, G., Willems, E., Poesen, J., Deckers, J., & De Bievre, B. (2003). Linking hydrological, infinite slope stability and land-use change models through GIS for assessing the impact of deforestation on slope stability in high Andean watersheds. *Geomorphology,* 52, 299-315.

Van Asch, T.W.J., Buma, J., & Van Beek, L.P.H. (1999). A view on some hydrological triggering systems in landslides. *Geomorphology, 30*, 25-32.

Van Den Eeckhaut, M. (2006). *Spatial and temporal patterns of landslides in hilly regions - The Flemish Ardennes (Belgium)*. PhD Thesis, Faculty of Sciences, Department of Geography–Geology, Leuven: K.U. Leuven, 250.

Van Den Eeckhaut, M., Marre, A., & Poesen, J. (2010). Comparison of two landslide susceptibility assessments in the Champagne–Ardenne region (France). *Geomorphology, 115*, 141-155.

Van Westen, C.J., Castellanos, E., & Kuriakose, S.L. (2008). Spatial data for landslide susceptibility, hazard, and vulnerability assessment: An overview. *Engineering Geology, 102*, 112-131.

Vanmaercke, M., Ardizzone, F., Rossi, M., & Guzzetti, F. (2017). Exploring the effects of seismicity on landslides and catchment sediment yield: An Italian case study. *Geomorphology, 278*, 171-183.

Varnes, D.J. (1978). Slope movement types and processes In: Schuster, R. L., & Krizek, R. J. (Eds.), *Special Report 176: Landslides: Analysis and Control*. Washington D.C: Transportation and Road Research Board, National Academy of Science, 11-33.

Vega, J.A., & Hidalgo, C.A.. (2016). Quantitative risk assessment of landslides triggered by earthquakes and rainfall based on direct costs of urban buildings. *Geomorphology, 273*, 217-235.

Villagran, M., Lindholm, C., Dahle, A., Cowan, H., & Bungum, H. (1996). Seismic hazard assessment for Guatemala City. *Natural Hazards, 14*, 189-205.

Wang, L.-J., Guo, M., Sawada, K., Lin, J., & Zhang, J. (2015). Landslide susceptibility mapping in Mizunami City, Japan: A comparison between logistic regression, bivariate statistical analysis and multivariate adaptive regression spline models. *CATENA, 135*, 271-282.

Wang, L.-J., Sawada, K., & Moriguchi, S. (2013). Landslide susceptibility analysis with logistic regression model based on FCM sampling strategy. *Computers & Geosciences, 57*, 81-92.

Weng, M.-C., Wu, M.-H., Ning, S.-K., & Jou, Y.-W. (2011). Evaluating triggering and causative factors of landslides in Lawnon River Basin, Taiwan. *Engineering Geology, 123*, 72-82.

Wu, Y., Li, W., Wang, Q., Liu, Q., Yang, Q., Xing, M., Pei, Y., Yan, S. (2016). Landslide susceptibility assessment using frequency ratio, statistical index and certainty factor models for the Gangu County, China. *Arabian Journal of Geosciences, 9*, 84.

Yalcin, A., Reis, S., Aydinoglu, A.C., & 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*, 274-287.

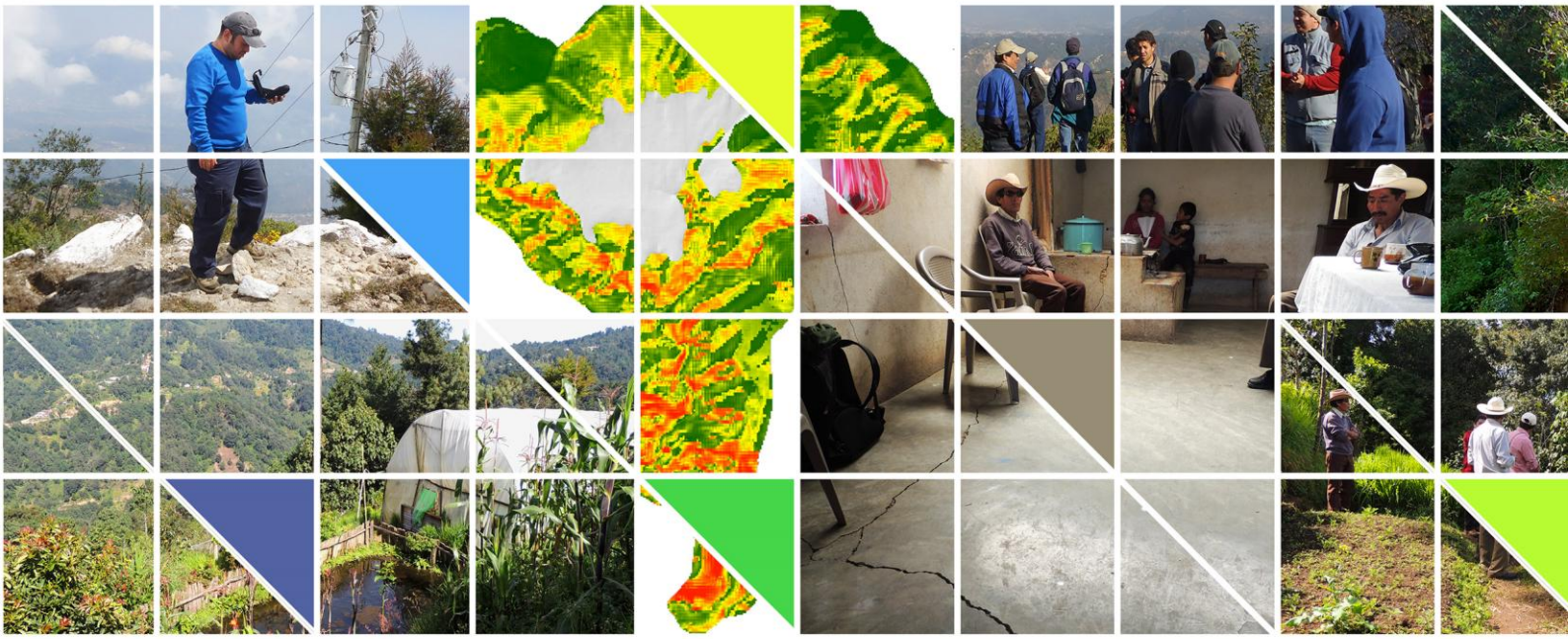
Yani-Quiyuch, R.O. (2015). Seismic Monitoring in Guatemala. In: Frassetto, A. *et al*. (Eds.), *National Geophysical Networks in Latin America – Best Practices, Challenges, and Opportunities for Collaboration,* May 25-29 2015, Santiago, Chile: IRIS Workshop, 8.

Zhang, S., Zhang, L.M., & Glade, T. (2014). Characteristics of earthquake- and rain-induced landslides near the epicenter of Wenchuan earthquake. *Engineering Geology, 175*, 58-73.

Zhu, D., Ren, Q., Xuan, Y., Chen, Y., & Cluckie, I. D. (2013). An effective depression filling algorithm for DEM-based 2-D surface flow modelling. *Hydrology and Earth System Sciences, 17*, 495.

9 List of figures

10 List of tables



In tropical mountain areas, landslides triggered by heavy rainfall represent a natural hazard, especially where meteorological events such as tropical cyclones are recurrent. Rainfall-triggered landslides may also increase in highly seismic active areas. This is the case of Guatemala, located in the Circum-Pacific Belt and between Pacific and Atlantic Oceans, sources of cyclonic events. Although Guatemala is a region where landslides are a widespread phenomenon, landslide inventories, maps and quantitative susceptibility assessments that are useful for land use planners and decision makers are hitherto non-existent. The aim of this research is to produce quantitative landslide susceptibility assessments, using logistic regression multivariate statistical method. The study area is a watershed located in the department of San Marcos (western Guatemala), impacted by a tropical storm event in October 2005. A total of 766 landslides were identified and mapped using orthoimages from 2005. Then, 99 landslides were mapped in 2011 based on field data. The comparison of the models and the associated susceptibility maps highlighted 6 significant input variables that are associated with landslide occurences. A susceptibility map was generated to display the results of the models in terms of probability values. A proposal and discussion on the operational use of susceptibility maps were also made. These will help stakeholders in decision-making and in implementing protective measures.

University of Liège - Department of Geography
3 Clos Mercator - 4000 Liège - Belgium